Lee County Invitational — January 26, 2008

Statistics Individual — Fort Myers High School

For each of the following questions mark the best answer on your scantron. If the correct answer is not present, choose choice E for "None of the above."

NOTE: For this test, use the following values for special areas under the normal distribution, for all other values use the provided Z-table. Let Z_a be defined so that $P(z>Z_a)=a$, then use the following values:

$Z_{0.1}= 1.282$
$Z_{0.05} = 1.645$
$Z_{0.025} = 1.96$
$Z_{0.01} = 2.326$

1. In the Major League Baseball World Series, the two teams that face one another must play a series of 7 games and win 4 to be crowned the champions. Once one team has won 4 games, the series is ended. Suppose that in the World Series the Yankees play the Mets. The Yankees have a $\frac{9}{10}$ chance of beating the Mets. What is the probability that the series will end without 7 games being played (rounded to 4 decimal places)?

   A. 0.7283 B.0.7285 C. 0.9542 D. 0.9854 E. NOTA

2. In a standard deck of 52 playing cards, what is the probability of drawing a prime number or a red card (NOTE: the Jack, Queen, King and Ace cards are not considered to be prime numbers)?

   A. $\frac{8}{13}$ B. $\frac{17}{26}$ C. $\frac{19}{26}$ D. $\frac{21}{26}$ E. NOTA

For questions 3 and 4, let X~N(μ,σ):

3. What is the mean of -2X-7?

   A. $-2\mu-7$ B. $\mu+7$ C. $\mu-7$ D. $-2\mu+7$ E. NOTA

4. What is the standard deviation of -2X-7?

   A. $4\sigma-7$ B. $2\sigma-7$ C. $-4\sigma-7$ D. $-2\sigma-7$ E. NOTA

5. Gianluigi Buffon, the goalkeeper for the Italian National Soccer Team, wants to determine the proportion of times he allows a shot on goal to score (enter the goal). At a recent practice, his teammates shot 96 times on goal and he allowed 10 shots in. What is the 90% confidence interval of the true proportion of shots he allows to score (to 4 decimal places)?

   A. (0.0431, 0.1653) B. (0.0529, 0.1554) C. (0.8347, 0.9569)
   D. (0.8445, 0.9471) E. NOTA

6. A triangle with sides 5, 6 and 9 has a circle inscribed in it. What is the probability a point will be placed inside the triangle but outside the circle, given that it lands somewhere in the triangle?

   A. $\frac{10-\sqrt{\pi}}{10}$ B. $\frac{10-\pi}{10}$ C. $\frac{10-\sqrt{2}\pi}{10}$ D. $\frac{5-\pi}{5}$ E. NOTA

7. Through extensive research, it has been found that Tyra Banks says the word “fierce” an average of 10 times on each episode of “America’s Next Top Model” with a standard deviation of 4. However, all of the data has recently been destroyed by an unknown source. What can be inferred about the probability of Tyra saying the word “fierce” between 4 and 16 times in a single episode?

   A. At least $\frac{4}{9}$ B. At most $\frac{4}{9}$ C. At least $\frac{5}{9}$
   D. At most $\frac{5}{9}$ E. NOTA

Use the following information for questions 8-10: The president of Pagong Incorporated, a company devoted to the marketing of sales of luxury island getaways, has claimed that sales people are averaging no more than 10 sales per week. As the representative of the people’s union, you decide to try and refute this claim made by the president. You take a random sample of 49 employees and find the mean and standard deviation for the sample is 12 and 7 sales, respectively. Assume $\alpha = .05$, let μ represent the average number of sales per week.

8. What are the null and alternative hypotheses for this study?

   A. H_0: $\mu \le 10$, H_a: $\mu > 10$
   B. H_0: $\mu \ge 10$, H_a: $\mu < 10$
   C. H_0: $\mu \ne 10$, H_a: $\mu = 10$
   D. H_0: $\mu = 10$, H_a: $\mu \ne 10$
   E. NOTA

9. What is the p-value for this study (rounded to 4 decimal places)?

   A. 0.0001 B. 0.0228 C. 0.0401 D. 0.0500 E. NOTA

10. If the null hypothesis is not rejected in this study, which of the following is true?

   A. A type I error has occurred
   B. A type II error has occurred
   C. A double blind error has occurred
   D. The correct conclusion has been made
   E. NOTA

11. For the following dataset, what is the spread of the middle 50% of the data:
   6, 9, 4, -12, 5, 9, 10, 10, 10, 3, -8, 9

   A. 3 B. 6 C. 13 D. 22 E. NOTA

12. Which of the following is determined true by an application of the Law of Large Numbers?

A. For any normal distribution, the mean, median and mode all have the same value.
B. The sampling distribution of the sample mean can be approximately modeled by a normal distribution.
C. For independent events A and B, P(A or B) = P(A) + P(B) – P(A and B)
D. Probability of getting a head when flipping a fair coin is $\frac{1}{2}$.
E. NOTA

13. Given that, for a discrete random variable X, $\sigma = 7$ and $E(X^2) = 55$, what is the mean of X?

A. 6 B. 104 C. $\frac{49}{55}$ D. $\frac{55}{49}$ E. NOTA

14. A local high school recently polled a random group of 400 students to determine their preferences on athletics offered at the school. When asked whether they thought if the school offered enough of a variety of programs, 300 said that they felt the school did offer enough of a variety. The (1-α)% confidence interval for the true proportion who felt that there was **not** enough of a variety is (.218,.282), rounded to 3 decimals. What interval contains the value of α?

