

Answer Key:

1. D

2. C

3. B

4. D

5. D

6. C

7. A

8. C

9. A

10. C

11. B

12. A

13. E

14. B

15. A

16. B

17. E

18. D

19. D

20. E

21. B

22. A

23. B

24. A

25. C

26. B

27. A

28. C

29. D

30. D

1. **D:** The calculations for this question can be simplified by using a linear transformation. Since the smallest value of the data set is 15, we can transform by subtracting 15 from each value in the dataset and then adding 15 back to the mean. This gives a new dataset: {6, 0, 4, 1, 2, 3, 2, 2, 13, 6}. The mean of this dataset is $\frac{0+1+2+2+2+3+4+6+6+13}{10} = \frac{39}{10} = 3.9$, so the original mean is $15 + 3.9 = 18.9$. The variance of the new set will be the same as the original set. The variance is $\frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{0+1+4+4+4+9+16+36+36+169 - \frac{39^2}{10}}{9} = \frac{279 - \frac{1521}{10}}{9} = \frac{1269}{90} = 14.1$. Then, $14.1 + 18.9 = 33$.

2. **C:** Jake takes an SRS of 5 strata (the couples) and samples everybody within each strata. Thus, he performed a cluster sample.

3. **B:** The number of ways to move from (0, 0) to (m, n) is equivalent to the number of ways to arrange m steps to the right and n steps to left, which is $\binom{m+n}{m}$. The number of ways to go from (0, 0) to (5, 6) passing through (2, 2) is the number of ways to go (0, 0) → (2, 2) times the number of ways to go from (2, 2) → (5, 6); $\binom{2+2}{2} \binom{3+4}{4} = \binom{4}{2} \binom{7}{4} = 6 * 35 = 210$. Similarly, the number of ways to go to (0, 0) → (5, 6) while passing through (3, 4) is $\binom{3+4}{4} \binom{2+2}{2} = 210$. The number of ways to get to (5, 6) while passing through both (2, 2) and (3, 4) is $\binom{2+2}{2} \binom{1+2}{2} \binom{2+2}{2} = \binom{4}{2} \binom{3}{2} \binom{4}{2} = 6 * 3 * 6 = 108$. The number of ways to get to (5, 6) while passing through (2, 2) or (3, 4) is the number of ways through (2, 2) plus the number of ways through (3, 4) minus the number of ways through both. This is $210 + 210 - 108 = 312$.

4. **D:** This is a binomial situation with $p = 0.4$. $np = 10 \Rightarrow n = \frac{10}{0.4} = 25$.

5. **D:** While it may appear that we have to find the variance of the *Exponential*($\lambda = 5$) distribution, it is not necessary. $\frac{Var(X_1+X_2)-Var(X_1)}{Var(2X_1)} = \frac{V[X_1]+V[X_2]+2\rho SD[X_1]SD[X_2]-V[X_1]}{4V[X_1]} = \frac{2\rho SD[X_1]SD[X_2]+V[X_2]}{4V[X_1]} = \frac{2\rho+1}{4}$ as X_1 and X_2 follow the same distribution and thus have the same standard deviation and variance. Then, $E\left[\frac{2\rho+1}{4}\right] = \int_0^1 \frac{2\sqrt{\rho^2+1}}{4} d\rho^2 = \int_0^1 \frac{2\sqrt{u+1}}{4} du = \frac{1}{4} + \frac{2}{4} \left[\frac{2}{3} u^{\frac{3}{2}}\right]_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$.

6. **C:** In order to be a valid probability distribution, $\int f_C(x)dx = \int_0^5 \frac{c}{1+\tan\sqrt{2}\left(\frac{\pi x}{10}\right)} dx = 1$. Substituting $u = \frac{\pi x}{10}$ gives $\frac{10c}{\pi} \int_0^{\frac{\pi}{2}} \frac{du}{1+\tan\sqrt{2}u}$. Calling just the integral $I = \int_0^{\frac{\pi}{2}} \frac{du}{1+\tan\sqrt{2}u}$ and substituting $u = \frac{\pi}{2} - v$ gives $I = \int_{\frac{\pi}{2}}^0 \frac{-dv}{1+\tan\sqrt{2}\left(\frac{\pi}{2}-v\right)} = \int_0^{\frac{\pi}{2}} \frac{dv}{1+\frac{1}{\tan\sqrt{2}v}}$ so, changing the auxiliary variable back to u yields $2I = \int_0^{\frac{\pi}{2}} \frac{du}{1+\tan\sqrt{2}u} + \int_0^{\frac{\pi}{2}} \frac{\tan\sqrt{2}u du}{1+\tan\sqrt{2}u} = \int_0^{\frac{\pi}{2}} \frac{1+\tan\sqrt{2}u}{1+\tan\sqrt{2}u} du = \int_0^{\frac{\pi}{2}} 1 du = \frac{\pi}{2}$. Thus, $\frac{10c}{\pi} \left(\frac{\pi}{2}\right) = 1 \Rightarrow c = \frac{2}{5}$

7. **A:** $CI = PE \pm Z^* * MoE = \bar{x} \pm Z_{0.95}^* * \frac{\sigma_x}{\sqrt{n}} = 475 \pm (1.645) * \frac{96}{\sqrt{576}} = 475 \pm 1.645 * 4 = 4.75 \pm 6.580$, where the value of $Z_{0.95}^*$ is achieved either by memory, or the Z table.

