FAMAT February Regional Statistics Test Solutions

1. $P(T|S) = 0.7 = \dfrac{P(T \cap S)}{P(S)} = \dfrac{\alpha}{\alpha + \delta} \Rightarrow \dfrac{\alpha}{\delta} = \dfrac{7}{3}$

   $P(T|\bar{S}) = 0.6 = \dfrac{P(T \cap \bar{S})}{P(\bar{S})} = \dfrac{\beta}{\beta + \gamma} \Rightarrow \dfrac{\beta}{\gamma} = \dfrac{3}{2}$

   $\dfrac{\alpha\beta}{\delta\gamma} = \dfrac{7}{3} \cdot \dfrac{3}{2} = \dfrac{7}{2}$ **(B)**

2. $Var(XY) = E((XY)^2) - (E(XY))^2 = E(X^2Y^2) - (E(X) \cdot E(Y))^2 =$
   $E(X^2) \cdot E(Y^2) - (0 \cdot 0)^2 = [Var(X) + (E(X))^2] \cdot [Var(Y) + (E(Y))^2] - 0 =$
   $(1+0) \cdot (1+0) = 1 \Rightarrow \ln(1+2) \approx 1.09 \Rightarrow Answer = 0$ **(B)**

3. $z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \dfrac{300/350 - 235/300}{\sqrt{\left(\dfrac{300+235}{350+300}\right)\left(1 - \dfrac{300+235}{350+300}\right)\left(\dfrac{1}{350} + \dfrac{1}{300}\right)}} \approx 2.4583$ **(B)**

4. Let m be the median.
   $1.5 \cdot IQR = 1.5 \cdot (2m - 0.5m) = 2.25m$
   $b = 2m + 2.25m = 4.25m$, $c = 0.5m - 2.25m = -1.75m$
   $b - c = 4.25m - (-1.75m) = 6m$
   Therefore $b - c$ is 6 times the median and thus $x = 6$. **(C)**

5. In order for the perimeter of the square to be greater than 10, S must be greater than 2.5 or less than -2.5.
   $P(S > 2.5 \cap S < -2.5) = P(S > 2.5) + P(S < -2.5) = 2P(z > 2.5) \approx 0.0124$ **(C)**

6. Ms. Henin can get at least one of her two serves in, in one of two ways. She can make here first serve or she can miss her first serve and make her second serve.
   $P(Gets\ at\ least\ one\ serve\ in) = P(1st\ serve\ in) + P(1st\ serve\ out) \cdot P(2nd\ serve\ in)$
   $= a + (1-a)b = (1-b) + b \cdot b = b^2 - b + 1 \Rightarrow A = 1, B = -1, C = 1 \Rightarrow A + B + C = 1$ **(B)**

7. When a log transformation is needed you will see a residual plot with positive residuals at the left and right hand sides of the independent variable range with negative residuals in the middle as the least squares regression line will over predict values in the middle and under predict values towards the edges of the independent variables range. **(B)**

8. A regression yields $E(y \mid x = X)$. If $X = 2$, then the stock is expected to decrease by 10% according to the regression. Therefore Nick will lose 10%, or \$1000, of the money in his 401(k) which is all in Bank of America stock. **(D)**

9. The square of the correlation coefficient is the coefficient of determination, which gives the percentage of variation in the dependent variable that is explained by the regression.
$\frac{s_y}{s_x} r = b \Rightarrow \frac{0.06}{1} r = -0.055 \Rightarrow r^2 \approx 84.03\%$ **(B)**

10. X is a geometric variable with probability of success $1/n$. Y is a binominal variable with parameters $(n, 1/n)$.
$E(X) = 4E(Y) \Rightarrow \frac{1}{1/n} = 4n\frac{1}{n}\left(1-\frac{1}{n}\right) \Rightarrow n = 4\left(1-\frac{1}{n}\right) \Rightarrow n^2 - 4n + 4 = 0 \Rightarrow n = 2$
$\ln 2 \approx 0.693147 \Rightarrow Answer = 6$ **(E)**

11. If 2.5% of the time it takes Ned longer than 60 seconds to touch the victim for the second time, then:
$\Phi^{-1}(0.975) = \frac{60-\mu}{\sigma} \Rightarrow 1.96 = \frac{60-55}{\sigma} \Rightarrow \sigma \approx 2.55$ **(E)**

12. The mean is to the left of the median when the distribution is skewed to the left. The mean is decreased by a few very small observations. **(C)**

13. This is an example of a systematic sample. **(C)**

14. There are 10! different ways to randomly place 10 people in 10 spots. If you group the girls into one cluster, then there are 8 "people" which can be place at random and therefore 8! ways to place them in a straight line with the girls being together. For each of these 8! ways, the girls can be placed in 3! ways within the cluster. Thus the answer is:
$\frac{8!\cdot 3!}{10!} = \frac{1}{15}$. **(A)**

15. Blocking divides subjects into groups which are more homogenous and thus reduces variation. **(A)**

16. $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow .872 \pm 1.96\sqrt{\frac{.872(.128)}{500}} \Rightarrow (.8427, .9013)$ **(A)**

