Important Instructions for this Test: Assume all questions ask for an exact answer unless otherwise specified. No calculators may be used on this test. Pay close attention to the wording of questions. As usual, NOTA stands for "None Of These Answers" are correct. It is recommended that you don't spend too long on a single question. Good Luck!

Use the following information for questions 1 and 2:
Jake is interested in determining the average number of donuts people think they could eat in one hour. He conducts a study into this by randomly choosing 5 couples and asking both members how many donuts they think they could eat. The data is shown below:

$$\{21, 15, 19, 16, 17, 18, 17, 17, 28, 21\}$$

1. What is the sum of the sample mean and sample variance for Jake's sample?

A: 4.8 B: 18 C: 22.7 D: 33 E: NOTA

2. What type of sampling method did Jake use?

A: Simple Random Sample
B: Stratified Random Sample
C: Cluster Sample
D: Convenience Sample
E: NOTA

3. Sharvaa is trying to travel from the point (0,0) to the point (5,6) along a rectangular grid consisting of only lattice points one unit at a time. If he can only move towards the finish line (ie. only right and up), how many ways can he do so while passing through either of the points (2,2) or (3,4)?

A: 462 B: 312 C: 210 D: 108 E: NOTA

4. Odin is fishing in the Atlantic Ocean. Each time he casts his line, he has a 0.4 probability of catching a fish. He will cast his line n times. If the expected number of fish he catches is 10 with a variance of 6, what is n? Naturally, assume each cast is independent of all other casts!

A: 6 B: 15 C: 9 D: 25 E: NOTA

5. Suppose X_1 and X_2 are identically distributed exponential distributions with rate parameter $\lambda = 5$ and positive correlation coefficient ρ. If the coefficient of determination, ρ^2, takes values according to a continuous uniform distribution over the interval [0,1], what is the expected value of $\frac{\mathrm{Var}(X_1+X_2)-\mathrm{Var}(X_1)}{\mathrm{Var}(2X_1)}$?

A: 1/2 B: 7/6 C: 5/12 D: 7/12 E: NOTA

6. Corbin is asking people to think of any real number between 0 and 5. The number that people think of is modeled by random variable C, which has probability density function:

$$f_C(x) = \begin{cases} 0 & x < 0 \\ \dfrac{c}{1 + \tan^{\sqrt{2}}\left(\dfrac{\pi x}{10}\right)} & 0 \le x \le 5 \\ 0 & 5 < x \end{cases}$$

What is the value of c such that this is a valid probability distribution?

A: $\frac{1}{2}$ B: $\frac{2}{\pi}$ C: $\frac{2}{5}$ D: $\frac{4}{\pi}$ E: NOTA

7. An SRS of 576 American households found an average of \$475 spent on groceries monthly. It is known that the true population standard deviation of the amount spent on groceries monthly for all American households is \$96 exactly. What is the 90% confidence interval for the true mean amount spent on groceries monthly for American households? Round the appropriate critical value to 3 decimal places.

A: 475 ± 6.580 C: 475 ± 6.768 E: NOTA
B: 475 ± 7.840 D: 475 ± 6.592

8. Let X be a Gamma distribution with shape parameter $\alpha = 4$ and rate parameter $\lambda = 1/2$. What is the moment-generating function of X, $M_X(t)$? Recall that the density function for the Gamma distribution is

$$f_X(x; \alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x \geq 0 \\ 0, \quad x < 0 \end{cases}$$

A: $\frac{1}{(2+t)^4}$ B: $\frac{2}{(2-t)^4}$ C: $\frac{1}{(1-2t)^4}$ D: $\frac{1}{(1+2t)^4}$ E: NOTA

Use the following information for questions 9 and 10:
Prateek wants to see if different parts of Florida support different universities. He takes three independent simple random samples of North, Central, and South Floridians and records which of 3 three universities they support the most. The data is summarized in the table below:

	North Florida	Central Florida	South Florida
UF	10	19	15
FSU	9	10	10
UM	3	15	30

9. Among individuals within Prateek's sample, what is the probability of choosing someone from North Florida, given that they support UM?

A: $\frac{1}{16}$ B: $\frac{2}{11}$ C: $\frac{3}{2}$ D: $\frac{48}{121}$ E: NOTA

10. Prateek then decides to run a chi-square test for homogeneity using the data he collected. What is the chi-square contribution of North Floridians that support UF? Assume all conditions for inference are met.

A: $\frac{2}{5}$ B: $\frac{1}{4}$ C: $\frac{1}{2}$ D: $\frac{1}{5}$ E: NOTA

11. Saathvik is giving out presents for Saathmas. He selected 5 different presents for his 5 best friends, but he forgot to put their names on their gifts when he was wrapping them! If he randomly gives each person one of the presents, what is the probability nobody receives the gift they were supposed to get?

A: $\frac{3}{10}$ B: $\frac{11}{30}$ C: $\frac{19}{30}$ D: $\frac{7}{10}$ E: NOTA

12. There are 6 boxes, indexed $1, 2, \dots, 6$, each of which contains 64 balls where each ball is either blue or red. The number of blue balls in box n is equal to 2^n. A box is randomly selected and a ball is randomly selected from the box. Given the ball is red, the probability that the 4th box was chosen is $\frac{m}{n}$ in simplified form. What is $m + n$?

