

TMTA Statistics Exam 2011

1. Last month, the mean and standard deviation of the paychecks of 10 employees of a small company were \$1250 and \$150, respectively. This month, each one of the 10 employees received a \$125 raise on his/her check. What is the mean and standard deviation of their paychecks this month?

- (A) Mean = \$1375, Standard Deviation = \$150
- (B) Mean = \$1375, Standard Deviation = \$275
- (C) Mean = \$1250, Standard Deviation = \$150
- (D) Mean = \$1262.50, Standard Deviation = \$162.50
- (E) Unable to be determined without having the actual data

2. Ali H. Mokdad, Ph.D., and colleagues from the Centers for Disease Control and Prevention, Atlanta, conducted a study to identify and quantify the leading causes of death in the United States. The researchers used year 2000 mortality data reported to the Centers for Disease Control and Prevention to identify the causes and number of deaths. The leading causes of death in 2000 are given below. Which type of graph is appropriate to display these data?

Leading Causes of Death	Counts
Tobacco	435,000
Poor Diet and Physical Inactivity	400,000
Alcohol Consumption	85,000
Microbial Agents	75,000
Toxic Agents	55,000
Motor Vehicle Crashes	43,000
Incidents Involving Firearms	29,000
Sexual Behaviors	20,000
Illicit Use of Drugs	17,000

- (A) Histogram
- (B) Stem plot
- (C) Scatterplot
- (D) Boxplot
- (E) Bar Chart

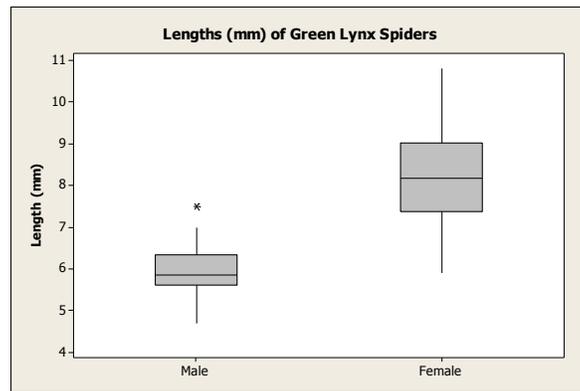
3. There are three senior citizens in a room, ages 68, 70, and 72. If a seventy-year-old person enters the room, the

- (A) mean age will stay the same, but the variance will increase.
- (B) mean age and variance will stay the same.
- (C) mean age will stay the same, but the variance will decrease.
- (D) mean age and variance will increase.
- (E) mean age and variance will decrease.

4. The Educational Testing Service reported that the mean Graduate Record Exam (GRE) score for all individuals who have taken the exam was 529. A student with a GRE score of 600 wants to know her relative standing in relation to the mean GRE. The most appropriate numerical summary for this purpose is

- (A) median.
- (B) interquartile range.
- (C) number of individuals that took the exam.
- (D) standard deviation.
- (E) correlation coefficient.

5. A biologist is interested in comparing the lengths of male and female green lynx spiders measured in millimeters. He makes side-by-side boxplots using data collected on the lengths of 30 male and 30 female green lynx spiders and produces the graph and summary statistics given below.



Group	Count	Mean	Median	Std Dev
Male	30	5.917 mm	5.850 mm	0.663 mm
Female	30	8.153 mm	8.175 mm	1.187 mm

The lengths of the female spiders tend to be

- (A) longer than the lengths of the males. The 25th percentile (Q1), which is 7.388 mm for the females, is longer than nearly all of the male lengths. The lengths of the females vary more than the lengths of the males.
- (B) shorter than the lengths of the males. The lengths of the females vary less than the lengths of the males.
- (C) about the same as the males since the summary statistics are nearly the same.
- (D) longer than the lengths of the males since the standard deviation of the female lengths is larger than the standard deviation of the male lengths.
- (E) shorter than the lengths of the males since the male lengths contain a high outlier and the female lengths do not.