A. (0.04, 0.08) B. (0.08,0.12) C. (0.12,0.16) D. (0.16,0.20) E. NOTA

15. If a number is chosen at random from the first 100 whole numbers, what is the probability that the number has an odd number of factors?

A. $\frac{9}{100}$ B. $\frac{1}{10}$ C. $\frac{2}{25}$ D. $\frac{11}{100}$ E. NOTA

16. At a student summer camp, heights among students are normally distributed. The mean height is 62 inches and the standard deviation of heights is 4 inches. What is the probability that a randomly selected student is taller than 68 inches in height (to 4 decimals)?

A. 0.0014 B. 0.1469 C. 0.2546 D. 0.2514 E. NOTA

17. A study investigating the relationship of voting patterns with age takes a random sample of 100 people in a certain precinct. If the standard deviation of the ages is 16 years, what is the probability that the mean age of the individuals is within 0.5 years of the mean age for all individuals in the precinct (to 4 decimal places)?

A. 0.0240 B. 0.2434 C. 0.7698 D. 0.9974 E. NOTA

Use the following table for questions 18-20: The following table gives the distribution of number of meltdowns per kid at Walt Disney World that occurred on October 9, 2007.

# of meltdowns a child has at WDW on 10/9/2007	Probability
0	.46
1	.23
2	X
3	.03
4	.01

18. What is the value of X in the above table?

    A. 0.27 B. 0.31 C. 0.43 D. 0.73 E. NOTA

19. What is the mean number of child meltdowns that occurred on October 9, 2007?

    A. 0.2 B. 0.9 C. 1 D. 2.5 E. NOTA

20. What is the median number of child meltdowns that occurred on October 9, 2007?

    A. 0 B. 1 C. 2 D. 3 E. NOTA

21. A company, made up of 10 men and 8 women, is randomly picking 5 people to go to a conference in New York. What is the probability that more men than women will go to the conference, given that at least one woman was selected?

    A. $\frac{7}{11}$ B. $\frac{10}{17}$ C. $\frac{21}{34}$ D. $\frac{20}{33}$ E. NOTA

22. The following stem and leaf plot gives a graphical representation of age (in years) of people who entered a local bookstore on a given day. Based on the stem and leaf (given below), which of the following describes the distribution?

    I. Skewed left
    II. Skewed right
    III. Normal
    IV. Bimodal

    0|
    1|1 2 5 9
    2|0 0 5 5 7 8 9
    3|0 0 0 0 1 2 8 8 8 9
    4|0 0 3 4 5 6 7 8 8 8 9
    5|3 4 5 5 6 7 7 8 8 8 9
    6|0 0 0 3 5 6 6 6 8 8 8 9 9 9
    7|0 0 4 4 4

    A. I only B. I and II only C. I and IV only D. III and IV only E. NOTA

23. A population of 9 butterflies in an artificial rainforest have wingspans (in inches) given in the following: {3, 9, 1, 5, 5, 10, 3, 6, 5}. If a researcher wants to take a random sample of 3 butterflies from the above population, what is the maximum magnitude for sampling error of the mean that can occur?

A. $\frac{26}{27}$ B. $\frac{26}{9}$ C. $\frac{28}{27}$ D. $\frac{28}{9}$ E. NOTA

24. When performing a significance test on a mean from a normal population distribution and you have a small sample size, what distribution is used to perform the test?

A. F-distribution B. t-distribution C. χ^2-Distribution
D. Z-distribution E. NOTA

25. Sailey Inc. has developed a new drug to reduce the aliments of the common cold. To test the strength of their new drug, they set up a double-blind experiment in which they given one random group their drug (group 1) and a second random group the current leading drug in the field (group 2). They want to show that the mean amount of time with symptoms is less using their drug over their competitor. Using the table of statistics below that the researchers recorded based upon the study, what is the p-value for the test (to 2 decimal places)?

	Group 1	Group 2
Number of participants	36	42
Average time with symptoms (in days)	2.7	3.2
Standard Deviation (in days)	1.6	1.4

A. 0.04 B. 0.93 C. 0.07 D. 0.96 E. NOTA

26. Janet wants to study the effect of occupation on happiness in Cape Coral, Florida. To do this, she determines a number of occupations to be studied: government job, teacher, banker, plumber and administrative assistant. She then retrieves a list of all people who work in one of the 5 fields above in Cape Coral (luckily, none of the people in the population are in more than one of the fields listed). She then takes a random sample within each of the occupations to sample. What type of sampling has Janet used?

A. Cluster Sample
B. Simple Random Sample
C. Stratified Random Sample
D. Systematic Random Sample
E. NOTA

27. A scatterplot has been drawn representing age versus amount of time to complete a marathon (in hours). A least squares line has been drawn and its equation is given by (time) = 3.2 – 0.42(age). If it is known that r^2=0.8843, what is the value of the correlation coefficient (to 4 decimal places)?

A. -0.9404 B. -0.7820 C. 0.7820 D. 0.9404 E. NOTA

28. In a recent study to determine the effect of amount of time playing Dance, Dance Revolution has on GPA, a researcher randomly studied 34 participants (numbering them 1-34). The results were plotted on a scatterplot putting time versus GPA. A line of best fit was then drawn based upon the results. Let D_a be the distance from the point representing the a^{th} participant to the line of best fit. If it is known that $D_1 + D_2 + \ldots + D_{30} = 4.5$, what is the value of $D_{31}+D_{32}+D_{33}+D_{34}$?

