**Problem Set #2 Answer Sheet**

**Geog 2000: Introduction to Geographic Statistics**

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

**#1) The Standard Normal Distribution in all its glory**

|  |
| --- |
| **Make a detailed drawing of a Standard Normal (e.g. N(0,1)) probability density function (aka pdf), and answer the following questions:** |
|  |

|  |
| --- |
| 1. **What fraction of observations drawn from a N(0,1) distribution are larger than 1 standard deviation above the mean?** |
|  |
| 1. **What is the mathematical equation describing the N(0,1) pdf?** |
|  |
| **C) What is the total area under the N(0, 1) curve? (e.g.  int_(-infty)^inftyP(x)dx=1.  ) where P(x) is the equation you identified in part ‘B’?** |
|  |
| 1. **What fraction of observations drawn from a N(0,1) distribution would you expect to land anywhere from -1 to +1? (e.g. -1 <= X <= +1)** |
|  |
| 1. **What fraction of observations drawn from a N(0,1) distribution would you expect to land anywhere from -2 to +2? (e.g. -2 <= X <= +2); -3 to +3? (e.g. -3 <= X <= +3)** |
|  |
| 1. **Assume women’s heights are distributed N(5’ 7”, 3”). A particular woman has a height of 5’ 10”. What percentage of women are taller than this woman?** |
|  |
| **G) Suppose there are 1,000 trout in Lake Goober. Assume the lengths of these trout are distributed N(10”, 2”). How many fish would you expect that lake to have that are anywhere from 11” to 13” long?** |
|  |

**#2) Measures of variability and Z-scores oh my!**

**Given the following observations of a random experiment calculate the following:**

**3, 4, 4, 6, 7, 8, 8, 8, 9, 10, 10, 12, 21, 28, 35, 52**

|  |  |  |  |
| --- | --- | --- | --- |
| **A) Mean** |  | **D) Variance** |  |
| **B) Median** |  | **E) inter-quartile range** |  |
| **C) Standard Deviation** |  | **F) Mode** |  |

|  |
| --- |
| 1. **Draw a histogram of these data and comment on the pattern it reveals.** |
|  |
| 1. **How is the standard deviation different than the Mean Absolute Deviation? Define both the standard deviation and the mean absolute deviation and calculate each of them for this set of numbers.** |
|  |
| 1. **Calculate a z-score (e.g. a normalized value for each of the 16 numbers). Does it make sense to use z-scores for this particular distribution? Explain** |
|  |
| **J) Suggest a random experimental observation that might produce a set of numbers like this?** |
|  |

**3) Measurement Scales:  *“These Tennis Balls are MUCH Fuzzier”.***

**Define the four measurement scales most commonly encountered in geographic measurement (listed below as ‘A’, ‘B’, ‘C’, and ‘D’). Provide an example of the observations of a ‘random process’ for each of these:**

|  |  |
| --- | --- |
| A) Nominal |  |
| B) Ordinal |  |
| C) Interval |  |
| D) Ratio |  |

1. **Classify each of the following according to their measurement scale:**

|  |  |
| --- | --- |
| **1) Student course Evaluation Scores here at DU. (i.e. Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)** |  |
| **2) Sex (e.g. Male or Female)** |  |
| **3) Number of text messages sent from randomly selected cell phones** |  |
| **4) SAT scores of the graduating high school students of the year 2000** |  |

**#4) Tumbling Dice**

**Use the example of the rolling of two six sided dice to define the ideas of:**

|  |
| --- |
| 1. **A Random Experiment** |
|  |
| 1. **Elementary Outcomes** |
|  |
| 1. **Sample Space** |
|  |
| 1. **Draw a probability mass function for the random experiment of rolling two dice and summing their face up numbers** |
|  |
| 1. **What is the probability of at least one of the die showing a ‘6’ on its face in a single roll of two dice?** |
|  |

**#5) Sometimes extreme events are very important & the Normal Distribution is not.**

**Consider the following numbers observed by a hypothetical hydrologist for a stream he has been observing for 3 years. The numbers represent the maximum flow of water (in cubic meters per minute), at a given point at the lower end of the stream right after a rainstorm.**