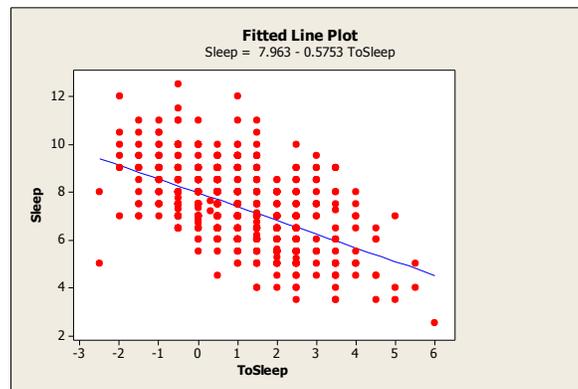
6. If the biologist in problem 5 decided that the Normal model was appropriate for the distribution of the lengths for the female spiders, what interval would contain the middle 95% of the female lengths?

- (A) 6.966 mm to 9.34 mm
- (B) 4.591 mm to 7.243 mm
- (C) 5.779 mm to 10.527 mm
- (D) 4.592 mm to 11.714 mm
- (E) 7.388 mm to 9.025 mm

7. Which one of the following statements involving correlation is possible and reasonable?

- (A) The correlation between hair color and eye color is 0.80.
- (B) The correlation between left foot length and right foot length is 2.35.
- (C) The correlation between hair color and age is positive.
- (D) The correlation between hair color and eye color is 0.
- (E) The correlation between the height of a father and the height of his first son is 0.6.

8. Can we predict the total amount of sleep an individual will receive based on the time he/she went to bed the night before? A fitted line plot of this relationship between the variables ToSleep (measured in hours with midnight equaling zero) and Sleep (measured in hours) for a simple random sample of individuals is given below. The correlation between the two variables is $r = -0.531$.



Which of the following is the correct interpretation of the slope of this regression equation?

- (A) Since the correlation is -0.531, amount of sleep decreases by 53.1%, on the average.
- (B) For every hour that passes before an individual goes to sleep, the amount of sleep that he/she receives decreases by 0.5753 hours, on the average.
- (C) As the amount of sleep that an individual receives decreases, the time that the individual goes to sleep decreases by 0.5753 hours, on the average.
- (D) For every hour that passes before an individual goes to sleep, the amount of sleep that he/she receives decreases by 7.963 hours, on the average.
- (E) For each additional hour of sleep that an individual receives, the time that he/she went to sleep decreases by 0.5753 hours, on the average.

9. For the relationship displayed in problem 8, what does the value of r^2 tell us?

- (A) The strength of the correlation between ToSleep and Sleep is 28.2%.
- (B) 28.2% of people get enough sleep each night.
- (C) Only 28.2% of the variability in the amount of sleep individuals get is explained by the variability in the times that they go to sleep the night before.
- (D) The relationship between the time an individual goes to sleep and the amount of sleep he/she gets is very strong.
- (E) There is a 28.2% chance that the regression equation will correctly predict the amount of sleep an individual receives given the time that he/she went to bed the night before.

10. A study found that 331 students who read the textbook before going to class had a better final semester grade, on the average, than 407 students who did not. Does the study seem to indicate reading the textbook before class has an effect on semester grade?

- (A) Yes, because better grades cause better study habits.
- (B) Yes, because the study showed that better study habits cause better grades.
- (C) No, but the study does seem to indicate an association between the two variables.
- (D) No, because it is unlikely that the two variables are connected.
- (E) No, because the correlation coefficient is close to 1.

11. A study published in the *New England Journal of Medicine* suggested that it is dangerous to enter a hospital in Ontario, Canada on a weekend. Researchers tracked over 4 million emergency admissions to hospitals over a 10-year period and found that patients admitted on weekends had a much higher risk of death than those who went to the emergency room on weekdays. The researchers said the difference in death rates was “statistically significant.” What does this mean?

