No Calculator Middleton High School Invitational – February 18, 2006 – Statistics Competition

Written by Jorge Uy- Middleton salutatorian 2005 currently attending University of "Go Gators" Florida

1. In Ms. Woolfenden's class of AP Statistics students, a test was conducted to determine the distribution of student GPA's. After a nerve-racking month of collecting and analyzing data, the test showed the student GPA's to be skewed to the left, with mean 2.40 and mode 2.96. Decide upon which value would most nearly fit the median GPA.

A. 2.14 B. 2.38 C. 3.23 D. 2.86 E. NOTA

2. The least squares regression line of y on x was determined to be $\hat{y} = mx + b$, where m and b are represented as constants. The coefficient of determination is equal to 0.25 and $\frac{s_x}{s_y} = \frac{1}{2}$. If the least squares regression line passes through $\overline{X} = 2$ and $\overline{Y} = 2$, find the value of $m^2 + b^2$. Round answers to its thousandths position.

A. 3.078 B. 2.313 C. 1.250 D. 1.000 E. NOTA

3. If variables U and F are independent and given that $p(U) = 1/3$ and $p(F) = 1/6$ determine the probability value of p(U or F).

A. $\frac{1}{2}$ B. $\frac{1}{3}$ C. $\frac{4}{9}$ D. $\frac{5}{9}$ E. NOTA

4. In a test conducted by the WhizKids of Middleton, they unbiasly rolled two, fair six-sided die and found the probability that their sum would equal a number greater than or equal to 7. Due to some outer forces distracting the concentration of the WhizKids though, they fear that their calculation might have been flawed. Determine the correct value for this test.

A. $\frac{5}{12}$ B. $\frac{7}{12}$ C. $\frac{1}{2}$ D. $\frac{1}{4}$ E. NOTA

5. In the following table, which shows the gender breakdown of students liking Inuyasha and Naruto for the cartoon competition bragging rights, what value of n would result in the table indicating a perfectly independent relationship between gender and cartoon?

	Inuyahsa	Naruto
Boys	20	n
Girls	8	16

A. 7 B. 10 C. 12 D. 40 E. NOTA

6. Which of the following is a characteristic of a binomial distribution, but not of a geometric distribution?

   A. There is a fixed number of observations.
   B. The observations are all independent.
   C. Each observation has only two outcomes: "success" or "failure."
   D. The probability of success is the same for each observation.
   E. NOTA

7. In a coloring contest held in Mr. Holt's classroom, students must randomly choose colors out of a crayon box to begin the contest. The box contains 9 red, 8 blue, 7 neon, 6 purple, 6 turquoise, 4 fuchsia, and 3 yellow. If two of the colors are randomly selected, determine the complement of the probability of attaining first a neon, then a fuchsia. (Assume that crayons are not replaced)

A. $\frac{2}{129}$ B. $\frac{6}{301}$ C. $\frac{28}{1849}$ D. $\frac{127}{129}$ E. NOTA

8. In a graph comparing student's head size to the level of math that they're taking, the correlation between them resulted r = 0.15. If 0.5 is added to the student's head size and then multiplied by 3, what would the new correlation be equal to? Round answer to the nearest hundredths position.

A. 0.15 B. 0.84 C. 0.95 D. 1.95 E. NOTA

9. Suppose P(passing the AP Stat exam) = 1/5 and P(passing the AP class) = 2/5. If

   P(passing the AP Stat exam| passing the AP class) = 3/25, what is

   P(passing the AP class| passing the AP Stat exam)?

   A. $\frac{1}{5}$ B. $\frac{2}{5}$ C. $\frac{3}{50}$ D. $\frac{6}{25}$ E. NOTA

10. Given a normal distribution centered at 11 with variance of 2.25, which of the following approximates the Z-score associated with the data point 12.25 to the nearest hundredth?