A: 51 B: 29 C: 7 D: 71 E: NOTA

13. The standard Cauchy distribution, also known as a t-distribution with one degree of freedom, is a continuous probability distribution commonly used to describe resonance behavior. The probability density function is $f(x) = \frac{1}{a(1+x^2)}, x \in (-\infty, \infty)$ where a is a constant used to normalize the distribution. If the expected value of the Cauchy distribution is μ and the median of the Cauchy distribution is M, what is $a + \mu + M$?

A: 0 B: $\pi/2$ C: π D: $2 + \pi$ E: NOTA

14. The Florida Panthers and the Edmonton Oilers are playing each other in the Stanley Cup Finals. The Finals is a best-of-seven series, with no ties. If the Panthers have a 0.75 probability of winning each individual game, the probability they win the series in 7 games is $\frac{m}{n}$ in simplified terms. What is $n - m$? Assume each game is independent of all other games and the series will end once either team has won four games.

A: 1 B: 3691 C: 3961 D: 13549 E: NOTA

15. Yash is currently shooting free throws in Cameron Indoor Stadium. He refuses to leave until he successfully makes one free throw. He is tired, so his probability of successfully making one free throw is 0.2 on each attempt. What is the probability it takes more than 3 attempted free throws before he can leave? Assume the probability of Yash successfully making free throw is constant and each attempted free throw is independent of all previous attempts.

A: 64/125 B: 61/125 C: 16/125 D: 81/125 E: NOTA

Use the following information for the next two questions:
Larissa wants to make sure that people in Florida are drinking enough water. She takes a random sample of $n = 81$ Floridians and finds a sample mean and sample standard deviation of $\overline{W} = 16.5$ and $S_W = 3$ cups per day, respectively. The CDC recommends drinking at least 16 cups per day. Thus, Larissa will test the hypotheses $H_0: \mu_W = 16$ vs. $H_a: \mu_W > 16$ to determine whether the average Floridian drinks enough water. Assume a significance level of $\alpha = 5\%$ is used and that all conditions for inference are satisfied.

16. What is the positive test statistic of the test described above and does Larissa's sample provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis? Answers are in the form ($Test\ Statistic$, Yes or No).

A: (3, Yes) B: (1.5, No) C: (3, No) D: (1.5, Yes) E: NOTA

17. Suppose Larissa's test was designed to have a power of 0.84 against a specific alternative. If the average Floridian does, in fact, drink enough water, what is the probability Larissa made a Type I error?

A: 0 B: 0.05 C: 0.16 D: 0.84 E: NOTA

18. A simple random sample of 100 Purdue students found that 36 believe that the moon landing was faked. An independent simple random sample of 100 students from Indiana University found that 64 students believe that the moon landing was faked. Test the hypotheses $H_0: P_P = P_I \; vs. H_a: P_P < P_I$ where P_P denotes the true proportion of Purdue students who believe that the moon landing was faked and P_I denotes the true proportion of Indiana University students who believes that the moon landing was faked and report the <u>absolute value</u> of the test statistic. Assume that all conditions for inference are satisfied.

A: $\frac{35}{12}$ B: $\frac{35\sqrt{2}}{12}$ C: $\frac{14}{5}$ D: $\frac{14\sqrt{2}}{5}$ E: NOTA

19. The number of flights that are delayed from Fort Lauderdale International Airport on any given day can be modeled using a Poisson distribution with a mean of 8. What is the probability that there are a total of exactly 10 delayed flights on two independently chosen days?

A: $\frac{64^5 e^{-16}}{(5!)^2}$ B: $\frac{8^{10} e^{-8}}{10!}$ C: $\frac{16^{10} e^{-8}}{10!}$ D: $\frac{16^{10} e^{-16}}{10!}$ E: NOTA

Use the following information for questions 20 and 21:
Christina is performing an experiment to try and improve students' SAT scores. She takes a simple random sample of $n_T = 100$ students and randomly assigns $n_i = 25$ students to exactly one of $m = 4$ potential Training Programs. She then records the difference between each student's score after and before the Training Program X. The data is summarized below. Assume all conditions for inference are satisfied.

Training Program i	$i = 1$	$i = 2$	$i = 3$	$i = 4$
Sample Mean	6	9	11	6
Sample Standard Deviation	2	2	3	1

20. How many blocks are there in Christina's experiment?

A: 2 B: 4 C: 25 D: 100 E: NOTA

21. Using the data, Christina runs a one-way ANOVA to test whether training program has an effect on score difference. What is the F-statistic of the ANOVA test?