A. 2.75 B. 4.5 C. 0 D. -4.5 E. NOTA

29. Which of the following is/are true regarding the Central Limit Theorem?

I. The sampling distribution of $\overline{Y}$ is approximately normal regardless of the distribution of Y.

II. The sampling distribution of $\overline{Y}$ looks more like the population distribution of Y as the sample size increases.

III. For a population distribution Y with mean μ and standard deviation σ, $\overline{Y}$ is distributed normally with mean μ and standard deviation $\sqrt{\frac{\sigma}{n}}$ for sample size n taken.

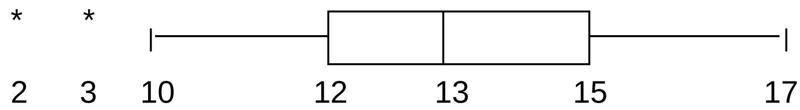
A. I only B. I, III only C. II, III only
D. III only E. NOTA

30. A recent poll with a large sample size looked at the age of the people who responded to their survey. The mean age for the sample was μ with standard deviation $\sigma \neq 0$. Spring and Paul each conducted a confidence interval for the average age. Spring took a (1-α)% confidence interval on the mean which resulted in the interval (X,Y). Paul took a (1-$\frac{\alpha}{2}$)% confidence interval on the mean which resulted in the interval (X’, Y’). Which of the following statements is true if $0 < \alpha < 1$?

A. X’>X and Y’>Y
B. X’<X and Y’<Y
C. X’>X and Y’<Y
D. X’<X and Y’>Y
E. NOTA

Southwest Florida Invitational January 26, 2008
Statistics Team Round 1

Answer the following questions based on the following box and whisker plot:



A. What is the range of the data set?
B. What is the interquartile range (IQR) of the data set?
C. If X=25th percentile and Y=50th percentile, what is the value of Y-X?
D. If all the values of the data points were increased by 7, what would the new value of the median be?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 2

In her English class, Shelly has grades of 95%, 64% and 80% on quizzes and 85% and 79% on tests.

A. If quizzes and tests are weighted the same, what grade must Shelly get on her next test in order to average an 83% for her tests and quizzes?
B. If tests are weighted twice and quizzes weighted once, what must Shelly get on her next quiz in order to average an 83% for her tests and quizzes?
C. Shelly’s teacher tells the class that the lowest quiz grade will be dropped from their overall average. Tests and quizzes are weighted equally. Shelly took her fourth (and final) quiz in the class, which caused her overall average to increase by 2.6%. What grade did Shelly get on the fourth quiz?
D. Shelly’s teacher was confused when determining the grades and input each of her totals backwards (if Shelly got a 58% it was input as an 85%). The tests and quizzes are weighted equally. What is the absolute value of the difference between Shelly's true average and the average her teacher got for her?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 3

A. What is the probability of drawing a black card, followed by a red card from a standard 52 card deck without replacement?
B. What is the probability of getting a full house when dealt 5 cards from a standard 52 card deck?
C. What is the probability of getting three of a kind when dealt 5 cards from a standard 52 card deck?
D. What is the probability of getting a flush when dealt 5 cards from a standard 52 card deck (to 4 decimals)?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 4

A. If you throw a dart at a circular target with a 3 inch radius, what is the probability of landing within 1.5 inches of the center, given that you hit the target?
B. A target is made of a square inscribed in a circle which is in turn inscribed in a second square. What is the probability of hitting the target in the small square region, given that you hit the target?
C. Tyler decides to test his luck at a game of chance at a carnival in a game that costs $2 to play. In the game he must break one of 10 balloons with a dart. In the balloons are dollar amounts he receives if he breaks the balloon and all balloons have an equal chance of being popped. Five of the balloons contain $0, two balloons contain $1, two balloons contain $2 and one balloon has $5. What is Tyler's expected game or loss in this game?
D. With regards to the game in part C, the owner of the game wants to change the cost of playing to the "fair price." What is the fair price of playing the game (ie. the price to play the game that makes the expected winnings $0)?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 5

Given that $X \sim N(12,4)$:

A. What is the Z-score associated with a data point of 15?
B. What is the data point that is 1.6 standard deviations below the mean?
C. If $P(9 \le X \le 17) = 0.6678$ and $P(X \le 9) = 0.2266$, what is $P(12 \le X \le 17)$?
D. The 20th percentile is determined by the data point of X. What is the data point that determines the value of the 80th percentile (in terms of X)?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 6

A. Determine the number of distinct arrangements of THEAMAZINGRACE
B. How many ways can a wedding party, composed of a bride, a groom, 3 bridesmaids and 3 ushers be photographed in a line if the bride and groom must be standing next to each other?
C. If ${}_xP_y = \frac{((25!)!)!}{((3!)!)!}$, then what is the value of x-y?
D. What is the sum of all possible values of n if $\frac{1}{4}({}_nP_2)^2 - 27({}_nC_2) - 28 = 0$?

