For each question, mark the best answer on your answer sheet.
For all questions, NOTA means "None of these answers."

For questions 1 - 2: Ms. Woolfenden's 12 students took their exam. The average score was 80 with a standard deviation of 5.

1.   If Nerissa earned a 72 and Dan earned a 78, what is the average score of the remaining students?

   A) 80   B) 81   C) 82   D) 88   E) NOTA

2.   Ms. Woolfenden realizes that one of the questions was not valid. She decides to add 5 points to each of the 12 exam scores. What is the new average and standard deviation?

   A) 80 & 5   B) 80 & 10   C) 85 & 5   D) 85 & 10   E) NOTA

3.   If it is known that the median of a distribution is much less than the mean of the distribution, what can be said of the shape of the distribution?

   A) It is skewed right.   B) It is skewed left.   C) It is roughly symmetric.
   D) It is uniform.   E) NOTA

4.   When comparing the t-distribution to the standard normal distribution, which of the following is **NOT** true?

   A)   Both distributions are symmetric about zero.
   B)   Both distributions are bell-shaped.
   C)   As the degrees of freedom increase, the t distribution approaches the standard normal curve ever more closely.
   D)   t distributions have less probability in the tails.
   E)   NOTA
   F)

5.   If A and B are independent events with P(A) = 0.2 and P(B) = 0.3, find the probability that A or B occur.

   A) 0.06   B) 0.44   C) 0.50   D) 0.56   E) NOTA

6.   Which of the following will **NOT** decrease the probability of a Type II error?

   A)   Increasing the level of significance.
   B)   Increasing the sample size.
   C)   Increasing the standard deviation.
   D)   Increasing the difference between the null value and the alternative value.
   E)   NOTA

For questions 7 - 9: A pizza company claims that their mean delivery time is 30 minutes with a standard deviation of 6 minutes. Believing the average delivery takes longer, a consumer advocate conducts a random sample of 9 deliveries that yields a mean delivery time of 32.8 minutes per delivery. Using a significance level of 5%, the advocate tests the company's claim.

7.   What are the appropriate hypotheses for this test?

A) H_o: $\mu = 30$, H_a: $\mu > 32.8$ B) H_o: $\mu = 32.8$, H_a: $\mu < 32.8$

C) H_o: $\mu = 32.8$, H_a: $\mu > 32.8$ D) H_o: $\mu = 30$, H_a: $\mu > 30$ E) NOTA

8.   What is the probability of a Type I error?

A) 0.028   B) 0.03   C) 0.05   D) 0.95   E) NOTA

9.   If the P-value is .08, we can conclude that

A)   There is sufficient evidence that the company's mean delivery time is more than 30 minutes.

B)   There is sufficient evidence that the company's claim is true.

C)   There is insufficient evidence that the company's mean delivery time is more than 30 minutes.

D)   There is insufficient evidence that the company's claim is true.

E)   NOTA

10.   Which of the following can establish a causal link between two variables?

A) A controlled experiment   B) A sample survey

C) An observational study   D) A census   E) NOTA

11.   Out of 100 students in a school, 20 do not take Calculus or Statistics and 20 take both Calculus and Statistics. The total number of students that take Calculus is equal to the total number of students that take Statistics. How many students take Calculus?

A) 20   B) 30   C) 50   D) 60   E) NOTA

12.   During the previous election, exit polls were conducted by selecting every fifth person that walked out of the various voting sites, handing them a survey asking how they voted, and letting them place it anonymously in a box. This is an example of what type of sampling?

A) Simple Random Sampling   B) Systematic Sampling

C) Stratified Sampling   D) Cluster Sampling   E) NOTA

13. Which of the following is **NOT** true for the chi-square distributions?

    A) Total area under the curve is equal to 1.
    B) Bell-shaped and symmetric about zero.
    C) As the number of degrees increase, the curve looks more and more like a normal curve.
    D) After it peaks, the curve approaches the x-axis asymptotically.
    E) NOTA

14. Which of the following is **NOT** true for a geometric distribution?

    A) There is a fixed number of observations.
    B) Each observation has only two possible outcomes, "success" or "failure."
    C) The probability of success is the same for each observation.
    D) Our interest is in the number of trials until the first success.
    E) NOTA

For questions 15 - 17: You are given the regression equation
        Temp = 30 - .5(*Dist*) where *Temp* is the temperature on a sensor in °C and *Dist* is
        the distance in centimeters from the sensor to a heat source.