- (A) The difference in death rates was important.
- (B) The null hypothesis is true.
- (C) The difference in death rates is higher than they would normally expect.
- (D) The P -value of the test must have been large.
- (E) Stay away from the emergency room on the weekends since it is causing deaths.

12. An educational software company wants to assess the usefulness of its software. It runs a poll on the Internet, asking users to indicate whether they like or dislike the software. Of 900 respondents, 610 said they liked the software. The results of the sample are probably

- (A) biased, because it is a voluntary response sample.
- (B) unbiased, because of the large sample size.
- (C) unbiased, because it is a simple random sample.
- (D) unbiased, but a larger sample should be used.
- (E) an example of a Normal distribution.

13. The weather reporter predicts that there is 90% chance of rain tomorrow for a certain region. What is meant by this phrase?

- (A) Rain occurs 90% of the time in this region.
- (B) It will rain 90% of the day tomorrow in this region.
- (C) 90% of the time it rains on this date in this region.
- (D) In circumstances “like this,” rain occurs 90% of the time in this region.
- (E) The occurrence of rain is “truly random” and will occur 90% of the time.

14. A consumer group surveyed the price for a certain item in five different stores and reported the median price is \$15. We visited four of the five stores and found the price of the item to be \$10, \$15, \$15, and \$25. Assuming that the consumer group correctly calculated the median, the price of the item at the store we did not visit

- (A) must be \$15.
- (B) must be below \$15.
- (C) must be above \$15.
- (D) must a value between \$10 and \$25.
- (E) could be any value.

15. The nutritional status of each of 1000 elementary school children was assessed and recorded as “poor,” “adequate,” or “excellent.” The academic performance for those children was rated as “below average,” “average,” or “above average.” The data are displayed in the table below.

		Poor	Adequate	Excellent	Totals
Academic	Below Average	70	95	35	200
Performance	Average	130	450	30	610
	Above Average	90	30	70	190
	Totals	290	575	135	1000

What is the probability that a randomly chosen child is below average academically, given that the child chosen had a poor nutritional status?

- (A) 0.3500
- (B) 0.2000
- (C) 0.2900
- (D) 0.2414
- (E) 0.0700

16. A student takes an eight question true-false quiz and guesses on each question. Find the probability of passing the quiz, if the lowest passing grade is six out of eight.

- (A) 0.7500
- (B) 0.1094
- (C) 0.5000
- (D) 0.1445
- (E) 0.8555

17. A survey was designed to study how the operations of a group of businesses vary with their size. Companies were classified as small, medium, and large. Questionnaires were sent to 200 randomly selected businesses of each size, for a total of 600 questionnaires. Since not all questionnaires in a survey of this type are returned, it was decided to examine whether the response rate varied with the size of the business. The data are given in the table below.

Size	Response	No Response	Total
Small	120	80	200
Medium	80	120	200
Large	40	160	200

Which of the following conclusions seems to be supported by the data in the table?

- (A) There are more small companies than large companies in the survey.
- (B) Small companies appear to have higher response rates than medium or big companies.
- (C) Exactly the same number of companies responded as didn't respond.
- (D) Small companies dislike larger companies.
- (E) The data should be displayed using percentages rather than counts since we are trying to make comparisons between the sizes of the companies.

18. A roulette wheel has 38 slots in which the ball can land. Two of the slots are green, 18 are red, and 18 are black. The ball is equally likely to land in any slot. The roulette wheel is going to be spun twice and the outcomes of the two spins are independent. The probability that the ball lands on red at least once is

- (A) 0.2244
- (B) 0.2493
- (C) 0.2770
- (D) 0.4986
- (E) 0.7230

19. A basketball player makes 80% of his free throws. At the end of a game, his team is losing by two points. He is fouled attempting a 3-point shot and is awarded three free throws. Assuming each free throw is independent, what is the probability that he makes at least two of the free throws?

