**Problem Set #1**

**Geog 2000: Introduction to Geographic Statistics**

**Instructor: Dr. Paul C. Sutton**

**Do it By Hand Exercises (i.e. Don’t Use A Computer)**

**#1) Strange Numbers from What or Why or Where or How?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Mean*** |  | ***Median*** |  | ***Mode*** |  |
| ***Range*** |  | ***Inter-Quartile Range*** |  | ***Shortest Half*** |  |
| ***Maximum*** |  | ***Minimum*** |  | ***Outliers*** |  |

**B)** Draw a histogram, a box-plot, and a stem-and-leaf plot for these numbers. Explain how the Stem and Leaf plot does not really improve on the histogram with respect to characterizing these numbers in this particular example but might in another dataset.

|  |  |
| --- | --- |
| ***Histogram*** |  |
| ***Box-plot*** |  |
| ***Stem and Leaf plot*** |  |
| ***Explanation*** |  |

**C)** Write a short 2-4 sentence explanation for each of the following random processes as to why each of them **do** or **do not** represent a *reasonable* and *probable* manifestation of the measurements represented by these numbers. ( I expect answers to vary ☺)

|  |
| --- |
| 1. ***The number of sexual partners of a 16 randomly selected 20 year old men in New York City in 1995.***
 |
|  |
| 1. ***The number of corporate logos (e.g. the McDonald’s ‘M’, The Nike ‘Swoosh’, British Petroleum’s ‘BP’) recognized from a presentation of 1000 logos to a randomly selected set of 16 sixth grade students from Boston MA.***
 |
|  |
| 1. ***The length in inches of Trout caught in a Fishing Derby in Lake Tahoe CA.***
 |
|  |
| 1. ***The actual Fertility rate (the number of live births) of 16 randomly selected American Women of the age of 50 in 2007 (Actual TFR then was 2.05).***
 |
|  |
| 1. ***The net worth (in millions of dollars) of sixteen people who paid $50 to have their photograph taken at the top of a ski lift in Aspen Colorado in 2001.***
 |
|  |

**#2) Fuzzy Dice from Hell and their Probability Density Functions**

**A)**

|  |  |
| --- | --- |
| ***What is the average result you would expect from the ‘sum’ value of the rolling of these two dice?*** |  |
| ***What is the range of possible outcomes from the sum of the rolling of these two dice?*** |  |
| ***What is a more likely outcome of the sum of a single roll of these two dice: Sum of Dice = 2 or Sum of Dice = 5? Explain.*** |  |

**B) Probability Density Function (Probability Mass Function)**

|  |
| --- |
|  |

**C)**

|  |  |
| --- | --- |
| ***1) Are you wise to play this game of chance?***  |  |
| ***2) If you did choose to play this game of chance how much would you expect to win or lose assuming this ‘game of chance’ worked out perfectly randomly?*** |  |

**D)**

|  |
| --- |
| ***Suppose you try to cheat on your roommate by sneaking in a ‘bogus’ die with ‘twos’ on every side. Did you improve your odds of winning? Explain.*** |
|  |

**E)**

|  |
| --- |
| ***What is the average value of the roll of one of these four sided die?***  |
|  |
| ***How might you do financially if you suggested to your roommate that she roll a SINGLE four sided die and you pay her $100 dollars every time she rolled the average value and you collected a mere twenty five cents every time she did not? Explain.*** |
|  |

**#3) The odds and oddity of heirs**

|  |
| --- |
| ***Assume that the probability of having a baby boy or a baby girl are equal and 50%.******Assume a woman has two children. What are her chances of having at least one female child? Explain.*** |
|  |

 **#4) Sometimes it is easier to think about what couldn’t happen.**

Assume you have been randomly selected from thousands of college age students to participate in an ‘American Idol’ competition. Also assume that there is an equal probability of all of these college age students having a birthday on any day of the year (No Leap Years – we’re not going there – e.g. 1 in 365 days of the year).

|  |
| --- |
| 1. ***How many students would have to be selected for which the probability of you having the same birthday as one of the other contestants would be a 50% chance? (Note: You can use a calculator or spreadsheet for this question – it does involve significant sequential arithmetic).***
 |
|  |
| 1. Assume the song you have decided to perform for this ‘American Idol’ competition is “Blue Valentines” by Tom Waits – What are the odds that out of 1,000 other contestants one or more of them would choose the same song?
 |
|  |
| 1. Characterize the probabilistic approach you took to parts ‘A’ and ‘B’ of this

Question as classical, relative frequency, monte carlo, or subjective. Explain. |
|  |

**#5) Conditional Probability can be REALLY tough to recognize: “Let’s Make a Deal”.**

|  |
| --- |
| Switch or Stick or does it not matter? Explain using what you have learned about probability.  |
|  |

**#6) Probabilistic Approaches**

|  |  |
| --- | --- |
| ***Classical*** |  |
| ***Relative Frequency*** |  |
| ***Subjective*** |  |
| ***Monte Carlo*** |  |

**#7) Will the Real Independent Random Variable Please step forward?**

|  |
| --- |
| ***What does it mean for observations of a random variable to be ‘Independent’?*** |
|  |
| ***Provide an example of independent observations and an example in which***  ***Observations are not independent.*** |
|  |
| ***Why is geography (and especially spatial Analyses) often criticized for using regular statistics on non-independent Measurements?***  |
|  |

**Computer Problems (use JMP and/or Excel for these exercises)**

#8) Generate two Normal distributions identical to those described above and

 Create a single column with all of the observations (e.g. 2000 from N(70,3) and 2000 from a N(7, 0.03). Plot them all on one histogram. What do you see?

|  |  |
| --- | --- |
| ***Histogram*** |  |
| ***What do you see?*** |  |

#9)

|  |
| --- |
| ***(Less than) One page Summary of the Uniform Distribution*** |
|  |

#10)

|  |
| --- |
| ***(Less than) One Page Summary of the Binomial Distribution*** |
|  |

#11)

|  |
| --- |
| ***(Less than) One Page Summary of the Exponential Distribution*** |
|  |

#12)

|  |
| --- |
| ***One Page Summary of the Poisson Distribution*** |
|  |

#13)

|  |
| --- |
| ***One Page Summary of the Chi-Square Distribution*** |
|  |

**How to Lie With Statistics Question**

|  |  |
| --- | --- |
| ***#14) Write a 4 to 7 sentence summary of Chapter 1: The Sample with built in Bias*** |  |
| ***#15) Write a 4 to 7 sentence summary of Chapter 2: The well chosen average*** |  |
| ***#16) Write a 4 to 7 sentence summary of Chapter 3: The little figures that are not there*** |  |
| ***#17) An apochryphal story about the department of Geography at the University of North Carolina (UNC) is that their undergraduate geography majors earn more money than the geography graduates of any other university in the country. It just so happens that Michael Jordan (an extremely well paid basketball player) got his degree from UNC in geography in 1986. Would Darrell Huff describe this as ‘Lying With Statistics’? If so, what category of your above essays would this ‘lie’ fall under. Explain.*** |  |