Southwest Florida Invitational
Statistics Team

January 26, 2008
Round 7

The table to the right lists certain statistics for a study which looked at amount of times the word alliance was used per episode on various reality TV shows. The table below shows the information based on when the show was aired: earlier seasons compose the shows shown before 2004 and the later seasons compose shows shown during or after 2004. Since there have been well over 1000 reality TV episodes, a sample has been taken. Let μ_E denote the mean number of times it was used during earlier seasons and let μ_L denote the mean number of times it was used during later seasons. (NOTE: For this question, use the following values for special areas under the normal distribution, for all other values use the provided Z-table. Let Z_a be defined so that $P(z>Z_a)=a$, then use the following: $Z_{0.1}= 1.282$, $Z_{0.05} = 1.645$, $Z_{0.025} = 1.96$, $Z_{0.01} = 2.326$)

	Earlier Seasons	Later Seasons
Sample average number of times alliance was said	12	7
Sample standard deviation of number of times alliance was said	3	5
Number of episodes	91	91

A. What is the 90% confidence interval for $\mu_E - \mu_L$? (round to 3 decimal places)
B. What is the 95% confidence interval for $\mu_E - \mu_L$? (round to 3 decimal places)
C. If we wish to test H_0: $\mu_E - \mu_L = 0$ vs. H_a: H_0: $\mu_E - \mu_L \neq 0$, what is the value of the test statistic? (round to 3 decimal places)
D. If the p-value for the above test (in part C) is Q, what is the p-value (in terms of Q) of H_0: $\mu_E - \mu_L > 0$ versus H_a: $\mu_E - \mu_L \leq 0$?

Southwest Florida Invitational
Statistics Team

January 26, 2008
Round 8

For each of the following statements determine if they are true or false. Please write either "TRUE" or "FALSE."

A. The 95% confidence interval for the mean length of ducks (in inches) in a pond is determined to be (4,18). This is interpreted that 95% of the ducks in the pond have a length between 4 and 18 inches.
B. As the number of degrees of freedom, n, for the t-distribution increases towards infinity, the t-distribution approaches the χ^2-distribution.
C. All else being the same, if the sample size quadruples for a 95% confidence interval for a mean, then the size of the interval increases by a factor of 2.
D. $X \sim N(\mu,\sigma)$ and suppose that the Z-score associated with data point a is Z_a, then $P(Z \leq 2Z_a) = 2P(Z \leq Z_a)$.

Southwest Florida Invitational January 26, 2008
Statistics Team Round 9

A. If $f(x) = 3x - 6$, what is the probability that $f(x) \geq -4$ if $x \in [-3,3]$?
B. If $f(x) = x^2 - 4x + 5$, what is the probability that $f(x) \geq 0$ if $x \in [-10,12]$?
C. If $f(x) = x^3 - 4x^2 - x + 2$, what is the probability that $f(x) \geq -2$ if $x \in [-3,5]$?
D. Find the sum of the possible values of a in the following: If $f(x) = x^2 + 3x - 18$, the probability that $f(x) \geq -8$ if $x \in [-7, a]$ is $\frac{3}{10}$.

Southwest Florida Invitational January 26, 2008
Statistics Team Round 10

Suppose a certain disease has a rate of infection of 2%. A certain test for this disease gives a false positive reading 3% of the time. This test also gives a false negative reading 4% of the time.

A. What is the probability that, in a group of 5 people, at most 1 person has the disease? (round to 3 decimals)
B. If a person actually has the disease, what is the probability that the test will give a positive reading for the disease? (round to 3 decimals)
C. If the test reads positive, what is the probability that the person actually does not have the disease? (round to 3 decimals)
D. If the test reads negative, what is the probability that the person actually does have the disease? (round to 3 decimals)

Southwest Florida Invitational January 26, 2008
Statistics Team Round 11

Consider the following data points: (3,6); (5,7); (6,7); (8,8); (10,11); (11,9)

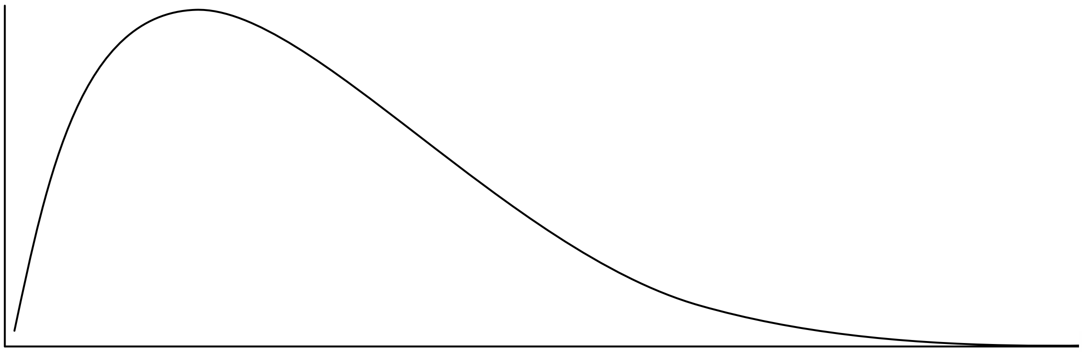
A. Determine the y-intercept of the line of best fit through the above data points (write as a coordinate and round to 2 decimals).
B. Determine the slope of the line of best fit through the above data points (round to 2 decimals).
C. With regards to the above data set, if each x-value was increased by 3 and each y-value was decreased by 7, what is the new value of the y-intercept of the line of best fit (round to 2 decimals)?
D. With regards to the above data set, if each x-value was increased by 3 and each y-value was decreased by 7, what is the new value of the slope of the line of best fit (round to 2 decimals)?

Southwest Florida Invitational January 26, 2008
Statistics Team Round 12

Let q be a real number. Determine the value of q in each of the following that makes each function a probability density function.