- (A) 0.8960
- (B) 0.1040
- (C) 0.0960
- (D) 0.8000
- (E) 0.6800

20. The number of undergraduates at College A is approximately 25,000 while the number at college B is approximately 40,000. At both schools, a simple random sample of about 3% of the undergraduates is taken. We conclude

- (A) the sample from College A has less sampling variability than that from College B.
- (B) the sample from College A has more sampling variability than that from College B.
- (C) the sample from College A has about the same sampling variability as that from College B
- (D) the number of individuals sampled from each college is the same.
- (E) it is impossible to make any statements about the samples or sampling variability of the two samples, since the students surveyed were different groups of people.

21. The average age of cars owned by residents of a small city is 4 years with a standard deviation of 2.2 years. A simple random sample of 400 cars is to be selected, and the sample mean age of these cars is to be computed. What is the approximate probability that the average age of the 400 cars is less than 4.1 years old?

- (A) 0.3183
- (B) 0.1814
- (C) 0.6817
- (D) 0.9091
- (E) 0.8186

22. A recent poll was conducted to determine the percent of U.S. citizens who approve of the President's job performance. A random sample of 1006 citizens was drawn and 453 citizens said that they approve of his job performance. The margin of error was $\pm 3\%$ with 95% confidence. Which of the following statements is true?

- (A) An approximate 95% confidence interval for the percentage of citizens in the sample that approve of his job performance is 42% to 48%.
- (B) An approximate 95% confidence interval for the percentage of all U.S. citizens that approve of his job performance is 42% to 48%.
- (C) We are 95% confident that 45% of all U.S. citizens approve of his job performance.
- (D) We can be 95% confident that 45% is between 42% and 48%.
- (E) 95% of all U.S. citizens approve of his job performance with a margin of error of $\pm 3\%$.

23. A simple random sample of 20 commuters to the Los Angeles metropolitan area are selected, and each is asked how far they commute to work each day. In the sample, the mean distance is 64 miles and the standard deviation is 12 miles. Assume that in the population of all commuters to the Los Angeles metropolitan area, daily commuting distance follows a normal distribution, with some mean μ . A 90% confidence interval for the population mean commuting distance based on these data is

- (A) (61.32 miles, 66.68 miles)
- (B) (59.36 miles, 68.64 miles)
- (C) (58.38 miles, 69.62 miles)
- (D) (57.16 miles, 70.84 miles)
- (E) (62.96 miles, 65.04 miles)

24. At what age (measured in months) do infants speak their first word of English? A SRS of twenty children was selected and the age at which they spoke their first word was recorded. A stemplot of the data is given below. In addition, a t-test was conducted to determine if this sample provides evidence that children take longer than 12 months to speak their first word of English. The output from the test is also given below.

Stem-and-leaf of age N = 20
Leaf Unit = 1.0

```

0 7
0 899
1 00001111
1 2
1 55
1 7
1 8
2 00
2
2
2 6

```

One-Sample T: age

Test of $\mu = 12$ vs > 12

Variable	N	Mean	StDev	SE Mean	95% Lower Bound	T	P
age	20	13.0000	4.9311	1.1026	11.0934	0.91	0.188

Does this sample data provide good evidence that the mean age at which children speak their first word of English is greater than 12 months, using $\alpha = 0.05$?

- (A) Yes, there is strong evidence that the mean age at first word is greater than one year.
- (B) Since the sample mean is 13 months, we should reject the null hypothesis.
- (C) There doesn't appear to be enough evidence to suggest that the mean age at first word is greater than one year.
- (D) Yes, since we are 95% confident that the mean age at first word is at least 11.0934 months.
- (E) Yes, since there is only an 18.8% chance that the null hypothesis is correct.

25. Looking at the analysis of the data provided in problem 24, is there any reason why we should question the use of the t-procedure?