    A. -0.83 B. -0.56 C. 0.83 D. 5.00 E. NOTA

11. Which of the following correctly describes the concept of a Type II error?

    A. Failure to reject a true null hypothesis.
    B. Rejecting a true null hypothesis.
    C. Failing to reject a false null hypothesis.
    D. Rejecting a false null hypothesis.
    E. NOTA

12. A continuous random variable, Z, is uniformly distributed between the values of 0 and 30, inclusively.

    What is the probability that Z > 25?

    A. $\frac{1}{6}$ B. $\frac{1}{4}$ C. $\frac{3}{4}$ D. $\frac{5}{6}$ E. NOTA

13. A statistical observation of the Middleton's superheroes, The Woolfenator and The Super Spasmodic Squirrel, provided that their success rate formed a geometric distribution with probability of success 1/5. What is the probability that it takes more than 5 battles (trials) to see their first victory (success)?

    A. $\frac{1}{5^4}$ B. $\frac{1}{5^5}$ C. $\frac{4^4}{5^4}$ D. $\frac{4^5}{5^5}$ E. NOTA

.

14. Refer to #13, what is the probability that the first success will come on the 5th battle (trial)?

A. $\frac{4}{5^4}$ B. $\frac{4}{5^5}$ C. $\frac{4^4}{5^5}$ D. $\frac{4^5}{5^5}$ E. NOTA

15. Please refer back to Question 13. If the distribution of Middleton's superheroes was limited to 5 battles (trials), what is the probability that they win exactly four battles (successes)?

A. $\frac{4}{5^4}$ B. $\frac{4}{5^5}$ C. $\frac{4^4}{5^5}$ D. $\frac{4^5}{5^5}$ E. NOTA

16. Which of the following is **<u>NOT</u>** true for chi-square distributions?

A. Each Chi-squared is skewed to the left.
B. Total area under the curve is 1.
C. The more degrees of freedom a chi-square has, the more normal it appears.
D. All chi-squares decrease asymptotically to the X-axis as x-values get large.
E. NOTA

17. The random variable J has a mean of 3.1 and a standard deviation of 3. The random variable U has a mean of 2.9 and a standard deviation of 4. Assume the correlation between J and U is 1. Determine the mean and standard deviation when summing J and U together. Answers will be written in the form: (mean, standard deviation)

A. (3, 5) B. (3, 7) C. (6, 5) D. (6, 7) E. NOTA

18. In an odd third dimensional realm, the length of ET-like creature's fingers has a standard deviation of 1.5 meter. What sample size would be needed in order for us to be 95.44% confident of knowing that the length of their fingers is within ± 1 meter?

A. 3 B. 7 C. 8 D. 9 E. NOTA

Use the following chart for questions 19 to 22. Note that this information is not fact, but made up for the simplicity of this problem.

In a series of tests, conducted by the Mueller is Weiss Company, to determine the number of cows that are infected with "Mad Cow Disease", they took a sample size of 100 cows from each of the states, California, Florida, Georgia, and New York, and determined how many cows were infected with the disease. The following table shows the results of their findings.

	California	Florida	Georgia	New York
Yes	24	56	10	36
No	76	44	90	64

19. Determine the probability of cows infected with "Mad Cow Disease" from the entire sample size.

A. $\frac{137}{200}$ B. $\frac{6}{25}$ C. $\frac{8}{25}$ D. $\frac{63}{200}$ E. NOTA

20. In carrying out a χ^2 test, determine the expected value of cows not having the disease and being from Florida.

A. 63/2 B. 137/2 C. 44 D. 100 E. NOTA

21. Out of all the cows that do not have the disease, what is the conditional probability that they are from Florida?

A. $\frac{4}{9}$ B. $\frac{22}{137}$ C. $\frac{11}{100}$ D. $\frac{11}{25}$ E. NOTA

22. How many degrees of freedom will there be if you carried out a χ^2 test?

A. 3 B. 6 C. 8 D. 16 E. NOTA

23. If variables P and G are independent, what is the probability of P given G, if P(P) =0.2 and P(G) =0.7?

A. 0.14 B. 0.2 C. 0.7 D. 0.9 E. NOTA

24. In a Calculus test administered by the infamous Pam Allison, Jorge scored a 96% and William scored an 89%. In a class of 8 students, if the average scored on the test was a 64%, what is the average of the remainder of the class, excluding Jorge and William? Round to the nearest percentage.

