Statistics Open Test Solutions FAMAT State Convention 2021

1. **D.** The mean of a geometric random variable is $\frac{1}{p}$ and the variance is $\frac{1-p}{p^2}$. The sum of these is $\mathbf{\frac{1}{p^2}}$.

2. **A.** P(Type I Error) + P(Type II error) = 0.17. The level of significance is 0.05 which is equivalent to P(Type I Error). Therefore, P(Type II Error) is equal to 0.12. The power of a test is 1 – P(Type II Error) = 1 – 0.12 = **0.88**

3. **A.** Number of arrangements for a keychain with such a clasp is $\frac{n!}{2} = \frac{6!}{2} = \frac{720}{2}$ = **360**

4. **B.** There are 12 places where BOURTS can be placed. The rest of the letters can be arranged $\frac{11!}{2!2!2!}$ ways because of the repeated letters 'A', 'L' and 'T'. so, the final answer is the product $\frac{12*11!}{2!2!2!}$ = **59,875,200.**

5. **E.** Some of these values can be negative and the ones that cannot be negative can be equal to zero. Therefore, **none** of the eight statistics are always positive.

6. **B.** There are three sections. One from 3 to 4, another from 4 to 5 and finally a section from 5 to k. From 3 to 4, there is a height of 0.1, from 4 to 5 there is a height of 0.4, and from 5 to k the height is 1. The area must sum to 1, so the area is 1(0.1) + 1(0.4) + x = 1 And x = 0.5. So, the value of k is 5 + 0.5 = **5.5**

7. **C.** I. True because $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$
   II. True because $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$
   III. True because $\mu_{X+Y} = \mu_X + \mu_Y$
   IV. **False because variances add but standard deviations do NOT!**

8. **E.** The computation of a confidence interval is NOT appropriate to estimate the parameter μ because Alex asks EVERY ONE of the 14 competitors and therefore μ can be computed directly and estimating it is unnecessary because he conducted a census.

9. **B**. The correlation coefficient is computed from $\sqrt{\frac{\Sigma(y-\bar{y})^2 - \Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2}} = \sqrt{\frac{4.2}{14}}$ = **0.548**

10. **A**. For the two balls to be the same color, they can both be white, black or red. Since Tomer is choosing both without replacement, the correct probability is found by
    P(WW or RR or BB) = $\frac{2}{9} * \frac{1}{8} + \frac{4}{9} * \frac{3}{8} + \frac{3}{9} * \frac{2}{8} = \mathbf{\frac{5}{18}}$

11. **B.** The prime numbers that Elijah can roll are 2, 3, 5, 7, 11, 13, 17, and 19. Therefore, the probability of rolling a prime number on the die is $\frac{8}{20} = 0.4$. The standard deviation of this geometric situation is given by $\sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{0.6}{0.4^2}}$. This simplifies to $\mathbf{\frac{\sqrt{15}}{2}}$.

12. **B.** The definition of a 90% confidence level is that we expect 90% of the confidence intervals constructed from statistics using random samples of the same size to capture the true parameter value. In this case, the parameter value we are looking for is the **population mean.**

13. **B**. Draw a three-circle Venn Diagram. After filling in the given information, there will be four sections that are unknown. Physics and Biology only (let this be a), Chemistry and Biology only (let this be b), Physics and Chemistry only (let this be c), and the intersection of all three classes, which we shall call x (and this is what we seek). Now, since 40 total students are in only two of the three classes, we have that a + b + c = 40. Also, x can be expressed in the following three different ways: x = 19 – a, x = 24 – b, and x = 15 – c. Thus, a + b + c = 19 – x + 24 – x + 15 – x = 40 which simplifies into 58 – 3x = 40 and so x = **6.**

14. **A.** When a fair coin is tossed five times, the probability of at least three heads is $\frac{1}{2}$. This can be seen by the following probability distribution:

X (number of heads)	0	1	2	3	4	5
P(X)	1/32	5/32	10/32	10/32	5/32	1/32

The probability of at least 25 successes is binomial with n = 40, p = 0.5. The correct probability can be found by 1 – binomialcdf (40, 0.5, 24) = **0.0769**

The five missing probabilities in the two tables are solved for as follows:
Since a + b = 1 – 0.65 = 0.35, we have b = 0.35 – a. Thus, given that ac = 0.08 and bc = 0.06, we have that 0.08 / a = 0.06 / (0.35 – a). Solving for a we get a = 0.02 and so b = 0.15. This helps us get c = 0.08 / 0.20 = 0.40 and d = 0.03 / 0.20 = 0.15 and e = 1 – 0.90 = 0.10. The final results:

A	1	3	5	7	9
P(A)	0.20	0.10	0.15	0.25	0.30

B	2	4	6	8	10
P(B)	0.40	0.15	0.32	0.03	0.10

Hence, the mean and variance for random variables A and B, in the form (mean, variance), are (5.7, 8.91) and (4.56, 6.6464) respectively.