- (A) No, the use of the t-procedure is appropriate when σ is unknown.
- (B) Yes, the z-test should have been used instead of the t-test.
- (C) No, the data analysis in problem 24 is appropriate to answer the research question since the data are based on a simple random sample.
- (D) Yes, the sample size needs to be at least 30 in order to use the t-procedure.
- (E) Yes, the use of the t-procedure should be questioned because the stemplot shows that the data are right-skewed with a high outlier.

26. The national mean IQ test score for seventh grade girls is 100. An educator wonders if seventh grade girls who attend charter schools have a mean IQ different from the national mean IQ score. She selects a simple random sample of 50 seventh grade girls from charter schools and based on these data, she calculates a 95% confidence interval for the mean IQ of all seventh grade girls in charter schools and finds it to be (98.305, 105.695). Which of the following statements is true?

- (A) The study does not provide enough evidence that charter school seventh grade girls have a mean IQ different from the national mean IQ of 100, at the 0.05 level.
- (B) The study definitely concludes that charter school seventh grade girls have a mean IQ score higher than the national mean of 100, at the 0.05 level.
- (C) We are 95% confident that all charter school girls have their IQ scores between 98.305 and 105.695.
- (D) The sample mean is equal to the population mean and this can be stated with 95% confidence.
- (E) 95% of girls in charter schools have their IQ scores between 98.305 and 105.695.

27. Joe takes a simple random sample of 50 pennies and finds a 95% confidence interval for the mean age (in years) to be 9.4 ± 0.91 . Which of the following comments is most appropriate?

- (A) Joe can be 95% confident that the mean age of all pennies in circulation is 9.4 years.
- (B) 95% of all pennies in circulation are between 8.49 and 10.31 years old.
- (C) Joe is 95% confident that the mean age of all pennies in circulation is between 8.49 and 10.31 years.
- (D) Joe could get a smaller confidence interval by increasing the confidence level.
- (E) Joe could get a larger confidence interval by taking a larger sample size.

28. Twelve golfers are asked to play a round of golf on each of two consecutive weeks. During one round, the golfers use one brand of clubs and during the second, another brand. Which brand they use in which round is determined at random. All scores are recorded, and each golfer is asked to try to play his or her best in each round. Based on their results, if you want to determine if there is enough evidence that the mean score using brand 2 is less than the mean score using brand 1, you should use a

- (A) one-sample t -test
- (B) matched pairs t -test
- (C) two-sample t -test
- (D) one sample z -test
- (E) any of the above are valid. It is at the experimenter's discretion.

29. I conduct a statistical test of hypotheses and find that the test is statistically significant at level $\alpha = 0.05$. I may conclude

- (A) the test would also be significant at level $\alpha = 0.10$.
- (B) the test would also be significant at level $\alpha = 0.01$.
- (C) the p -value of the test is greater than 0.05.
- (D) the p -value of the test is less than 0.05.
- (E) both (A) and (D) are true.

30. A random sample of 1400 Internet users was selected from the records of a large Internet provider and asked whether they would use the Internet or the library to obtain information about health issues. Of these, 872 said they would use the Internet. Based on this data, a 95% confidence interval for the true proportion of all Internet users that would use the Internet to get information about health issues is

- (A) (35.6%, 39.8%)
- (B) (35.2%, 40.2%)
- (C) (60.2%, 64.4%)
- (D) (59.8%, 64.8%)
- (E) (59.0%, 65.6%)

31. The mean area of the several thousand apartments in a new development is advertised to be 1250 square feet. A tenant group thinks that the apartments are smaller than advertised. They hire an engineer to measure a sample of apartments to test their suspicion. Which of the following are the appropriate null hypothesis, H_0 , and alternative hypothesis, H_a ?