A. 33 B. 36 C. 54 D. 55 E. NOTA

25. What is the definition of power?

A. The probability that we will reject the false null hypothesis.
B. The probability that we will reject the false alternate hypothesis.
C. The probability that we will accept the false null hypothesis.
D. The probability that we will accept the false alternate hypothesis.
E. NOTA

26. If statisticians were given a 99% confidence intervals based on a sample size N, what would be a correct interpretation of this result?

A. 99% of the sample means of size N will fall within this interval.
B. 99% of the sample means of size $\sqrt{N}$ will fall within this interval.
C. This sampling process is 99% effective in finding the true population mean.
D. If this sampling process is repeated many times, the resulting confidence intervals will contain the population mean 99% of the time.
E. NOTA

27. How many of the following types of measurements will yield a result within the interval $[0, \infty)$?

I. Correlation II. Coefficient of Determination III. Standard Deviation

IV. Variance V. Chi-Squared Value VI. Probability

A. 2 B. 3 C. 4 D. 5 E. NOTA

28. A really smart mathematician conducts a linear regression procedure on a data set X, where it measures the size of human brains to determine Y, the size of the person's head. If the least-squares regression procedure yields X = 1.5Y + 2, what is the correct interpretation of the slope?

A. For every 1 unit decrease in Y, X decreases by 1.5 units.
B. For every 1 unit increase in Y, X increases by 1.5 units
C. For every 1 additional unit of a person's brain size, the person's head size is predicted to increase by 1.5 units.
D. For every 1 additional unit of a person's brain size, the person's head size is predicted to increase by 2/3 unit.
E. NOTA

29. Pedro decides to hit the slots, to test if lady luck's with him. He sits down on a $5 slot machine and the expected net payouts are as follows: winning $100 with a probability of 0.01, winning $20 with a probability of 0.05, and losing $1 with a probability of 0.5 (It's a special type of slot machine.). Assuming that the remaining probability values result in him losing only the original $5, determine the expected net value payout.

A. -3.5 B. -0.7 C. 1.5 D. 2.5 E. NOTA

30. Assume a statistical test yields a p-value < α, where α is a predetermined statistical significance level. Which of the following is a statistically correct conclusion of the result?

A. Reject the null hypothesis.
B. Fail to reject the null hypothesis.
C. Reject the alternate hypothesis.
D. Fail to reject the alternate hypothesis.
E. NOTA

Attach Z table

SOLUTIONS FOR STATISTICS

1. Answer must lie between 2.4 and 2.96. Therefore, the answer is **D**.
2. $r\dfrac{s_y}{s_x} = m = 0.5(2) = 1$; Therefore, y=x + b; 2=2+b; b=0. $1^2 + 0^2 = 1$ **D.**
3. Since the variables aren't disjoint,

   P(U or F) = p(U) + p(F) – p(U and F) = $\dfrac{4}{9}$ **C**
4. The table below shows the possible outcomes in summing up the two die. From counting the bold lettered numbers, we can see that 21/36 or 7/12, of the outcomes result in a sum of 7 or greater. Therefore **B**.

2	3	4	5	6	**7**
3	4	5	6	**7**	**8**
4	5	6	**7**	**8**	**9**
5	6	**7**	**8**	**9**	**10**
6	**7**	**8**	**9**	**10**	**11**
7	**8**	**9**	**10**	**11**	**12**