A. $f(x) = q, \ -1 \le x \le 5$
B. $g(x) = qx + \frac{1}{6}, \ 0 \le x \le 3$
C. $h(x) = \sqrt{q^2 - x^2}, \ -q \le x \le q$
D. $k(x) = \sqrt{1 - \frac{x^2}{q^2}}, \ -q \le x \le q$

For the following density curve, determine the following:



A. Between the mean, median and mode what value is the smallest?
B. Between the mean, median and mode what value is second largest?
C. Between the mean, median and mode what value is the largest?
D. Which of the following terms best describes the above density curve: skewed left, normal, bimodal, skewed right?

The probability of rain on a given day in Fort Myers, Florida is $\frac{2}{3}$ and the probability of wind on a given day in Fort Myers, Florida is $\frac{3}{5}$. The probability of rain and wind are independent.

A. What is the probability of rain and wind?
B. What is the probability of rain or wind?
C. What is the probability of at least one of rain or wind?
D. What is the probability of, in the next 5 days, it will rain at least 3 days?

Let β = {4,7,0,3,-3,5,5,5,5,7,3,7,9}

A. What is the mean of β?
B. What is the mode of β?
C. What is the interquartile range of β?
D. What is the range of β?

1. Notice that, for example, 6 games to be played with the Yankees winning the first 5 games will have 3 Yankee wins and 2 Mets wins in any order, but the sixth game MUST have a Yankee win

   So now P(end without 7 games) = 1 – P(end with 7 games)

   = 1 – [P(Yankees win in 7) + P(Mets win in 7)]

   $= 1-[{}_6C_3(\frac{9}{10})^3(\frac{1}{10})^3 * (\frac{9}{10}) + {}_6C_3(\frac{9}{10})^3(\frac{1}{10})^3 * (\frac{1}{10})] = 0.9854$

2. There are 4 prime numbers in each suit, so the probability of drawing a prime is $\frac{16}{52}$. Also, there are 26 red cards, so the probability of drawing red is $\frac{26}{52}$. Now, there are 8 prime, red cards, so that probability is $\frac{8}{52}$. So,

   P(red or prime) = $\frac{16}{52}+\frac{26}{52}-\frac{8}{52}=\frac{17}{26}$.

3. By definition $aX+b$, has mean $a\mu+b$. So the mean is $-2\mu-7$.

4. By definition $aX+b$ has variance $a^2\sigma^2$. So the variance is $4\sigma^2$, so the standard deviation is 2σ.

5. In this problem $\hat{\pi}=\frac{10}{96}$. Also, $Z_{\frac{.10}{2}}=Z_{.05}=1.645$ from the table at the top of the test. So, the confidence interval is

   $$\left(\frac{10}{96}-1.645\sqrt{\frac{(\frac{10}{96})(\frac{86}{96})}{96}},\frac{10}{96}+1.645\sqrt{\frac{(\frac{10}{96})(\frac{86}{96})}{96}}\right)\equiv(0.0529,0.1554)$$

6. First, by Heron's Formula (with $s=\frac{5+6+9}{2}=10$), the area of the triangle is $\sqrt{10(10-5)(10-6)(10-9)}=10\sqrt{2}$. Now the radius of the inscribed circle can be found by $r=\frac{\text{area}\Delta}{\text{semiperimeter}}=\frac{10\sqrt{2}}{10}=\sqrt{2}$. So the area of the circle is $(\sqrt{2})^2\pi=2\pi$. So the probability it lands inside the triangle but outside the circle is

   $\frac{10\sqrt{2}-2\pi}{10\sqrt{2}}=\frac{20-2\sqrt{2}\pi}{20}=\frac{10-\sqrt{2}\pi}{10}$

7. Notice that

   $$P(4\le X\le 16)=P(|X-10|\le 6)=P(|X-10|\le 4(\frac{3}{2}))=P(|X-10|\le\frac{3}{2}\sigma)\ge 1-\frac{1}{(3/2)^2}$$

So the probability is at least 5/9.

8. Since you are trying to refute the claim that no more than 10 sales a week are being made, that is the alternative hypothesis. Thus, H_0: $\mu \le 10$, H_a: $\mu > 10$.

9. The test statistic for the study is $\frac{12-10}{7/\sqrt{49}} = 2$. So the p-value is $P(Z \ge 2) = .0228$

10. Type II

11. The spread of the middle 50% of the data is the interquartile range (IQR) of the data. Rearrange the data in ascending order:
-12, -8, 3, 4, 5, 6, 9, 9, 9, 10, 10, 10. Now the median of the data is between 6 and 9. So $Q_1 = \frac{3+4}{2} = 3.5$ and $Q_3 = \frac{9+10}{2} = 9.5$. So, IQR = 9.5 – 3.5 = 6.

12. The Law of Large numbers is a theorem that describes the long-run stability of a random variable. Thus, in the long run a coin will come up heads $\frac{1}{2}$ of the time.

13. Using the formula
$\sigma^2 = E(X^2) - [E(X)]^2 \Rightarrow 7^2 = 55 - [E(X)]^2 \Rightarrow [E(X)]^2 = 6 \Rightarrow E(X) = \sqrt{6}$

14. Using the left endpoint, the formula is $\frac{100}{400} - Z_{a/2}\sqrt{\frac{.25(1-.25)}{400}} = 0.218$. Solving yields that $Z_{a/2}$ = 1.478. So a/2 is between .0694 and .0708. So a is between 0.1388 and 0.1416. So a falls in (.12,.16).

