TMTA Statistics Exam

1. The scoring formula for this exam is 4R – W + 40, where R is the number of correct answers, W the number of wrong answers. If a student completely guesses on each of the forty questions on this exam, then their expected score is
   a. 0
   b. 8
   c. 16
   d. 40
   e. 100

2. An air quality specialist wants to assess the overall threat of vehicle pollution on the air quality of the Great Smoky Mountains. As part of her report, she gives the "average" level of ozone emissions from vehicles entering the National Park. To assess the overall threat of emissions, she should use the
   a. Mean emissions.
   b. Median emissions.
   c. Mode of emissions.
   d. Maximum of emissions
   e. Standard deviation of emissions.

3. Which of the following will *not* change the standard deviation?
   a. Adding 5 to every value in a data set.
   b. Multiplying every value in a data set by 5.
   c. Dividing every value in a data set by 5.
   d. Squaring each value in a data set.
   e. Taking the logarithm of each value in a data set.

4. The distribution of grade point averages for freshmen at a college is clearly skewed left. When thinking of the "average" GPA for freshmen, which of the following is necessarily true?
   a. Mean GPA > Median GPA.
   b. Mean GPA < Median GPA.
   c. The Mean and Median are essentially equal.
   d. 68% of students will be within a standard deviation of the Mean GPA.
   e. More than one of the above.

5. Suppose that Gallup finds that 44% of American adults say they are baseball fans with a margin of error of 4% at the 95% confidence level. Which of the following accurately states a valid inference from these results?
   a. A minority of American adults are baseball fans.
   b. A majority of American adults are baseball fans.
   c. Exactly 44% of American adults are baseball fans.
   d. Between 42% and 46% of American adults are baseball fans.
   e. Exactly 95% of American adults are baseball fans.

6. The following table gives the expected incidence and mammography results for breast cancer in women ages 25-64.

	Breast Cancer	No Breast Cancer
Mammography Positive	309	12,256
Mammography Negative	51	87,384

Given that a randomly selected woman has a positive mammography, what is the probability that she has breast cancer?
   a. .00309
   b. .02459
   c. .02521
   d. .85833
   e. 1.0000

7. The least squares regression line is often referred to as the "line of best fit." It is best in the sense that it
   a. Goes through the most points.
   b. Minimizes the sum of the squared vertical distances between the points and the line.
   c. Minimizes the sum of the squared horizontal distances between the points and the line.
   d. Minimizes the sum of the vertical distances between the points and the line.
   e. Minimizes the sum of the horizontal distances between the points and the line.

8. The regression between child mortality and literacy rates for 111 countries yields a correlation of - 0.76. This implies that
   a. No relationship exists between child mortality and literacy rates.
   b. If literacy can be increased in countries with low literacy rates, then child mortality will decrease.
   c. If literacy can be increased in countries with high child mortality rates, then child mortality will decrease.
   d. Countries with higher literacy rates tend to have lower child mortality rates.
   e. More than one of the above.

9. Donut holes from a local donut shop have weights that are approximately normally distributed with a mean of 6.2 grams and a standard deviation of 0.7 grams. We expect that almost all donuts (95%) will weigh
   a. At least 6.2 grams.
   b. At least 5.5 grams.
   c. At least 4.8 grams.
   d. Between 5.5 and 6.9 grams.
   e. Between 4.8 and 7.6 grams.

10. Length of human pregnancies is approximately normally distributed with a mean of 266 days and a standard deviation of 16 days. The 25th percentile of pregnancy lengths is about
   a. 235 days.
   b. 255 days.
   c. 262 days.
   d. 270 days.
   e. 277 days.

11. In a study of a potential drug for treating patients with migraines, volunteers are randomly assigned to take either the new drug or placebo. If the new drug significantly (p-value<.05) reduces migraines compared to placebo, then which of the following is a valid inference?
    a. The study is not useful because it used volunteers.
    b. The study shows a cause-and-effect relationship exists between the new drug and migraines for the participants in the study.
    c. The study shows a cause-and-effect relationship exists and the results can be generalized to all migraine sufferers.
    d. A claim of causation is not possible, though the association between the new drug and migraines can be generalized to all migraine sufferers.
    e. More than one of the above.