5. $\dfrac{20\times16}{8} = 40 = n$ **D.**
6. Definition of Binomial Distribution. **A.**
7. The complement of an answer is equal to 1 subtract the answer, which is equal to: $1-\left(\dfrac{7}{43}\times\dfrac{4}{42}\right)$ **D.**
8. Correlation coefficient is not affect by addition and multiplication. Therefore, the correlation coefficient remains the same. **A.**
9. p(passing the AP Stat exam and passing the AP class) = 3/25 x 2/5 = 6/125

   $$\text{p(passing the AP class| passing the AP Stat exam)} = \frac{\text{p(passing the AP Stat exam and passing the AP class)}}{p\text{(passing the AP Stat exam)}} = \frac{\frac{6}{125}}{\frac{1}{5}} = \frac{6}{25}$$ **D.**
10. $\dfrac{12.25-11}{1.5} = 0.83$ **C.**
11. Definition of Type II error. **C.**
12. $1-\dfrac{25}{30} = \dfrac{1}{6}$ **A.**
13. $p(x>5) = \left(1-\dfrac{1}{5}\right)^5 = \dfrac{4^5}{5^5}$ **D.**
14. $p(x=5) = \left(1-\dfrac{1}{5}\right)^4 \times \dfrac{1}{5} = \dfrac{4^4}{5^5}$ **C.**
15. $(5C4)x\left(\dfrac{1}{5}\right)^4\left(\dfrac{4}{5}\right)^1 = \dfrac{4}{5^4}$ **A.**
16. Chi-Squared is always skewed to the right. **A.**

17. The mean is equal to 3.1 +2.9 = 6. Since the two variables are not independent, the standard deviation i equal to $\sigma^2_{J+U} = 3^2 + 4^2 + 2(1)(3)(4)$. $\sigma_{J+U} = 7$. **D.**
18. $Z^*\left(\frac{\sigma_X}{\sqrt{n}}\right) = 2\left(\frac{1.5}{\sqrt{n}}\right) = \pm 1, n = 9$ **D.**
19. Definition of marginal distribution. $\frac{sumof\ "yes"}{overalltotal} = \frac{126}{400} = \frac{63}{200}$ **D.**
20. The expected value is equal to $\frac{rowtotal \times columntotal}{tabletotal}$ which yields 137/2. **B.**
21. $\frac{44}{76+44+90+64} = \frac{22}{137}$ **B.**
22. Degrees of freedom is equal to (r-1)(c-1) = (2-1)(4-1) = 3 **A.**
23. If P and G are independent, p(P|G) = p(P) = 0.2 **B.**
24. ((0.64x8)-0.96 – 0.89 = 3.27/6 = 0.545 = 55% **D.**
25. Definition of power. **A.**
26. Definition of a confidence interval. **D.**
27. II, III, IV, V,and VI are the only ones that will yield a result in the interval $[0, \infty)$. **D.**
28. If we rewrite the equal to where we solve for Y, we can conclude that D. is the correct answer. **D.**
29. 100(0.01) + 20(0.05) – 1(0.5) – 5(0.44) = -.7 **B.**
30. Definition of significance level. **A.**

Statistics Team Round #1 Middleton Invitational February 18, 2006

Researchers at The Wind&Rain Channel believes there is a relationship between the latitude measure of a city (independent variable) and its average annual rainfall in inches. To test their assumption they gathered data from 20 weather stations in California. The data for latitude measurements covers the range [30,45]. The regression line generated is $\hat{y} = 5 + 0.7x$

SUMMARY OUTPUT

Regression Statistics	
R	0.83
R Square	0.689
Standard Error	13.0
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2,664.89	2,664.89	13.96	0.000849528
Residual	18	5,346.77	190.96		
Total	19	8,011.65			

	Coefficients	*Standard Error*
Intercept	5	1.5
X Variable	0.7	0.3

Part (A): Give an interpretation for your estimate of $\hat{\beta}_0$ (the constant). Are you confident with this interpretation, why or why not?

Part (B): If appropriate, give an interpretation for your estimate of $\hat{\beta}_1$. If not appropriate, explain why.

Part (C): Interpret the sign of the correlation coefficient.

Part (D): What percentage of variation in the average rainfall can be explained by the variation in latitudes?

Part (E): What would the value of the correlation coefficient be if you measured the average rainfall variable in centimeters instead of inches?

Part (F): Based on the given data, is it reasonable to attempt to predict the average rainfall of a station that has latitude of 24? Why or why not? If it is reasonable, perform the calculation and in addition to your explanation, provide the predicted value.