15. A number has an odd number of factors if it is a perfect square (for example, 4 has the factors of 1, 2 and 4). The first 100 whole numbers are {0,1,2,…,99}. So the numbers with an odd number of factors are 1, 4, 9, 16, 25, 36, 49, 64 and 81 (0 has an infinite number of factors). So the probability is $\frac{9}{100}$.

16. The Z-score associated with a height of 68 is $\frac{68-62}{4} = 1.5$. So
$P(X \ge 68) = P(Z \ge 1.5) = .0668$

17.

18. Since the probabilities must sum to 1: $.46 + .23 + X + .03 + .01 = 1 \Rightarrow X = .27$

19. To get the mean we take a weighted average:

$$\frac{0(0.46)+1(0.23)+2(0.27)+3(0.03)+4(0.01)}{0.46+0.23+0.27+0.03+0.01}=0.9$$

20. The median is the number which 50% of the data falls below. Since 1 has 69% of the data at, or below, it, we conclude that 1 is the median.

21. P(more men than women|at least 1 woman) =

$$\frac{P(\text{moremen} \cap \text{atleast1woman})}{P(\text{atleast1woman})}=\frac{P(\text{moremen} \cap \text{atleast1woman})}{1-P(\text{nowomen})}$$

$$=\frac{\frac{({}_{10}C_4)({}_8C_1)}{{}_{18}C_5}+\frac{({}_{10}C_3)({}_8C_2)}{{}_{18}C_5}}{1-\frac{{}_{10}C_5}{{}_{18}C_5}}=\frac{20}{33}$$

22. I only (Skewed right would have a heavy low end tail, normal would be symmetric and mound shaped and bimodal would have 2 extremes)

23. First, the mean wingspan is $\frac{3+9+1+5+5+10+3+6+5}{9}=\frac{47}{9}$. Now the maximum magnitude for sampling error will occur if we select the 3 highest values or the 3 lowest values. Now the mean when selecting the three highest values is $\frac{6+9+10}{3}=\frac{25}{3}$ and the magnitude of the sampling error is $|\frac{47}{9}-\frac{25}{3}|=\frac{28}{9}$. The mean when selecting the three lowest values is $\frac{1+3+3}{3}=\frac{7}{3}$ and the magnitude of the sampling error is $|\frac{47}{9}-\frac{7}{3}|=\frac{26}{9}$. So the maximum magnitude of the sampling error is $\frac{28}{9}$.

24. By definition, t-distribution.

25. The test statistic is $t=\frac{2.7-3.2}{\sqrt{\frac{1.6^2}{36}+\frac{1.4^2}{42}}}=-1.46$. So the p-value is $P(Z \le -1.46)=0.07$ to 2 decimals.

26. By definition, stratified random sample

27. The correlation coefficient, r, is the square root of the coefficient of determination and has the same sign (positive or negative) as the least-squares line. So $r = \sqrt{r^2} = \sqrt{0.8843} = \pm 0.9404$, but since the slope of the least squares line is negative r = -0.9404.

28. The sum of the residuals to a line of best fit is always 0. So if the sum of the first 30 residuals is 4.5, the sum of the remaining 4 must be -4.5 so that the sum of all of the residuals is equal to 0.

29. I only (II should say $\overline{Y}$ looks more normal as the sample size increases, and III should say $\overline{Y}$ has standard deviation of $\frac{\sigma}{\sqrt{n}}$)

30. Using algebra we can show that $(1-\frac{\alpha}{2}) > (1-\alpha)$ for all values of α between 0 and 1. We know that as the confidence level of the interval is increased, the width of the confidence interval decreases. Thus, X'<X and Y'>Y

Answers Statistics Individual – January 2008 Invitational

1. D
2. B
3. A
4. E
5. B
6. C
7. C
8. A
9. B
10. B
11. B
12. D
13. E
14. C
15. A
16. E
17. C

18. A
19. B
20. B
21. D
22. A
23. D
24. B
25. C
26. C
27. A
28. D
29. A
30. D

1A. Range = (largest value) – (smallest value) = 17 – 2 =15

1B. $Q_1 = 12, Q_3 = 15 \Rightarrow IQR = 15 - 12 = 3$

1C. X = 12, Y = 13, so Y – X = 13 – 12 = 1

1D. If all the values of the data points increased by 7, the value of the median would increase by 7. So the new value of the median is 20.

2A. Let X be the value of her next quiz score. Then

$$\frac{95+64+80+85+79+X}{6} = 83 \Rightarrow X = 95$$

2B. Let X be the value of her next quiz score. Then

$$\frac{95+64+80+2(85)+2(79)+X}{8} = 83 \Rightarrow X = 97$$

2C. Her initial average was $\frac{95+64+80+85+79}{5} = 80.6$. If the average increased by 2.6, then her new average is 83.2 and also the quiz grade of 64 was dropped. Let X be her new quiz grade, then $\frac{95+80+85+79+X}{5} = 83.2 \Rightarrow X = 77$

2D. Shelly's true average is 80.6 as determined from part C. Her teacher's average for her is $\frac{59+46+8+58+97}{5} = 53.6$. So the positive difference is 80.6 – 53.6 = 27.