15. Which of the following is **NOT** a reasonable conclusion?

    A) The predicted temperature of the heat source is 30°C.
    B) The predicted temperature decreases approximately .5°C for each centimeter the sensor is moved away from the heat source.
    C) We can predict that the sensor displays a temperature of 25°C when the sensor is 10 centimeters away from the heat source.
    D) There is a positive association between temperature and distance.
    E) NOTA

16. Which of the following **MUST** be a nonnegative?

    A) The slope of the regression line.
    B) The correlation coefficient.
    C) The coefficient of determination.
    D) The residual when the sensor is 10 centimeters from the heat source.
    E) NOTA

17. If the residual when the sensor is a distance of 8 centimeters from the heat source is –3, find the temperature reading on the sensor?

    A) 22 °C   B) 23°C   C) 26°C   D) 29°C   E) NOTA

18. While examining a set of bivariate data, you suspect the data will fit a power law model. Which of the following transformations will yield a linear model?

A) log y vs. x   B) x vs. log y
C) log x vs. y   D) log y vs. log x   E) NOTA

19. The 95% confidence interval for the mean of a population is (-12, 14). Which of the following is a reasonable interpretation of this interval?

A)   There is a 95% probability that the population mean is between -12 and 14.
B)   There is a 95% probability that the true mean is 1 and that the true margin of error is 13.
C)   If we took many, many additional random samples, approximately 95% of the means would be between -12 and 14.
D)   If we took many, many additional random samples and from each computed a 95% confidence interval, approximately 95% of them would contain the true population mean.
E)   NOTA

For questions 20 & 21: You pay $1 to play a game in which a fair die is tossed. If the number of spots showing is either 4 or 5 you win $1, if the number of spots showing is 6 you win $4, and if the number of spots showing is 1, 2, or 3 you win nothing.

20. What is the probability that you win more than the cost of playing the game?

A) 0   B) 1/6   C) 1/3   D) 2/3   E) NOTA

21. Including the cost of the game, what is your expected value of this game?

A) -$0.50   B) $0   C) $0.50   D) $1   E) NOTA

22. A stratified random sample corresponds to which of the following experimental designs?

A)   A block design
B)   A double-blind experiment
C)   An experiment with a placebo
D)   A confounded, nonrandomized study
E)   NOTA

23. If A and B are mutually exclusive events with P(A) = 0.4 and P(B) = 0.3, find the probability that A and B both occur at the same time.

A) 0.00   B) 0.10   C) 0.12   D) 0.70   E) NOTA

24. A researcher is working on a new treatment for a disease. The average survival time after standard treatment is two years. In a trial on three subjects using the new treatment, the average survival time is four years. Although the survival time has doubled, the results are not statistically significant at any reasonable level because

   A) the placebo effect is present, which limits the statistical significance.
   B) the sample size is too small.
   C) although the survival time has doubled, it's still only two years.
   D) the sample size for the standard treatment is not known.
   E) NOTA

25. Which of the following is a valid probability density function on the interval [0, 2]?

   A) $f(x) = 0.5x$
   B) $f(x) = x$
   C) $f(x) = 2x$
   D) $f(x) = 3 - 1.5x$
   E) NOTA

26. In determining the ratings of TV shows, a company gathers data on many variables. Which of the following is categorical?

   A) Age of viewers in years
   B) Family income of viewers
   C) Gender of viewers
   D) Number of TVs in the household
   E) NOTA

27. Which of the following is **NOT** possible?

   A) The standard deviation is greater than the mean.
   B) The 5-number summary has 3 identical values.
   C) The mean is negative and the standard deviation is positive.
   D) The mean is positive and the standard deviation is negative.
   E) NOTA

28. Each time Mower Depot receives a shipment of 50 mowers, one of them is defective. How many mowers would we expect Joe's Lawn Service to buy before it comes across a defective mower?

   A) 1   B) 2   C) 25   D) 50   E) NOTA

For questions 29 & 30: The following table displays the results of a sample of 100 in which the subjects indicted their favorite sport and age group.

Age	Football	Baseball	Soccer
Over 40	15	8	7
Between 20 and 40	20	15	15
Under 20	8	4	8

29. When performing a Chi-Square Test for Independence, what degrees of freedom should be used?

   A) 1   B) 3   C) 4   D) 6   E) NOTA

30. When performing a Chi-Square Test for Independence, what is the expected number of Under 20s who choose Soccer as their favorite sport?

   A) 6   B) 8   C) 20   D) 30   E) NOTA

Solutions

1. B   12 X 80 = 960. 960 – 72 – 78 = 810. 810/10 = 81.

2. C   Mean will increase by 5 but standard deviation is not changed.

3. A   The mean is being influenced by one or a few very large values.

4. D   The t distribution has MORE probability in the tails.

5. B   P(A or B) = P(A) + P(B) – P(A and B) = .2 + .3 – (.2)(.3) = .5 – .06 = .44

6. C   Increasing standard deviation will increase the P(Type II error).

7. D   Hypotheses use the company's claim of 30 minutes as the null value.

8. C   P(Type I error) = level of significance (alpha) = 5% = .05

9. C   The pvalue > alpha meaning there is insufficient evidence for the H_a.

10. A   Controlled experiment is the only way to show causation.

11. C   100 – 20 (neither) – 20 (both) = 60 who take one or the other. Split evenly, that means that 30 take Calculus only. Add to the 20 also taking Statistics, and 50 total students are taking Calculus.