17. The sample mean is normally distributed with mean 70 and standard deviation of $\sqrt{25}/\sqrt{n}$. So n must equal 50 for the variance of the sample mean to be 0.5. And $\ln(10 \cdot 50) \approx 6.21 \Rightarrow Answer = 2.$ **(C)**

18. The probability of at least one complete pass in the next five attempts is equal to 1 minus the probability of no complete passes in the next five attempts. The number of complete passes in the next five attempts is a binomial random variable with parameters (5, p)**.**
$1 - {}_5C_0 p^0 (1-p)^5 = 0.99 \Rightarrow p = 0.6019$ **(B)**

19. $H_0 : \mu = 6, H_a : \mu > 6$

$$p - value = P\left(z > \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right) = P\left(z > \frac{7-6}{4.8/\sqrt{144}} \right) = P(z > 2.5) \approx 0.0062 \quad \textbf{(C)}$$

20. Let *a* be the length of leg one and let *b* be the length of side two. Both *a* and *b* are discrete random variables with the same distribution and they are independent. So:

$$E(a) = E(b) = 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{1}{5} = 3$$

$$E(Area) = E(0.5ab) = 0.5 \cdot E(a) \cdot E(b) = 0.5 \cdot 3 \cdot 3 = 4.5 \quad \textbf{(E)}$$

21. Let Y_i equal the number rolled on die *i*. We know that $E[Y_i] = 3.5$. Your expected profit on any iteration of the game is $E\left[\sum_{i=1}^{n} Y_i - x\right]$. If you are expected to break even in the long run, then this expected profit must equal 0.

$$E\left[\sum_{i=1}^{n} Y_i - x\right] = 0 \Rightarrow \sum_{i=1}^{n} E[Y_i] - E[x] = 0 \Rightarrow nE[Y_i] = E[x] \Rightarrow n3.5 = x \Rightarrow \frac{n}{x} = \frac{1}{3.5} \Rightarrow \frac{n}{x} = \frac{2}{7}$$

(B)

22. $P(A \cup B) = P(A) + P(B) - P(A \cap B) \Rightarrow 1 = P(A) + P(A) - P(A)^2 \Rightarrow 0 = P(A)^2 - 2P(A) + 1$
$\Rightarrow P(A) = 1 \Rightarrow P(A \cap B) = P(A)P(B) = 1 \cdot 1 = 1$ **(D)**

23. $\mu_{x+y} = \mu_x + \mu_y \Rightarrow 10 = \mu_x + \mu_y$

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2} \Rightarrow 10 = \sqrt{\mu_x^2 + \mu_y^2}$$

$$100 = \mu_x^2 + (10 - \mu_x)^2 \Rightarrow 0 = 2\mu_x^2 - 20\mu_x \Rightarrow \mu_x = 0 \text{ or } \mu_x = 10 \Rightarrow |\mu_x - \mu_y| = 10 \quad \textbf{(E)}$$

24. If the two variables were independent we would see the following cell counts.

280(100)/450 = 62.2	110(100)/450 = 24.4	60(100)/450 = 13.3
280(350)/450 = 217.8	110(350)/450 = 85.6	60(350)/450 = 46.7

$$\chi^2 = \frac{(57-62.2)^2}{62.2} + \frac{(26-24.4)^2}{24.4} + \frac{(17-13.3)^2}{13.3} + \frac{(223-217.8)^2}{217.8} + \frac{(84-85.6)^2}{85.6} + \frac{(43-46.7)^2}{46.7} \approx 2.02 \quad \textbf{(B)}$$

25. Under the null hypothesis of independence, the χ^2 statistic has a χ^2 distribution with (2-1)(3-1) = 2 degrees of freedom. Looking at the χ^2 table we see critical value for p=0.25 is 2.77. Since 2.02 is less than 2.77, the p-value for the hypothesis test is greater than 0.25. **(A)**

26. The differences are 7, -1, 3, 4, -5, 4, and 5. The mean of the differences 17/7 and the sample standard deviation of the differences is approximately 4.07664.

$$t^* = \frac{17/7 - 0}{4.07664 / \sqrt{7}} \approx 1.58$$ **(D)**

27. The difference of the sample means is normally distributed with $\mu = 115 - 123 = -8$ and $\sigma = \sqrt{\frac{30^2}{60} + \frac{27^2}{60}}$.

$$z = \frac{0 - (-8)}{\sqrt{\frac{30^2}{60} + \frac{27^2}{60}}} = 1.54 \Rightarrow P(z > 1.534) \approx .0618$$ **(D)**

28. The level of significance is equal to the probability of committing a Type I error, so x = 0.05. Thus y = 0.07. The power is equal to 1 minus the probability of committing a Type II error given a particular true alternative hypothesis. Thus the power equals 0.07. **(C)**

29. The slope of the least squares regression line follows a t-distribution with n-2 degrees of freedom. **(C)**

30. ${}_{14}C_{12} \cdot 0.9^{12} \cdot 0.1^2 \approx 0.2570$ **(E)**