12. Conducting a hypothesis test at the 5% significance level is equivalent to saying
    a. The risk of a Type I Error is 5%.
    b. The risk of a Type II Error is 5%.
    c. Power is 5%.
    d. Power is 95%.
    e. More than one of the above.

13. A statistician wishes to conduct a hypothesis test for a population mean using a random sample of 20 observations. The population standard deviation is not known. Which of the following statements is correct?
    a. A t-test may be used regardless of the distribution of the population.
    b. A t-test may be used provided that the population is unimodal, not strongly skewed and does not contain outliers.
    c. The statistician does not have enough information to perform a t-test.
    d. The statistician should perform a z-test rather than a t-test.
    e. The statistician should perform a χ^2 test rather than a t-test.

14. When planning a hypothesis test, the power of the test is
    a. The probability that we reject a true null hypothesis, and power increases as sample size increases.
    b. The probability that we reject a false null hypothesis, and power increases as sample size increases.
    c. The probability that we accept a true null hypothesis, and power increases as sample size increases.
    d. The probability that we reject a true null hypothesis, and power decreases as sample size increases.
    e. The probability that we reject a false null hypothesis, and power decreases as sample size increases.

15. A lottery has the following prizes and corresponding probabilities:

Prize	$1	$10	$100	$1000
Probability	.25	.05	.005	.0001

If each ticket is sold for $2, then the expected profit per ticket for the organization is
    a. $0.50
    b. $0.65
    c. $1.00
    d. $1.35
    e. $2.00

16. A regression of college GPA vs. high school GPA yielded a correlation of .35. This correlation can be interpreted as
    a. We expect college GPA to be within .35 of high school GPA.
    b. We expect college GPA to increase by .35 for each increase of 1 in high school GPA.
    c. About 35% of the variation in college GPA can be explained by the linear relationship with high school GPA.
    d. About 12% of the variation in college GPA can be explained by the linear relationship with high school GPA.
    e. About 59% of the variation in college GPA can be explained by the linear relationship with high school GPA.

17. In the regression line $y = .23x - .45$ we conclude that
    a. Y is expected to increase by .23 as x increases by 1.
    b. Y is expected to decrease by .45 as x increases by 1.
    c. Y is expected to increase by 1 as x increases by .23.
    d. Y is expected to decrease by 1 as x increases by .45.
    e. The correlation between x and y is .23.

18. A student wondered whether songs by The Beatles were, on average, shorter than 3 minutes. If the student were to conduct a hypothesis test, the alternative hypothesis would be
    a. $\bar{x} = 3$
    b. $\bar{x} < 3$
    c. $\mu = 3$
    d. $\mu < 3$
    e. $\sigma = 3$

19. A student wondered whether songs by The Beatles were, on average, shorter than 3 minutes. In a study of songs by The Beatles the mean length for a random sample of 40 songs was found to be 2.7 minutes with a standard deviation of 0.8 minutes. The correct inference for these results would be to conclude that
    a. The mean length of Beatles' songs is less than 3 minutes with a p-value between .01 and .05.
    b. The mean length of Beatles' songs in not less than 3 minutes with a p-value between .01 and .05.
    c. The mean length of Beatles' songs is less than 3 minutes with a p-value greater than .05.
    d. The mean length of Beatles' songs in not less than 3 minutes with a p-value greater than .05.
    e. The mean length of Beatles' songs is equal to 3 minutes with a p-value greater than .05.

20. A biology student collected data on how far female turtles moved in a given time period and found for a sample of 6 turtles a mean of 10 meters with a standard deviation of 4.6 meters. Consider the following conditions.

    I   The sample is a simple random sample from the population of turtles.
    II   The population distribution of distance moved is normal.
   III   The sample standard deviation, s, is equal to the population standard deviation, σ.