3A. P(black followed by red) = $\frac{26}{52}(\frac{26}{51}) = \frac{13}{51}$

3B. P(full house) = $\frac{({}_4C_2)({}_4C_3)(13)(12)}{({}_{52}C_5)} = \frac{6}{4165}$

3C. Probability of getting three of a kind means getting three cards of any one rank and two other cards that are both different. From this

$$P(3kind) = (\frac{{}_4C_3[({}_4C_1)({}_4C_1)({}_{12}C_2)]}{{}_{52}C_5})(13) = \frac{88}{4165}$$

3D. $P(flush) = \frac{{}_{13}C_5}{{}_{52}C_5}(4) = 0.0020$

4A. $P = \frac{Area_{small}}{Area_{large}} = \frac{\pi(1.5)^2}{\pi(3)^2} = \frac{1}{4}$

4B. Let x be the side length of the larger square. Then the radius of the circle is $\frac{x}{2}$, which means the diameter is x. Now, the smaller inscribed square has a diagonal of x, so the side length of the smaller square is $\frac{x}{\sqrt{2}} = \frac{x\sqrt{2}}{2}$. So the probability is

$$\frac{\text{Areasmallsquare}}{\text{Arealargesquare}} = \frac{(\frac{x\sqrt{2}}{2})^2}{x^2} = \frac{1}{2}$$

4C. $E(X) = \frac{5}{10}(0-2) + \frac{2}{10}(1-2) + \frac{2}{10}(2-2) + \frac{1}{10}(5-2) = -0.9$

4D. Let x be the fair price: $\frac{5}{10}(0-x) + \frac{2}{10}(1-x) + \frac{2}{10}(2-x) + \frac{1}{10}(5-x) = 0 \Rightarrow x = 1.1$

5A. $Z = \frac{15-12}{4} = 0.75$

5B. A data point that is 1.6 standard deviations below the mean corresponds to a Z-score of -1.6. So let x be the data point, then $-1.6 = \frac{x-12}{4} \Rightarrow x = 5.6$

5C. Notice that $P(X \le 12) = 0.5$. Now this can be rewritten as

$$P(X \le 12) = P(X \le 9) + P(9 \le X \le 12) \Rightarrow 0.5 = 0.2266 + P(9 \le X \le 12) \Rightarrow P(9 \le X \le 12) = 0.2734$$

Now, $P(9 \le X \le 17) = P(9 \le X \le 12) + P(12 \le X \le 17) \Rightarrow 0.6678 = 0.2734 + P(12 \le X \le 16)$
$\Rightarrow P(12 \le X \le 16) = 0.3944$

5D. Notice that the 80th percentile has 80% of the area below that value which means it has 20% of the area above the value. Thus, the 80th percentile is the same distance from the mean as the 20th percentile. So, the 20th percentile is (12-X) from the mean. So the 80th percentile is at 12+(12-X) = 24 – X.

6A. $\frac{14!}{2!*3!} = 7264857600$

6B. Notice one way this can happen is: B G 6! Where the 6! represents the number of ways the bridesmaids and ushers can be arranged. Now, these same arrangements can be used if the bride and groom change places. So there are 2(6!) arrangements of this form. Now, move the bride and groom down the line, they can stand next to each other in 7 locations. So the total number of combinations is 7[2(6!)] = 10080

6C. First, notice that ${}_xP_y = \frac{x!}{(x-y)!}$, so the denominator will solely determine the value of x-y. By expanding ((3!)!)! = (6!)! = 720!. This is the value of the denominator, and thus x – y = 720.

6D.

$$\frac{1}{4}({}_nP_2)^2 - 27({}_nC_2) - 28 = 0 \Rightarrow (\frac{1}{2}\bullet\frac{n!}{(n-2)!})^2 - 27({}_nC_2) - 28 = 0 \Rightarrow (\frac{n!}{2!(n-2)!})^2 - 27({}_nC_2) - 28 = 0$$

$$\Rightarrow ({}_nC_2)^2 - 27({}_nC_2) - 28 = 0$$

Now let u = ${}_nC_2$

So $u^2 - 27u - 28 = 0 \Rightarrow (u-28)(u+1) = 0 \Rightarrow u = -1,28$

Now u = -1 is extraneous because a combination cannot be negative.

For u = 28, we substitute to get 28 = ${}_nC_2$ and solve to get n = 8. So the sum of the solutions is 8.

7A. $(12-7) \pm Z_{.05}\sqrt{\frac{3^2}{91}+\frac{5^2}{91}} \equiv 5 \pm 1.645\sqrt{\frac{34}{91}} \equiv (3.994,6.006)$

7B. $(12-7) \pm Z_{.025}\sqrt{\frac{3^2}{91}+\frac{5^2}{91}} \equiv 5 \pm 1.96\sqrt{\frac{34}{91}} \equiv (3.802,6.198)$

7C. $\text{TS} = \frac{12-7}{\sqrt{\frac{3^2}{91}+\frac{5^2}{91}}} = 8.180$

7D. From the information $2P(Z \geq 8.108) = Q \Rightarrow P(Z \geq 8.108) = \frac{Q}{2}$. The P-value for the Lower tail, one-sided significance test is $P(Z \leq 8.108) = 1 - P(Z \geq 8.108) = 1 - \frac{Q}{2}$

8A. FALSE: An interpretation of the confidence interval could be that if all confidence intervals of n were taken then the true mean length would fall in those intervals 95% of the time.

8B. FALSE: As the number of degrees of freedom for the t-distribution goes to infinity, the t-distribution approaches the Z-distribution.

8C. FALSE: The width decreases by a factor of 2.