Part (G): Based on the given data, is it reasonable to attempt to predict the average rainfall of a station that has latitude of 40? Why or why not? If it is reasonable, perform the calculation and in addition to your explanation, provide the predicted value.

Part (H): Based on the given data, is it reasonable to attempt to predict the average rainfall of a station that has latitude of 38 but is located 290 miles east of California in the Nevada Desert (longitude of 100^o west)? Why or why not? If it is reasonable, perform the calculation and in addition to your explanation, provide the predicted value.

Part (I): Suppose the residual plot is that of figure 1. What can you conclude?

Part (J) Suppose the residual plot is that of figure 2. What can you conclude?

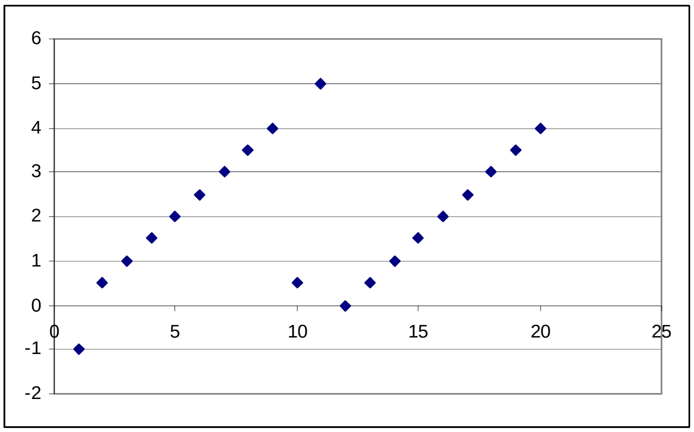
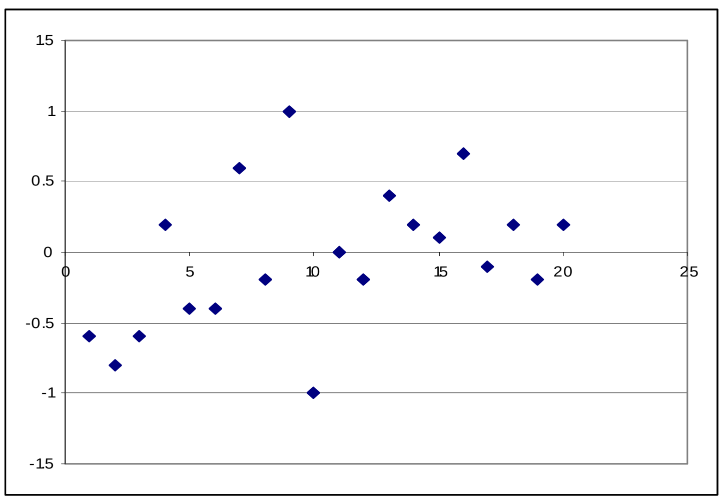


Figure 1

Figure 2



Statistics Team Round #2 Middleton Invitational February 18, 2006

At the University of Ol' Yonder, investigators are looking into the effects of age and gender on short-term memory for adults. Each in a sample of 193 adults were asked to look at a list of 25 words and remember as many as they could. After a social hour and recreation time, they were asked to recall as many of the 25 words as they could. The number of words remembered (out of 25) was recorded.

Part (A): The investigators believe that women have a higher mean memory score than men. Set up the null and alternative hypothesis to test this idea (specify if this is a one or two sided test). Use the symbols μ_m and μ_f to denote mean memory scores for men and women.

Part (B): The 50 men in the study had a mean memory score of 20 with variance of 10 and the 150 women in the study had a mean memory score of 21 with variance 7.5. Using the information from part (A) and the given information in part (B), conduct the test. Provide the value of your test statistic, the critical value for comparison, and CLEARLY state your conclusion, write out an explanation. Also, $\alpha = 0.05$.