8. C: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} e^{(\lambda-t)x} x^{\alpha-1} dx$. Using the substitution $u = (\lambda - t)x$ gives

$M_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\lambda-t}\right)^\alpha \int_0^{\infty} e^{-u} u^{\alpha-1} du$. Note that $\frac{e^{-u} u^{\alpha-1}}{\Gamma(\alpha)}$ is the Gamma distribution with $\lambda = 1$, so

$$M_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\lambda-t}\right)^\alpha \int_0^{\infty} e^{-u} u^{\alpha-1} du = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\lambda-t}\right)^\alpha (\Gamma(\alpha)) = \lambda^\alpha \left(\frac{1}{\lambda-t}\right)^\alpha = \left(\frac{\lambda}{\lambda-t}\right)^\alpha = \left(\frac{1/2}{1/2-t}\right)^4 = \frac{1}{(1-2t)^4}.$$

9. A: $P(\text{North FL} | UM) = \frac{P(\text{North FL} \cap UM)}{P(UM)} = \frac{3/121}{48/121} = \frac{1}{16}$.

10. C: $Exp = \frac{Row * Col}{Table} = \frac{22 * 44}{121} = 8$. Then, the contribution is $\frac{(10-8)^2}{8} = \frac{4}{8} = \frac{1}{2}$.

11. B: This is the probability of a derangements of 5 objects, so the formula $\sum_{r=0}^5 \frac{(-1)^r}{r!} = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} = \frac{11}{30}$ can be used. Alternatively, the total number of ways to distribute the presents is $5! = 120$. The number of ways to do so such that present i is correctly gifted is $(5 - 1)! = 4! = 24$. Similarly, the number of ways such that i and j are fixed is $(5 - 2)! = 3! = 6$; i, j, k fixed is $(5 - 3)! = 2$; i, j, k, l fixed is $(5 - 4)! = 1$; and all fixed is 1. The total number of ways to distributed such that at least one gift is given correctly is $24\binom{5}{1} - 6\binom{5}{2} + 2\binom{5}{3} - \binom{5}{4} + \binom{5}{5} = 76$. Thus, the probability of a derangement is $1 - \frac{76}{120} = \frac{44}{120} = \frac{11}{30}$.

12. A:

$$\begin{aligned} P(4|R) &= \frac{P(4|R)P(4)}{P(R)} \\ &= \frac{P(4|R)P(4)}{P(1)P(R|1) + P(2)P(R|2) + P(3)P(R|3) + P(4)P(R|4) + P(5)P(R|5)} \\ &= \frac{\left(1 - \frac{1}{4}\right) \left(\frac{1}{6}\right)}{\frac{1}{6} \left[\left(1 - \frac{2}{64}\right) + \left(1 - \frac{4}{64}\right) + \left(1 - \frac{8}{64}\right) + \left(1 - \frac{16}{64}\right) + \left(1 - \frac{32}{64}\right) + \left(1 - \frac{64}{64}\right) \right]} \\ &= \frac{\frac{3}{4}}{\frac{129}{32}} = \frac{8}{43} \Rightarrow 8 + 43 = 51 \end{aligned}$$

13. E: The mean of the Cauchy distribution is undefined as $\int_{-\infty}^{\infty} \frac{x}{a(1+x^2)} dx$ does not converge, so the answer is E: NOTA.

14. B: This is a negative binomial situation. $P(\text{Win in 7}) = \binom{7-1}{4-1} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^3 = \frac{405}{4096} \Rightarrow 4096 - 405 = 3691$

15. A: The probability of more than 3 attempts is the complement of the probability of at most 3 attempts. $P(X > 3) = 1 - P(X \leq 3) = 1 - \left(\frac{1}{5} + \left(\frac{4}{5}\right)\frac{1}{5} + \left(\frac{4}{5}\right)^2 \frac{1}{5}\right) = 1 - \frac{61}{125} = \frac{64}{125}$.