- (A) $H_0: \mu = 1250, H_a: \mu < 1250$
- (B) $H_0: \mu = 1250, H_a: \mu > 1250$
- (C) $H_0: \bar{x} = 1250, H_a: \bar{x} < 1250$
- (D) $H_0: p = 1250, H_a: p < 1250$
- (E) $H_0: 1250, H_a: < 1250$

32. An insurance company tracks life expectancy information to assist in determining the cost of life insurance policies. The insurance company will only change its premium structure if there is evidence that people who buy its policies are living longer than 77 years on the average. For the following null and alternative hypotheses, which of the statements is a correct conclusion of the test if H_0 is rejected?

H_0 : The average life expectancy of the insurance company's policy holders is 77 years.

H_a : The average life expectancy of the insurance company's policy holders is more than 77 years.

- (A) The insurance company should not change their premiums.
- (B) The average life expectancy is less than 77 years.
- (C) The insurance company needs to change its premiums.
- (D) The insurance company should change the premiums when the average age is 77 years.
- (E) The life expectancy of all policy holders is more than 77 years.

33. A pack of a certain brand of cigarettes displays the statement, "1.5 mg nicotine average per cigarette." Let μ denote the mean nicotine content per cigarette for all cigarettes of this brand. People who smoke this brand would probably be disturbed if it turned out that the true average nicotine content exceeded the claimed value. We carry out a test of hypotheses: $H_0: \mu = 1.5$ and $H_a: \mu > 1.5$. Which is a correct description of the Type II error?

- (A) We believe the advertisement but the true average content is above 1.5 mg.
- (B) The true average nicotine content is 1.5 mg, but we decide that it is higher.
- (C) The true average nicotine content is 1.5 mg, but we decide that it is lower.
- (D) The true average nicotine content is 1.5 mg, but we decide that it is different.
- (E) The p-value of the test must be smaller than α .

34. A particular election is approaching and we would like to estimate ahead of time what percent of registered voters plan to vote for Candidate X. We plan to take a SRS from the list of registered voters in the state and ask them who they plan to vote for. How many people should we have in the sample if we want to estimate that percent with a margin of error of 3% and we want to feel 95% confident in our estimation?

- (A) $n = 33$
- (B) $n = 534$
- (C) $n = 1068$
- (D) $n = 400$
- (E) $n = 251$

35. A medical researcher is working on a new treatment for a certain type of cancer. The average survival time after diagnosis on the standard treatment is two years. In an early trial, she tries the new treatment on three subjects who have an average survival time after diagnosis of four years. Although the survival time has doubled, the results are not statistically significant, even at the 0.10 significance level. The explanation is

- (A) the sample size is small.
- (B) the placebo effect is present, which limits statistical significance.
- (C) that although the survival time has doubled, in reality the actual increase is really 2 years.
- (D) the calculation was in error. The researchers forgot to include the sample size.
- (E) that the p-value is smaller than the 0.10 significance level.

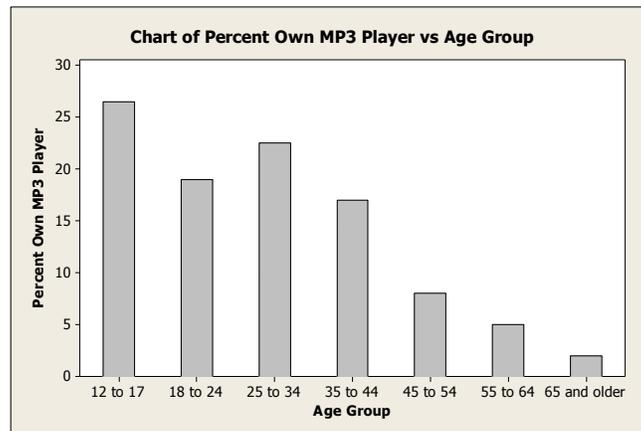
36. In general, which of the following is not a true statement about the sample standard deviation, s ?

- (A) s is always less than or equal to zero.
- (B) s is always greater than or equal to zero.
- (C) s is equal to zero only when all of the observations are the same value.
- (D) s is not resistant to outliers.
- (E) s is the square root of the sample variance.