Part (C): Continuing this case, researchers want to construct an α% confidence interval for the difference between male and female memory scores (μ_m - μ_f). Report your interval as [Lower Bound , Upper Bound] NOT Point Estimate ± Variance. (Note: You should use large sample procedures). Round your bounds to three decimal points. Use the appropriate α so that the accompanying T multiplier in your confidence interval equation is 2.0 (e.g. $D \pm (2.0)SE$.

Part (D): Fill in the blank with the choices in parenthesis. Statistical tests that yield a p-value less than alpha lead to _________ of the null hypothesis (failing to reject , rejection).

Part (E): Fill in the blank with the choices in parenthesis. Statistical tests that yield a test statistic less than the critical value lead to _____________ of the null hypothesis (failing to reject , rejection).

A) Suppose that you want to study whether an athletic training program actually helps the athletes score better times in track & field events, so you gather data on a random sample of athletes who attended the training program. Suppose that you find that 95% of the sample scored better in the events after attending the program than before attending the program. Moreover, suppose that you calculate that the sample mean of the improvements was a substantial 21.4 seconds.

- i) Explain why you cannot legitimately conclude that the athletic training program *caused* these students to improve their scores. Suggest some other explanations for the improvement.
- ii) Identify the explanatory and response variables in the training study.
- iii) Is the athletic training program as described above a controlled experiment or an observational study/survey? Justify your answer.

B) Suppose that you want to study whether clowns provide a therapeutic relief to children in the hospital. Furthermore, you decide to investigate whether burn victims who are visited by a clown tend to recover faster or more often than those who do not. You randomly select a sample of burn victims from a large urban hospital and follow them for one year. You then compare the proportions and find that 92% of those with clown visits made a full recovery while only 64% of those without clown visits made a similar full recovery.

- i) Identify the explanatory and response variables.
- ii) Is this an experiment or an observational study? Justify your answers.
- iii) What are some of the critical flaws in this design?
- iv) How would you design this study to investigate the proposition that patients who get a clown visit receive a therapeutic effect for burn victims?

C) A study is being performed to investigate whether or not the drug XYZ reduces the risk of a person passing on meningitis to their spouse by being contagious.

Note: Parts (ii), (iii), and (iv) are independent. Thus part (iii) is NOT to take your answer from part (ii) and redesign the study to include both principles, rather just the one referenced in part (iii), etc.

- i) Identify the explanatory and response variables in this study. Are these measurement or categorical variables?
- ii) Explain how the study should be designed to incorporate the principle of *randomization*.
- iii) Explain how the study should be designed to incorporate the principle of *comparison*.
- iv) Explain how the study should be designed to incorporate the principle of *blinding*.

A teacher for a microbiology class provides a summary for grades on her last exam. They are as follows:

GRADES SUMMARY STATS >>> DESCRIPTIVE >>> MINITAB OUTPUT	
N = 28	SEMEAN = 2.38
MEAN = 74.71	MIN = 35
MEDIAN = 76.00	MAX = 94
TRMEAN = 75.50	Q1 = 68
STDEV = 12.61	Q3 = 84

Fortunately for you (unfortunately for others) she left her grade book open and you saw the lowest grades were 35, 57, 59, 60 --- but you couldn't see any others. Nevertheless, a knowledgeable user of statistics can tell a lot about the data set simply by studying the set of descriptive statistics above.

(A) Write a brief description about the distribution of grades based on the summary statistics in the box above. Be sure to address:
  a. The general shape of the distribution.
  b. Unusual features, including possible outliers.
  c. The middle 50% of the data.

(B) Describe in detail a procedure for outlier detection and perform it on the data. Identify any outliers (specifically if you know the exact grades or whether there are any on the high side of the distribution.

(C) Construct a modified box-and-whisker plot for the test grades.

(D) Using your procedure in part (B), what percentage of the standard normal distribution would your procedure flag as "potential outliers"? (round to tenths place)

Statistics Team Round #5 Middleton Invitational February 18, 2006

A study is conducted to find out the relationship between people claiming to be religious and their college education. The following frequency table is found.

		Religious? Yes	Religious? No
College Education?	Yes	12	8
	No	4	6

Carry out a chi squared test to determine whether there is sufficient evidence to indicate that religiousness is dependent of getting a college education. Use $\alpha = 0.05$. You must show your matrix of expected values, the test statistic, and clearly state your conclusion.