16. **B:** $t = \frac{16.5-16}{3/\sqrt{81}} = \frac{0.5}{1/3} = 1.5$. The critical value for a 1-tailed t-test with $n = 81 \Rightarrow df = 80$ is 1.664. $1.5 < 1.664$, so we fail to reject the null hypothesis, for an answer of (1.5, No).

17. **E:** Because we failed to reject the null, we lacked evidence to conclude Floridians drink enough water. However, we know they do, so a Type I error was necessarily committed. I.e., $P(\text{Type I error}) = 1$, which is none of the answer choices.

18. **D:** This is a two-proportion z-test with $\hat{p}_1 = \frac{36}{100}, \hat{p}_2 = \frac{64}{100}, n_1 = n_2 = 100$. Then, $\hat{p}_{pooled} = \frac{36+64}{100+100} = \frac{1}{2}$, so the test statistic is $z = \frac{(\frac{64}{100} - \frac{36}{100}) - 0}{\sqrt{\frac{1}{2}(1-\frac{1}{2})(\frac{1}{100} + \frac{1}{100})}} = \frac{\frac{7}{25}}{\frac{1}{2}\sqrt{\frac{1}{50}}} = \frac{14\sqrt{2}}{5}$. NOTE: A dispute for using a χ^2 -test should be denied as the alternative hypothesis is one-sided, and a χ^2 -test of independence on a 2x2 table is only equivalent to a 2-proportion Z-test when the 2-proportion Z-test has a 2-sided alternative hypothesis.

19. **D:** The Poisson distribution has mass function $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, and if X_1, X_2 are independent Poisson distributions with means λ_1, λ_2 , respectively, then $X_1 + X_2$ is Poisson distributed with mean $\lambda_1 + \lambda_2$. In our case, we have $\lambda_1 = \lambda_2 = 8$, so $P(X_1 + X_2 = 10) = \frac{16^{10} e^{-16}}{10!}$

20. **E:** There is no blocking in this design, so the answer is 0.

21. **B:** The error sum of squares is $SSE = \sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i (n_i - 1) s_i^2 = 24(4 + 4 + 9 + 1) = 432$. The treatment sum of square is $SSTR = \sum_i r(\bar{x}_i - \bar{x})^2 = 25((6 - 8)^2 + (9 - 8)^2 + (11 - 8)^2 + (6 - 8)^2) = 25(4 + 1 + 9 + 4) = 450$. $df_E = m - 1 = 4 - 1 = 3$ and $df_{Tr} = n_T - m = 100 - 4 = 96$. The F-statistic is $F = \frac{SSTR/df_{Tr}}{SSE/df_E} = \frac{450/3}{432/96} = \frac{100}{3}$.

22. **A:** The equation for the slope is $\hat{\beta}_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{40(20) - (10)(10)}{40(30) - (10)^2} = \frac{700}{1100} = \frac{7}{11}$. The y-intercept is $\bar{y} - \hat{\beta}_1 \bar{x}_1 = \frac{1}{4} - \frac{7}{11}(\frac{1}{4}) = \frac{1}{11}$. Thus, the predicted value for (3, 5, 5) is $\hat{y} = \frac{1}{11} + \frac{7}{11}(3) = 2$ with residual $5 - 2 = 3$.

23. **B:** $SST = (n - 1)S_y^2 = n\sum y_i^2 - (\sum y_i)^2 = 40\left(\frac{245}{11}\right) - 10^2 = \frac{8700}{11}$; $r^2 = \frac{(n\sum x_i y_i - (\sum x_i)(\sum y))^2}{(n\sum x_i^2 - (\sum x_i)^2) * SST} = \frac{700^2}{(1100)\left(\frac{8700}{11}\right)} = \frac{49}{87}$. Then, $r^2 = 1 - \frac{SSE}{SST} \Rightarrow SSE = \left(1 - \frac{49}{87}\right)\left(\frac{8700}{11}\right) = \frac{3800}{11}$. $s^2 = \frac{SSE}{n-2} = \frac{\frac{3800}{11}}{40-2} = \frac{100}{11}$.

$SE[\hat{\beta}_1] = \sqrt{\frac{s^2}{(n-1)S_x^2}} = \frac{10}{\sqrt{11}\sqrt{1100}} = \frac{1}{11}$, so $t = \frac{\hat{\beta}_1}{SE[\hat{\beta}_1]} = \frac{7/11}{1/11} = 7$.

Alternatively, one could test for $\rho_{x_1 y}$ using only r^2 with the formula $t = r\sqrt{\frac{n-2}{1-r^2}}$ and get $t = 7$ that way.