12. B   Choosing every 5th person is Systematic.

13. B   The chi-square distribution starts at zero on the left, taking on only positive values.

14. A   There is not a fixed number of observations. The number of observations depends on the occurrence of the first success.

15. D   The temperature decreases as the distance increases, indicating a negative association.

16. C   The coefficient of determination is R^2 and must be positive.

17. B The predicted value is 30 - .5(8) = 26. The residual –3 = y – 26.
    Therefore, y = -3 + 26 = 23.

18. D You must take the log of both variables to transform a power model.

19. D The others refer specifically to the sample mean. Only D refers to
    our confidence in the method of constructing the interval.

20. B Only one out of the 6 possible outcomes will win more than the $1
    needed to play the game.

21. B Considering the cost of the game, 3/6(-$1) + 2/6($0) + 1/6($3) = $0.

22. A Both stratified and block require the division of subjects based on a
    common characteristic.

23. A By definition, if two events are mutually exclusive, they cannot occur
    at the same time.

24. B A sample size of 3 is too small – one of the survivors could have been
    unusually large.

25. A The triangular area created will have a base of 2 and height of 1.
    Area = ½ bh = 1. All others have an area under the curve other than 1.

26. C Gender is the only variable that is not quantitative.

27. D The standard deviation can never be negative.

28. D This is a geometric distribution with p = 1/50. The mean or expected
    number of trials is 1/p = 50.

29. C There are 3 row variables and 3 column variables. (r – 1)(c – 1) = 4.

30. A (Row total x Column total)/Table total = (20 x 30)/100 = 6

Middleton Invitational - February 2007   Statistics Team Question #1

Given the following set of data,   16   8   20   -5   17   7

A = the mean, B = the median, C = the first quartile, and D = the third quartile.

Middleton Invitational - February 2007   Statistics Team Question #2

Suppose events X and Y are independent, the probability of X is 0.5, and the probability of Y is 0.2. Let A = the probability that both X and Y occur, B = the probability that either X or Y occur, and C = the probability that X occurs given Y.

Middleton Invitational - February 2007   Statistics Team Question #3

Given the set of ordered pairs below and its linear model in $y = a + bx$ form.

(0, 14) (1, 11) (2, 8) (3, 5) (4, 2)

Let $A = a$, $B = b$, C = the correlation coefficient, and D = coefficient of determination.

Middleton Invitational - February 2007   Statistics Team Question #4

A coin is flipped until a head appears. Let A = the average number of flips until heads appears, B = the probability that it takes more than 3 flips until a head appears, and C = the probability that the third flip is the first time a head appears.

Middleton Invitational - February 2007   Statistics Team Question #5

Assume the scores on the 2002 Mathematics SAT are normally distributed with a mean of 550 and a standard deviation of 110. Using the Empirical Rule, let A = the score of a student at the 97.5^{th} percentile, B = the score of a student at the 16th percentile, and C = the score of a student at the 50th percentile.

Middleton Invitational - February 2007   Statistics Team Question #6

The table summarizes the political affiliations of 40 adults.

	Female	Male
Democrat	6	8
Republican	4	8
Independent	10	4

Suppose one of these adults is randomly selected. Let A = the probability that the adult is a Republican, B = the probability that the adult is female, and C = the probability that the adult is a female Republican.
Now suppose one of these female adults is randomly selected. Let D = the probability that she is a Republican.

Middleton Invitational - February 2007   Statistics Team Question #7

Random variable X has a mean of 10 and a standard deviation of 3. Random variable Y has a mean of 20 and a standard deviation of 4. The correlation between X and Y is -1.

Middleton Invitational - February 2007   Statistics Team Question #8

A bivariate set of data gives a linear regression equation of $\hat{y} = 15 + 40x$. The point $(3.5, y)$ has a residual of -60.   Find y.

Middleton Invitational - February 2007   Statistics Team Question #9

An experiment using 60 college students was conducted to explore the nature of the relationship between a person's heart rate (measured in beats per minute) and the frequency at which that person stepped up and down on steps of various heights. There were three rates of stepping and two different step heights used. Each subject was randomly assigned to perform the activity at one of the three stepping rates and at one of the two possible heights for three minutes. Heart rates were measured at the end of the three-minute period.

Let A = the number of experimental units, B = the number of factors, and C = the number of treatment groups needed.