37. An investigator has a computer file showing family incomes for 1,000 subjects of a certain study. These range from \$5,800 a year to \$98,600 a year. By accident, the highest income in the file gets changed to \$986,000. Which of the following is correct?

- (A) The mean, median, standard deviation, and IQR will all be affected.
- (B) The mean, median, standard deviation, and IQR will not be affected.
- (C) The median and IQR will increase, but the mean and standard deviation will not be affected.
- (D) The mean and standard deviation will increase, but the median and IQR will not be affected.
- (E) The five-number summary will not be affected.

38. Portable MP3 players, such as Apple iPod, are popular, but not equally popular with people of all ages. The following bar graph shows the percents of several age groups who own portable MP3 players.



Which feature best describes the association between age group and MP3 players?

- (A) The distribution of ages looks skewed right with a center at the 35 to 44 age group.
- (B) The distribution of MP3 players is centered at the 35 to 44 age group.
- (C) MP3 players are popular mainly among people less than age 35.
- (D) The distribution is bimodal, skewed right with an outlier at 65 and older.
- (E) The distribution is slightly left skewed with a center at the 35 to 44 age group.

39. According to a recent poll, about 56% of all American adults owned a cell phone. The results are based on interviews with a randomly selected national sample of 801 adults that are 18 years and older. The margin of error is reported to be 3.5% with a level of confidence of 95%. Which of the following correctly interprets the reported margin of error of 3.5%?

- (A) In about 3.5% of all random samples from this population, the sample percent will equal the population percent.
- (B) The probability that a 95% confidence interval based on this poll does not cover the population proportion is 3.5%.
- (C) In about 95% of all random samples of this size from the same population, the difference between the population percent and the sample percent will be more than 3.5%.
- (D) In about 3.5% of all random samples of this size from the same population, the difference between the population percent and the sample percent will be less than 95%.
- (E) In about 95% of all random samples of this size from the same population, the difference between the population percent and the sample percent will be less than 3.5%.

40. A certain test is used to screen blood specimens for the presence of antibodies to HIV, the virus that causes AIDS. Antibodies indicate the presence of the virus. The test is quite accurate, but is not always correct. The table below gives approximate probabilities of positive and negative test results when the blood tested does and does not actually contain antibodies to HIV.

	Positive Result	Negative Result
Antibodies Present	0.9985	0.0015
Antibodies Absent	0.0060	0.9940

Suppose that 1% of a large population carries antibodies to HIV in their blood. What is the probability that the test is negative for a randomly chosen person from this population?

- (A) 0.99000
- (B) 0.984075
- (C) 0.000015
- (D) 0.015925
- (E) 0.9954

TMTA Statistics Exam – Additional Questions

1. Coin collecting is a popular hobby but some older coins can be very rare. If you examine the coins carried by ordinary individuals in their pockets or purses you will find that most of the coins have relatively recent dates but occasional coins will be much older. Suppose the variable X is the 'date on the individual coins' found in a random person's possession. Which of these descriptions of the distribution X is likely to be closest to the truth?

- (A) skewed to the right with mean date more recent than the median
- (B) skewed to the right with median date more recent than the mean
- (C) symmetric with mean date less recent than the median
- (D) symmetric with mean date equal to the median
- (E) skewed to the left with mean date less recent than the median

2. A random sample of 900 thirteen to seventeen-year-olds found that 411 had a computer in their room with Internet access. Let p be the proportion of all teens in this age range that have a computer in their room with Internet access. Suppose you wished to see if the majority of teens in this age range have a computer in their room with Internet access. To do this, you test the hypotheses: $H_0: p = 0.50$, $H_a: p > 0.50$. The P -value for your test is

- (A) greater than 0.10.
- (B) between 0.05 and 0.10.
- (C) between 0.01 and 0.05
- (D) below 0.01.
- (E) unable to be determined without looking at the actual data.