24. **A:** $Cov_S(X_1, X_2) = \frac{\sum((x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))}{n-1} = \frac{1}{39} (\sum x_{i1}x_{i2} - \bar{x}_1 \sum x_{i2} - \bar{x}_2 \sum x_{i1} + n\bar{x}_1\bar{x}_2) = \frac{1}{39} \left(0 - \frac{10}{40}(20) - \frac{20}{40}(10) + 40 \left(\frac{10}{40} \right) \left(\frac{20}{40} \right) \right) = -\frac{5}{39}.$

25. **C:** The β matrix is $\beta = (X^T X)^{-1} X^T Y$. $X^T X = \begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{pmatrix} = \begin{pmatrix} 40 & 10 & 20 \\ 10 & 30 & 0 \\ 20 & 0 & 20 \end{pmatrix}$. The inverse of a symmetric matrix has entries equal to the cofactor matrix divided by the original determinant, which is $(X^T X)^{-1} = \frac{1}{10000} \begin{pmatrix} 600 & -200 & -600 \\ -200 & 400 & 200 \\ -600 & 200 & 1100 \end{pmatrix} = \begin{pmatrix} 0.06 & -0.02 & -0.06 \\ -0.02 & 0.04 & 0.02 \\ -0.06 & 0.02 & 0.11 \end{pmatrix}$. $X^T Y = \begin{pmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ -10 \end{pmatrix}$. Thus, $\beta = (X^T X)^{-1} X^T Y = \begin{pmatrix} 0.06 & -0.02 & -0.06 \\ -0.02 & 0.04 & 0.02 \\ -0.06 & 0.02 & 0.11 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ -10 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.4 \\ -1.3 \end{pmatrix}$. So, the regression line has equation $\hat{y}_i = \frac{4}{5} + \frac{2}{5}x_{i1} - \frac{13}{10}x_{i2}$.

26. **B:** When performing OLS, $\hat{\beta}_1$ is NOT resistant and the sum of SQUARED residuals are minimized. This is the correct definition of a high-leverage point and homoscedasticity. Thus, 2 statements are true.

27. **A:** The density function for our gamma distribution is $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{x^3 e^{-\frac{x}{2}}}{16 \cdot 6}$. Then, $P(6 \leq X \leq 10) = \int_6^{10} \frac{1}{96} x^3 e^{-\frac{x}{2}} dx = \frac{1}{96} \left[(-2x^3 - 12x^2 - 48x - 96)e^{-\frac{x}{2}} \right]_6^{10} = 13e^{-3} - \frac{118}{3}e^{-5}$.

28. **C:** The second-largest observation has value y , then we need 4 observations to be less than y and one to be greater than y , which has probability $g_Y(y) = \binom{6}{4 \ 1 \ 1} F(y)^4 f(y) (1 - F(y))$ of occurring, where $f(y)$ is the density function and the $\binom{6}{4 \ 1 \ 1}$ multinomial coefficient comes from ordering the 6 observations. $f(y) = F'(y) = \frac{d}{dy} \ln y = \frac{1}{y}$, so $g_Y(y) = \frac{6!}{4!1!1!} (\ln y)^4 \left(\frac{1}{y}\right) (1 - \ln y) = \frac{30}{y} (\ln^4 y - \ln^5 y)$. Therefore, $E[Y] = \int_1^e \frac{30}{y} (\ln^4 y - \ln^5 y) y dy = 30 \int_1^e \ln^4 y - \ln^5 y dy = 30(53e - 144) \Rightarrow n(a + b) = 1 * (30(53 - 144)) = -2730$.

29. **D:** We are told the order of the states is locked in, tweaking, and crashing out. The transition matrix is $P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.5 & 0.25 & 0.25 \\ 0 & 0.4 & 0.6 \end{pmatrix}$. The stationary distribution must satisfy $[\pi_1, \pi_2, \pi_3]P = [\pi_1, \pi_2, \pi_3]$ with $\pi_1 + \pi_2 + \pi_3 = 1$. Expanding the matrix multiplication gives $\begin{cases} 0.6\pi_1 + 0.5\pi_2 = \pi_1 \\ 0.3\pi_1 + 0.25\pi_2 + 0.4\pi_3 = \pi_2, \text{ which,} \\ 0.1\pi_1 + 0.25\pi_2 + 0.6\pi_3 = \pi_3 \end{cases}$ upon solving, gives the vector $\left[\frac{20}{51}, \frac{16}{51}, \frac{5}{17} \right]$.

30. **D:** Each individual (independent) exponential distribution has density $f(x_i) = \lambda e^{-\lambda x_i}$. The likelihood function is $L(\lambda|x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i) = \lambda^n e^{-\lambda \sum x_i}$. Taking the natural log gives log-likelihood function $\ell(\lambda|x_1, \dots, x_n) = n \ln \lambda - \lambda \sum x_i$. Differentiating with respect to λ and setting it equal to 0 gives $\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum x_i = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum x_i}$ as the maximum likelihood estimator.