A: $\frac{25}{24}$ B: $\frac{100}{3}$ C: $\frac{104}{3}$ D: 32 E: NOTA

Use the following information for questions 22 to 25:
Soha is investigating the relationship between dependent variable Y and predictor variables X_1 and X_2 by taking a random sample of $n = 40$ observation. For the ith data point, she records the value of the dependent variable y_i and the independent variables x_{i1} and x_{i2}. Assume all conditions for linear regression and inference are satisfied. The data is summarized below:

$$\sum_{i=1}^{n} x_{i1} = 10, \sum_{i=1}^{n} x_{i2} = 20, \sum_{i=1}^{n} x_{i1}^2 = 30, \sum_{i=1}^{n} x_{i2}^2 = 20, \sum_{i=1}^{n} x_{i1}x_{i2} = 0,$$
$$\sum_{i=1}^{n} y_i = 10, \sum_{i=1}^{n} y_i^2 = \frac{245}{11}, \sum_{i=1}^{n} y_i x_{i1} = 20, \sum_{i=1}^{n} y_i x_{i2} = -10$$

22. Soha first develops a one-variable least-squares regression line: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$. What is the residual for an observation with $x_1 = 3, x_2 = 5, y = 5$?

A: 3 B: -2 C: 2 D: -3 E: NOTA

23. Soha then decides to perform a linear regression t-test to test the null hypothesis $\beta_1 = 0$ against the alternative hypothesis of $\beta_1 \neq 0$. What is the absolute value of the test statistic?

A: 77 B: 7 C: 1 D: $\frac{1}{7}$ E: NOTA

24. Now, compute the sample covariance between the two the predictor variables, X_1 and X_2.

A: $-\frac{5}{39}$ B: $-\frac{1}{8}$ C: $-\frac{79}{312}$ D: $\frac{1}{8}$ E: NOTA

25. Finally, Soha performs multiple least-squares linear regression to form a regression line of the form $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$. What is the equation of the line?

A: $\hat{y}_i = \frac{233}{22} + \frac{7}{11}x_{i1} - \frac{3}{2}x_{i2}$ C: $\hat{y}_i = \frac{4}{5} - \frac{13}{10}x_{i1} + \frac{2}{5}x_{i2}$ E: NOTA

B: $\hat{y}_i = \frac{885}{88} - \frac{3}{2}x_{i1} + \frac{7}{11}x_{i2}$ D: $\hat{y}_i = \frac{4}{5} + \frac{2}{5}x_{i1} - \frac{13}{10}x_{i2}$

26. How many of the following statements are true regarding least-squares linear regression with a single predictor variable?

I. The slope of the regression line is resistant.
II. The least-squares regression line minimizes the sum of residuals.
III. A point with "high leverage" has an extreme predictor value.
IV. Homoscedasticity means that the standard deviation of the response variable is constant for all predictor values.

A: 1 B: 2 C: 3 D: 4 E: NOTA

27. You are the fifth person in the lunch line, and each person on average takes 2 minutes to get served lunch. The amount of time until you start getting served lunch can be modeled using a Gamma distribution with $\alpha = 4$ and $\lambda = 1/2$. What is the probability it takes between 6 and 10 minutes to get to start getting served lunch?

A: $\frac{118}{3}e^{-5} - 13e^{-3}$
B: $13e^{-3} - \frac{118}{3}e^{-5}$
C: $\frac{9}{4}e^{-3} - \frac{125}{12}e^{-5}$
D: $\frac{125}{12}e^{-5} - \frac{9}{4}e^{-3}$
E: NOTA

28. Uncle Ben decides to sample 6 independent observations from a probability distribution whose cumulative density function is $F(x) = \begin{cases} 0, & x < 1 \\ \ln x, & 1 \le x \le e \\ 1, & x > e \end{cases}$. If the expected value of the second-largest observation in his sample is of the form $ae^n + b$ for rational numbers a and b and positive integer n, what is $n * (a + b)$?

A: -135 B: $-\frac{9}{2}$ C: -2730 D: -91 E: NOTA

29. Jake's mental state can be modeled as a stochastic process, moving between three states. The first is locked in, the second is tweaking, and the third is crashing out. If he is currently locked in, he has a probability of 0.6 of staying locked in, 0.3 of tweaking, and 0.1 of crashing out. If he is currently tweaking, he has a probability of 0.5 of becoming locked in, 0.25 of crashing out, and 0.25 of staying in the tweaking state. If he is crashing out, he has a probability of 0.6 of continuing to crash out and 0.4 of moving to tweaking. What is the stationary distribution of Jake's mental state? Answers are in the form [Locked in, Tweaking, Crashing out].

A: $\left[\frac{16}{51}, \frac{20}{51}, \frac{5}{17}\right]$ B: $\left[\frac{5}{17}, \frac{16}{51}, \frac{20}{51}\right]$ C: $\left[\frac{16}{51}, \frac{5}{17}, \frac{20}{51}\right]$ D: $\left[\frac{20}{51}, \frac{16}{51}, \frac{5}{17}\right]$ E: NOTA

30. What is the maximum likelihood estimator for the rate parameter, λ, of an exponential distribution?

A: $\hat{\lambda} = \frac{1}{\sum x_i}$ B: $\hat{\lambda} = \frac{\sum x_i}{n}$ C: $\hat{\lambda} = \sum x_i$ D: $\hat{\lambda} = \frac{n}{\sum x_i}$ E: NOTA