Middleton Invitational - February 2007   Statistics Team Question #10

The power for a hypothesis test run at a 5% significance level was 0.85. Let A = the probability of a Type I error and B = the probability of a Type II error.

Middleton Invitational - February 2007   Statistics Team Question #11

Kim is running for class president at her very large high school. Brian, the manager of her campaign, wants to know Kim's chance of winning. He randomly surveys 100 students and finds that 80 will vote for Kim.
Using a critical value of 2, find the confidence interval for the percent of students that will vote for Kim.

Middleton Invitational - February 2007   Statistics Team Question #12

A die is rolled 240 times with the following results.

Face	1	2	3	4	5	6
Frequency	50	60	40	40	20	30

A Goodness of Fit test is conducted to assess the fairness of the die. Let A = the degrees of freedom for this test, B = the expected number of 2's, and C = the X^2 component for 5's.

Middleton Invitational - February 2007   Statistics Team Question #13

A vaccine for use against a dangerous virus has been developed. You have 8 available rats (named below), which you will expose to the virus. You decide to use 4 rats as a control group by not giving them the vaccine. You will treat the other 4 rats with the vaccine.

Start at the beginning of the list of random digits.

Random digits –   81868 71035 09001 43367 49497 54580 81507 27102

Use the labels attached to the rats.

Rats - 1. Andrew   2. Beth   3. Chi Chi   4. David
        5. Ernie   6. Floyd   7. Grant   8. Hannah

Write the names of the rats assigned to the treatment group in the order in which they were selected.

Middleton Invitational - February 2007   Statistics Team Question #14

A discrete random variable X takes four values and has the probability distribution shown below where A and B are constants. The expected value of X is 1.4.

X	0	1	2	3
$P(X)$	0.4	A	B	$A + B$

Middleton Invitational - February 2007   Statistics Team Question #15

The following are statistics from a set of bivariate data:

$$\bar{x} = 90, s_x = 5, \bar{y} = 42, s_y = 15, r = -0.2$$

Let A = the slope of the least squares regression line and B = the y-intercept of the least squares regression line.

Solutions Statistics Team

1. A = the mean = 10.5, B = the median = 12, C = the first quartile = 7, and D = the third quartile = 17.

2. $A = P(X \text{ and } Y) = 0.5 \times 0.2 = 0.1$, $B = P(X \text{ or } Y) = 0.5 + 0.2 - 0.1 = 0.6$, and because the events are independent, the condition of Y has no effect on $P(X)$ so $C = 0.5$. A = the y-intercept = 14, B = slope = -3, C = perfect negative correlation, $r = -1$, $D = r^2 = 1$.

3. $p = \frac{1}{2}$, $A = \frac{1}{p} = 2$, $B = (1-p)^n = (1 - \frac{1}{2})^3 = \frac{1}{8}$, and $C = (1-p)^{n-1}p = (\frac{1}{2})^2 \frac{1}{2} = \frac{1}{8}$.

4. $A = 550 + 2(110) = 770$, $B = 550 - 1(110) = 440$, and $C = 550$.

5. $A = \frac{12}{40}$, $B = \frac{20}{40}$, $C = \frac{4}{40}$, and $D = \frac{4}{20}$.

6. $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y} = \sqrt{3^2 + 4^2 + 2(-1)(3)(4)} = 1$, **Answer is 1.**

7. Residual $= y - \hat{y}$. Therefore, y = residual $+ \hat{y} = -60 + [15 + 40(3.5)] = -60 + 155 = 95$. **Answer is 95.**

8. A = the college students = 60, B = step rate and step height = 2, and C = 3 rates x 2 heights = 6.

9. $A = .05$ and $B = 1 - .85 = .15$.

10. $CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = .8 \pm 2\sqrt{\frac{(.8)(.2)}{100}} = .8 \pm 2\sqrt{\frac{.16}{100}} = .8 \pm 2\left(\frac{.4}{10}\right) = .8 \pm .08 = .72 - .88$. **Answer is 72% to 88%.**

11. A = degrees of freedom = number of categories $- 1 = 6 - 1 = 5$, $B = \frac{1}{6}(240) = 40$, and $C = \frac{(O-E)^2}{E} = \frac{(20-40)^2}{40} = \frac{400}{40} = 10$.

12. **Answer is Hannah, Andrew, Floyd & Grant.**

13. Solving the system of equations, $0(0.4) + 1A + 2B + 3(A + B) = 1.4$ and $0.4 + A + B + A + B = 1$, we find that $A = 0.1$ and $B = 0.2$.

15. $A = r\frac{s_y}{s_x} = -0.2\left(\frac{15}{5}\right) = -0.6$ and $B = \bar{y} - b\bar{x} = 42 - (-0.6)(90) = 96$.