SOLUTIONS STATISTICS TEAM MIDDLETON 2-18-06

Question #1 Linear Regression

Part (A) IF the latitude is zero the expected annual rainfall is 5 inches. NO, in this case it would mean a point with latitude zero, which is not applicable given the data set.
Part (B) For every one point (or degree) increase in latitude we expect an additional 0.7 inches of rain.
Part (C) The correlation coefficient means that latitude and rainfall (over this interval) have a positive association (one increases and the other increase).
Part (D) 68.9%
Part (E) 0.83. Units of measure do not change "r"
Part (F) No. 24 is outside the domain of the existing data, [30,45] which requires extrapolation, which is not appropriate here given other environmental factors.
Part (G) Yes. 33. 5+.7(40)
Part (H) No. 38 is within the literal domain of the data, but is not really within the practical scope of the data. An observant statistician would recognize that and exclude that point because of the different environmental conditions.
Part (I) A linear regression is not appropriate for this data given the pattern in the residuals.
Part (J): A linear regression is appropriate given the random layout of the residuals.

Question #2 Two Sample Mean Analysis

Part (A): Ho: $\mu_f = \mu_m$ and Ha: $\mu_f > \mu_m$

---OR---

Ho: $\mu_f - \mu_m \le 0$ and Ha: $\mu_f - \mu_m > 0$

This is a one sided test

Part (B): Test statistic: $z = \dfrac{(21-20)-0}{\sqrt{\dfrac{10}{50}+\dfrac{7.5}{150}}} = 2 >$ T-CV = 1.645 (one sided test). Thus we reject Ho. We conclude that we can reject the idea that $\mu_f \le \mu_m$.

Note: Also can be stated as there is evidence that females have better short term memory.

Part (C): $(20-21) \pm 2(.5)$ = [-2,0]

Part (D) Rejection

Part (E) Failing to reject.

Part (A)

Part (i) Causation is different from correlation. Other variables (outside training, weather conditions, team morale, etc) could have contributed to the change. These variables could be confounding the conclusion of causation, thus it is not appropriate. Also you could have mentioned that causation cannot be concluded since it is an observational rather than a controlled experiment.

Part (ii) Explanatory --- Participation in the training program. Response --- athletic performance (time).

Part (iii) Observational study. No treatment was imposed OR The researcher did not isolate the variables of interest and have a randomly selected control group.

Part (B)

Part (i) Explanatory --- Status of clown visit (happened or not). Response --- Full recovery or not.

Part (ii) Observational Study. There was not a control group, thus it cannot be an experiment. To be an experiment, you would have to randomly assign the subjects to a group (control or experimental) and allow visits to the experimental group and withhold them from the control group. Here there was no control group or control over the visit schedule or no treatment imposed is acceptable.

Part (iii) Failure to control the clowns visit schedule. Failure to stratify the subjects by initial injury severity (could have a large effect on recovery status) OR other reasonable confounding variables.

Part (iv) The design should contain two critical features: 1) a control group that does not receive visits and an experimental group that does and 2) randomization of subject assignment to the groups. Also, stratification of subjects by initial injuries would be acceptable.

Part (C)

Part (i) Explanatory --- Use of drug XYZ. Response --- Meningitis transmission. Explanatory variable is categorical (taking or not taking) and the response variable is also categorical (transmitted or not).

Part (ii) Randomization could be incorporated by assigning patients randomly to a control/experimental group and then testing their transmission rate. Any valid explanation here for random assignment is acceptable.

Part (iii) Comparison could be achieved by doing a paired experiment where subjects are paired based on similar characteristics (weight, height, gender, age, health history) and the only non-similar variable is drug XYZ consumption, then compare the rates between the paired subjects. OR placebo as the second group is acceptable.

Part (iv) Blinding could be achieved by the subjects (or drug administrators) not being aware of which subjects were or were not being given the placebo and the real drug. This ensures no bias among participants.