Which of the following must be true to use the t-confidence interval formula?
    a. I only
    b. II only
    c. I and II
    d. II and III
    e. I, II, and III

21. A biology student collected data on how far female turtles moved in a given time period and found for a sample of 6 turtles a mean of 10 meters with a standard deviation of 4.6 meters. The appropriate 95% confidence interval for this problem is
    a. (6.32, 13.68)
    b. (5.40, 14.60)
    c. (5.17, 14.83)
    d. (6.22, 13.78)
    e. (6.35, 13.65)

22. The correct interpretation of a 95% confidence interval for the mean distance female turtles move in an hour is
    a. 95% of the time a turtle will move a distance within this interval during the given time period.
    b. 95% of the time the population mean distance moved will fall in this interval.
    c. We can be 95% confident that the population mean distance moved is contained within this interval.
    d. We can be 95% confident that the sample mean is contained within this interval.
    e. 95% of the sample observations fall within this interval.

23. For a sample standard deviation of 4.6, to reduce the standard error of the sample mean to 1 would require a sample size of about
    a. 140
    b. 22
    c. 82
    d. Any sample size—the standard error isn't dependent on n.
    e. Not enough information is given.

24. A student has several routes that s/he may take from home to school. To avoid being late it is decided to take the route that is most consistent in terms of time. The most relevant statistic is the
    a. Mean time.
    b. Median time.
    c. Upper quartile of time.
    d. Lower quartile of time.
    e. Standard deviation of time.

25. Three students are selected from a class of 25 students to fill the three leadership roles of a council: president, vice president, and secretary. How many ways can the three roles be assigned if students are selected randomly?
    a. 72
    b. 75
    c. 2300
    d. 13800
    e. 15625

26. A college student takes 15 total credit hours in a given semester. In her 4-hour class she makes an A which equates to 4.00 grade points. In her 2-hour class she makes a C which equates to 2.00 grade points. In her three 3-hour classes she makes an A, a B (3.00 grade points), and a C. Her cumulative grade point average for the semester is
    a. 3.25
    b. 3.13
    c. 3.00
    d. 2.87
    e. 2.75

27. The Gallup organization was interested in determining the percentage of American adults who said they were baseball fans. If Gallup wanted a margin of error of ± 4% for a 90% confidence interval, how many adults would need to be surveyed?

    a. 423
    b. 1692
    c. 601
    d. 2401
    e. Not enough information is given

28. A county water authority conducts a random survey of households to determine if there has been a change in the proportion of homes with high lead levels in drinking water from the previous year. A Type II Error would involve
    a. Finding that the proportion has changed when in truth it hasn't.
    b. Finding that the proportion has not changed when in truth it has.
    c. Laboratory errors in the measurement of lead levels.
    d. More than one of the above.
    e. None of the above.

29. Past studies have found that in phone surveys about 60% of households are eventually reached. A study looked at whether leaving a message on an answering machine would lead to an increase in the percentage of households eventually reached. If 200 out of 291 randomly sampled households where a message is left are eventually reached, what is an appropriate conclusion about the percentage of homes that can be reached when a message is left?
    a. There is sufficient evidence at the 95% significance level to say that the population percentage is greater than 68%.
    b. There is sufficient evidence at the 95% significance level to say that the population percentage is not greater than 68%.
    c. There is sufficient evidence at the 95% significance level to say that the population percentage is not greater than 60%.
    d. There is not sufficient evidence at the 95% significance level to say that the population percentage is greater than 60%.
    e. There is sufficient evidence at the 95% significance level to say that the population percentage is greater than 60%.

30. If 3 independent research studies are to be conducted to compare a new drug with a placebo, with each study leading to a 95% confidence interval for the true difference, the probability that *at least* one of the confidence intervals does not contain the true difference is
    a. .05
    b. .000125
    c. .15
    d. .142625
    e. .857375

31. The scatterplot given below shows the average MPG for a new fleet of vehicles vs. the percent of the fleet that is classified as trucks (pickups and SUVs) for the years 1984-2001.