15. **A.** Mean is found by 2(5.7) + 3(4.56) and variance is found by 4(8.91) + 9(6.6464).
Therefore, **(25.08, 95.4576).**

16. **C.** The variance of random variable (A + B) is $8.91 + 6.6464 + 2(-0.92)\sqrt{8.91} \times \sqrt{6.6464} =$ 1.396834968… So, the standard deviation is $\sqrt{1.396834968\ldots} = 1.1819$

17. **A.** The confidence interval is $12 \pm \frac{2.009575199 * 0.69}{\sqrt{50}}$, which makes the interval (11.804, 12.196). 2.009575199 is found from invT (0.975, 49) because this is a t confidence interval with df = 49.

18. **C.** $\bar{p} = \frac{p_1 * n_1 + p_2 * n_2}{n_1 + n_2}$, so the standard error for a two-proportion significance test is given by $\sqrt{\bar{p}\,(1 - \bar{p}\,) * (\frac{1}{n_1} + \frac{1}{n_2})}$. Plugging in the appropriate values gives $\mathbf{\frac{\sqrt{3}}{16}}$**.**

19. **A.** P(A) = $\frac{1}{28}$ and P(B) = $\frac{1}{2}$. P(A U B) = P(A) + P(B) – P(A and B) = $\frac{1}{28} + \frac{1}{2} - \frac{1}{28} \times \frac{1}{2} = \frac{29}{56}$
$P(A^C | B^C) = P(A^C)$ because if events A and B are independent, then their complements are also independent. $P(A^C) = \frac{27}{28}$. Finally, $\left| \frac{29}{56} - \frac{27}{28} \right| = \mathbf{\frac{25}{56}}$

20. **C.** t = $\frac{b-3.5}{SE_b}$. We need to find SE_b. Using the calculator test menu and the fact that when running the test with β=0 we get t = $\frac{4.25}{SE_b}$ = 14.8347057. Thus SE_b = .2864903481. Going back to the equation for this problem, we get t = $\frac{b-3.5}{SE_b} = \frac{.75}{.2864903481}$ = 2.6179. P-value = tcdf(2.6179, 99999, 7) = **0.017.**

21. **C.** There are 10! ways to arrange the fruit. There are 9! Ways to arrange all the fruit except for one pear. There are six possible choices for the last pear. The probability is $\frac{6*9!}{10!} = \frac{2}{5} = \mathbf{0.6000}$.

22. **E.** Statement II is the only true statement. Statement I is false because the p-value of 0.178 / 2 = 0.089 is still greater than the alpha level of 0.05. Statement III is false because statement II is true. Statement IV is false because power is equal to 1 – P(Type 2 error), NOT 1 – P(Type 1 error). Finally, statement V is false because the alpha level is the probability of incorrectly rejecting the null hypothesis when it is in fact true.

23. **B.** There are 2020 students and the combined number of Calculus and Statistics students is 2180. Therefore, the minimum number of students that take both Calculus and Statistics is 160. Not all students necessarily take only one of the two courses. It is possible that ALL 930 Statistics students also take Calculus, and this would be the maximum value of x. Therefore, max – min = 930 – 160 = **770.**

24. **A.** The expected cell count in a two-way Chi-Square test for independence is
$\frac{(row\ total)(column\ total)}{grand\ total} = \frac{5(17)}{39} = \mathbf{2.18}$

25. **A.** This is one-sample t test because the population standard deviation is unknown and the statistic is computed by $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{2.65-2}{\frac{0.75}{\sqrt{14}}} = 3.24277\ and\ p-value = tcdf(3.24277, 999999, 13) = 0.0032.$

26. **B.** This question is trying to determine if the student understands "confidence level." If so, the only answer that makes sense is choice B. If the population mean is 29, then our constructed confidence interval does not include it. Yet, this should occur 99% of the time in repeated samples. So, for it to not be within the interval constructed using our sample mean would indeed be unlikely. *I'm imagining possible disputes to E because of the phrase "must be true" but it being unlikely MUST be true if it's only going to occur 1% of the time. This question is modeled after a MC question from the 2002 AP Stats exam.

27. **D.** The farther away the "alternative" value of the parameter is from the hypothesized one in H_O, the easier it is to detect. Also, increasing sample size increases power, as does increasing the alpha level. Choice D maximizes these factors.

28. **A**. The LSRL is $\widehat{weight}$ = -266.53 + 6.1376(height in inches). Plugging in 75 inches we get a predicted weight of 193.79 pounds. Residuals are computed by observed – expected, so 285 – 193.79 = **91.21.**

29. **B.** $SE_b = \frac{s}{s_x\sqrt{n-1}}$ so $S_x = \frac{s}{SE_b\sqrt{n-1}} = \frac{8.641}{.7353\sqrt{31}}$ = **2.111**

30. **A.** The sample slope and SE_b are found in the computer printout. The critical t value is found using the invT function with 30 df and the value needed is 2.75. Plugging into the confidence interval formula b $\pm\ tSE_b$ = 6.1376 $\pm$ 2.75(.7353) = **6.1376 $\pm$ 2.022075**