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

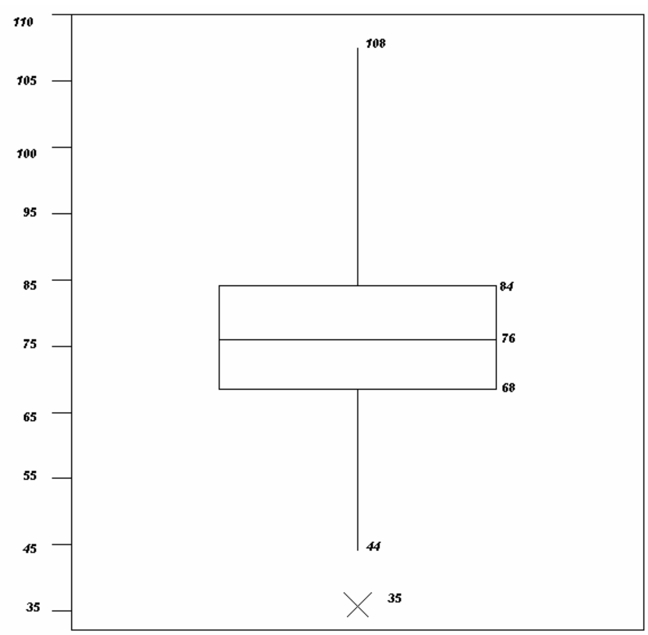
Question #4 Distributions

Part (A) The distribution exhibits properties of approximate normality with a slight left skew. Mean < Median reveals the left skew property whereas the fact that Q3 – Median = Median – Q1 shows the same percentage of data points in the quartiles over the same ranges. The IQR test reveals the outlier score of 35. The middle 50% of the data is contained between 68 and 84.

Part (B) Construct the IQR. Q3 – Q1 = 84-68 = 16. Q3 + 1.5(IQR) = 108 and Q1 – 1.5(IQR) = 44. Anything not between 44 and 108 would be considered an outlier. Thus here, the score of 35 is an outlier.

Part (C) The box plot should have box borders at 68 and 84 with center line at 76. The whiskers should extend to 94 and 57. The outlier at 35 is shown with an "X".

Part (D) For N(0,1) distribution, Q1= -0.675 and Q3 = 0.675. IQR = 1.35. Thus the bounds for Outliers is +/- 2.7. This translates to (from the table) Approximately 0.7% or 0.007.



Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

Question #5 Data Fit

Null: College education and religiousness is independent of each other

Alternative: College education and religiousness are dependent of each other

Expected values:

$10\frac{2}{3}$	$9\frac{1}{3}$
$5\frac{1}{3}$	$4\frac{2}{3}$

Test statistic: $\frac{\left(12-10\frac{2}{3}\right)^2}{10\frac{2}{3}}+\frac{\left(8-9\frac{1}{3}\right)^2}{9\frac{1}{3}}+\frac{\left(4-5\frac{1}{3}\right)^2}{5\frac{1}{3}}+\frac{\left(6-4\frac{2}{3}\right)^2}{4\frac{2}{3}}=\frac{15}{14}<\chi^2_{CV}=42.5.$

Thus we fail to reject Ho. We cannot conclude dependency.

attach Z,T and chi 2

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

School ______________________________ ID ______________

#1 A

#1 B

#1 C

#1 D

#1 E

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

#1 F

#1 G

#1 H

#1 I

#1 J

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team
School ______________________________ ID ______________

2 A

#2 B

#2 C

#2 D

#2 E

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

School ________________________________ ID _______________

#3 A *i)*

#3 A *ii)*

#3A *iii)*

#3 B *i)*

#3 B *ii)*

#3 B *iii)*

#3 B *iiii)*

#3 B *iiii)*

#3 C *i)*

#3 C *ii)*

#3 C *iii)*

#3 C *iiii)*

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

School ______________________________ ID ______________

#4 A

#4 B

#4 C

Middleton Invitational 2-18-2006 NO CALCULATOR Statistics Team

School ______________________________ ID ______________

#5