Scatterplot of MPG vs %trucks
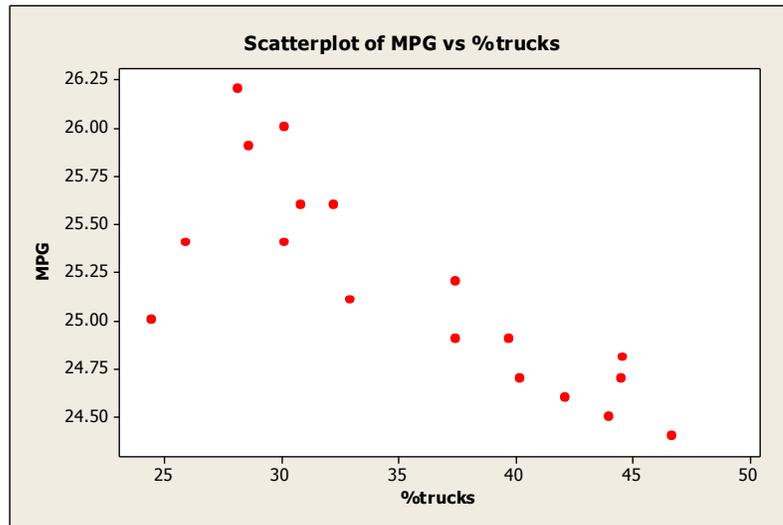
The correlation for this plot is approximately
    a.   -.80
    b.   -.40
    c.   0
    d.   .40
    e.   .80

32. The term *regression analysis* was first used by Francis Galton in the late 1800's when he looked at the relationship between father's height (x) and son's height (y). Galton's research led to the concept of regression to the mean. Regression to the mean would imply that the son of a father who was at the 90th percentile for height would be expected to
    a.   Also be at the 90th percentile for height.
    b.   Be in a higher percentile for height than his father.
    c.   Be above average in height, but not as tall as his father.
    d.   Be at the 50th percentile in height.
    e.   Be below average in height.

33. A coin is tossed 100 times and comes up heads 57 times. Which of the following is most accurate?
    a.   57 heads is the expected value of the experiment.
    b.   Getting 57 heads implies that, for the next 100 tosses, we should expect only 43 heads.
    c.   There is significant evidence at the .01 level that the coin is unfair.
    d.   There is not significant evidence at the .01 level that the coin is unfair.
    e.   More than one of the above.

34. A coin is tossed 400 times and comes up heads 228 times. Which of the following is most accurate?
    a.   228 heads is the expected value of the experiment.
    b.   Getting 228 heads implies that, for the next 400 tosses, we should expect only 172 heads.
    c.   There is significant evidence at the .01 level that the coin is unfair.
    d.   There is not significant evidence at the .01 level that the coin is unfair.
    e.   More than one of the above.

35. Billy makes 60% of his free throws in basketball, and he decides to shoot until he makes his next shot. Assuming that each shot is independent, the probability that it takes him more than three shots to make his next basket is
    a. .064
    b. .216
    c. .4
    d. .784
    e. .936

36. A malaria vaccine protects 1/4 of the people that receive it. In a random sample of 5, what is the probability that more than 3 are protected from malaria?
    a. 0
    b. .003906
    c. .014648
    d. .015625
    e. .632813

37. If ACT scores are approximately normal with a mean of 20 for graduating high school students in Tennessee, and only 20% of such students score 25 or higher, then we can also say that only 10% of students score at approximately what level or higher?
    a. 26
    b. 28
    c. 30
    d. 34
    e. More information is needed.

38. If World Cup goals follow a Poisson process with a mean of 2.7 goals per first-round match, find the probability that at least two goals will be scored in a first-round match.
    a. .75134
    b. .24866
    c. .80085
    d. .19915
    e. .40601

39. If six means are to be compared, for example mean responses for six treatment conditions, pairs of means could be compared using t-tests to determine if there are any differences. In this case, how many comparisons would need to be made to compare all pairs of six means.
    a. 3
    b. 6
    c. 12
    d. 15
    e. 30

40. In the previous scenario, the six means could be compared with a single analysis of variance (ANOVA) F-test to determine if any pairs of means differ. Which of the following statements is most accurate (assume each test would be done at the .05 significance level).
    a. Multiple t-tests and the single F-test are equally good ways to determine if any means differ.
    b. Multiple t-tests are preferable to a single F-test.
    c. A single F-test is preferable to multiple t-tests.
    d. Which one is preferable depends on whether the data are observational or experimental.
    e. None of the above.