8D. FALSE: For example $P(Z \leq 1) = 0.8413$ but $P(Z \leq 2) = 0.9772 \neq 2(0.8413)$

9A. First, notice that $3x-6=-4 \Rightarrow x=\frac{2}{3}$, so $f(x)\geq -4$, when $x\geq \frac{2}{3}$. So the probability is $\frac{3-2/3}{3-(-3)}=\frac{\frac{7}{3}}{6}=\frac{7}{18}$

9B. Notice that the discriminant of f(x) is $(-4)^2-4(1)(5)=-4$ which means that there are no real roots. Because of this, it will never cross the x-axis and since the coefficient of the square term is positive, the graph opens upwards. Because of all of this the graph of f(x) is always above the x-axis, thus the probability is 1.

9C. Solving for the roots reveals
$x^3-4x^2-x+4=0 \Rightarrow x^2(x-4)-1(x-4)=0 \Rightarrow (x^2-1)(x-4)=0$ this shows that the roots of the equation are x = -1, 1, 4. By graphing the equation, f(x) is greater than 0 between -1 and 1 and also between 4 and 5. So the probability is $\frac{(1-(-1))+(5-4)}{5-(-3)}=\frac{3}{8}$

9D. Solving for the roots of f(x) yields that $x^2-3x-10=0 \Rightarrow x=-2,5$. So since the parabola is upward opening, f(x) is greater than 0 when x is less than -2 or greater than 5. There are two cases that need to be considered: if a is between the roots or if a is greater than the roots.
<u>Case 1</u>: If a is between the roots, then the probability that f(x) is greater than 0 is
$\frac{(-2-(-7))}{(a-(-7))}=\frac{3}{10} \Rightarrow \frac{5}{a+7}=\frac{3}{10} \Rightarrow a=\frac{29}{3}$.
<u>Case 2</u>: If a is greater than both of the roots then the probability that f(x) is greater than 0 is $\frac{(-2-(-7))+(a-5)}{(a-(-7))}=\frac{3}{10} \Rightarrow \frac{a}{a+7}=\frac{3}{10} \Rightarrow a=3$.
So the sum of the possible values of a is $\frac{29}{3}+3=\frac{38}{3}$

10A. P(at most one person) = P(none) + P(1) = $(.98)^5+{}_5C_1(.98)^4(.02)=0.996$

10B. P(test positive if they have disease) = 1 – P(false negative) = 1 – (.04) = 0.960

10C. P(no disease|positive reading) =
$\frac{P(positive\,|\,nodisease)*P(nodisease)}{P(positive)}=\frac{0.03(0.98)}{(0.98)(0.03)+(0.02)(0.96)}=0.605$

10D. P(has disease|negative reading) =
$\frac{P(neg\,|\,hasdisease)*P(hasd)}{P(negative)}=\frac{(.04)(.02)}{.04(.02)+.98(.96)}=0.001$

11A. Using the calculator: 4.33

11B. Using the calculator: 0.51

11C: Using the calculator: -4.21

11D: Using the calculator: 0.51

12A. Notice first, that for any probability density function, the area enclosed by the curve and the x-axis must be 1.
The function f(x) = q, is just a rectangle with a length of 5 – (-1) = 6 and a height of q. So the area is $6q = 1 \Rightarrow q = \frac{1}{6}$.

12B. Notice that the graph between the given values and the x-axis is a trapezoid. The base lengths are $\frac{1}{6}$ and $3q + \frac{1}{6}$ and the height is 3-0=3. So the area is $\frac{1}{2}(3)(\frac{1}{6} + 3q + \frac{1}{6}) = 1 \Rightarrow q = \frac{1}{9}$

12C. This is the graph of a semicircle with radius of q. So the area is $\frac{1}{2}(\pi q^2) = 1 \Rightarrow q = \sqrt{\frac{2}{\pi}} \Rightarrow q = \frac{\sqrt{2\pi}}{\pi}$

12D. Notice that this graph is half of an ellipse with semimajor axis q and semiminor axis 1. So the area is $\frac{1}{2}(1)(q)(\pi) = 1 \Rightarrow q = \frac{2}{\pi}$

13A. Mode by definition (and highest point)

13B. Median, by definition (and equal area)

13C. Mean, by definition (and center of gravity)

13D. Right Skewed

14A. P(rain and wind) = $(\frac{2}{3})(\frac{3}{5}) = \frac{2}{5}$

14B. P(rain or wind) = P(rain) + P(wind) – P(rain and wind) = $\frac{2}{3} + \frac{3}{5} - \frac{2}{5} = \frac{13}{15}$

14C. P(at least one ran or wind) = 1-P(no rain and no wind) = $1 - (\frac{1}{3})(\frac{2}{5}) = \frac{13}{15}$

14D. P(rain at least 3 days) = ${}_5C_3(\frac{2}{3})^3(\frac{1}{3})^2 + {}_5C_4(\frac{2}{3})^4(\frac{1}{3}) + {}_5C_5(\frac{2}{3})^5 = \frac{64}{81}$

15A. Mean $= \frac{4+7+0+3+-3+5+5+5+5+7+3+7+9}{13} = \frac{57}{13}$

15B. Mode = most often = 5

15C. Arrange the values in ascending order: -3, 0, 3, 3, 4, 5, 5, 5, 5, 7, 7, 7, 9. From this the median is 5. From here we can find that Q_1=3 and Q_3=7. So IQR = 7-3=4

15D. Range = biggest – smallest = 9 – (-3) = 12