**1.0, 1.0, 1.5. 1.6, 1.8, 2.0, 2.1, 2.1, 2.1, 2.5, 2.5, 2.6, 2.7, 2.7, 2.7, 3.0, 3.3, 3.5,**

**3.5, 3.6, 3.6, 3.7, 3.7, 3.8, 3.9, 4.1, 4.3, 4.9, 5.5, 6.3, 7.1, 7.5, 8.8, 10.3, 16.5, 59.6**

|  |
| --- |
| **A) Draw a histogram of these numbers (use JMP if you wish). Do these observations look normally distributed? How might the Normal distribution not be an appropriate model for these observations?** |
|  |
| **B) Suppose that sediment transport (e.g. kilograms of soil, gravel etc.) that are moved by these storms varies directly as a function of the flow of water raised to the 3rd power (e.g. a cubic function (i.e. Flow = 1 , sediment transport = 1; flow = 2, sediment transport = 8; flow = 3, sediment transport = 27). Calculate the predicted sediment transport for each of the storms and calculate what percentage of the sediment transport that took place over these three years was accomplished by the single largest storm. Paste a table in here** |
|  |
| **C) Assume these are reasonable numbers (as observations) and the stated relationship between stream flow and sediment transport is also correct. How important are extreme events? How well do you think the Normal distribution will serve as a model for these kind of hydrologic phenomena?** |
|  |

**#6) Shooting Craps - The odds are not that easy to calculate….**

|  |  |
| --- | --- |
| **A) In any case – summarize the rules of ‘craps’ right here.** |  |
| **B) What is the probability of ‘winning’ on your first roll? (e.g. a ‘7’ or an ‘11’)** |  |
| **C) What is the probability of ‘losing’ on your first roll?** |  |
| **D) Why is it difficult to calculate the ‘overall’ odds of winning at craps?** |  |

**#7) A Tale of two distributions**

|  |
| --- |
| **Use the normal and the binomial distribution to explain the difference between a discrete random variable and a continuous random variable. What are the parameters of these two distributions? Why does one have a probability distribution function whereas the other has a probability mass function? Describe the relationship between these two distributions? (Larry Gonick’s ‘Cartoon Guide to Statistics’ Should REALLY help ☺).** |
|  |

**#8) NASCAR, fishing derbies, and Random Variables**

|  |
| --- |
| **Suppose there is a big fishing derby for NASCAR fans in Kentucky. Lake Goober is stocked with trout whose lengths are distributed N(10, 3). Assume the IQ of the fishing derby participants is distributed N(80, 10). Let’s say we create a random variable that is simply the sum of the IQ of the fisherman and the length of the first fish he or she catches. How will that random variable be distributed?** |
|  |

**#9) Testing for rare diseases and the *False Positive Paradox***

**Suppose a rare disease infects one out of every 3,000 people in a population. And suppose that there is a good, but not perfect, test for this disease. If a person has the disease the test comes back positive 99 % of the time. On the other hand, the test also produces some *false positives*. About 1 % of the uninfected patients also test positive. Assume you just tested positive.**

|  |
| --- |
| **A) What are your chances of having this disease?** |
|  |
| **B) Also, assume all administrations of this test are independent of one another regardless of who is tested. What should you do if you test positive?** |
|  |

**#10) Pascal’s Triangle, Permutations, Combinations, and the Binomial Distribution**

|  |
| --- |
| **Draw the first few lines of Pascal’s triangle and explain the rule for drawing additional rows indefinitely. How is this triangle related to the Binomial distribution? What is the binomial distribution again (include formulas)? Define, and explain applications, of the formulas for both *permutations P(n,r)*  and *combinations* *C(n,r)* shown below:** |
|  |

**#11) Going from Univariate to Bivariate Characterization of Random Manifestations of distributions: Point patterns in ‘Spaaaaace’.**

**YOW! This is fun.**

|  |
| --- |
| **A) Plot N(100, 30) vs. N(100,30). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain.** |
|  |
| **B) Plot N(100, 30) vs. an Exponential (1). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain.** |
|  |
| **C) Plot a Uniform (100) vs. a N(100, 30). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain.** |
|  |
| **D) Plot a Gamma(7) vs. A Uniform (100). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain.** |
|  |
| **E) Plot a Uniform (100) vs. a Uniform (100). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain** |
|  |
| **F) Plot a N(100,30) vs. N(100,3). Describe the pattern you see and suggest a random phenomena that might produce such a pattern. Explain.** |
|  |

**#12) What is *‘Complete Spatial Randomness’* (aka CSR) for Point Patterns?**

|  |
| --- |
| 1. **Which of the spatial distributions you produced in Question #1 of the Computer problems do you think *most closely* represents ‘Complete Spatial Randomness’ as we might theoretically expect for the ‘x’, ‘y’ locations of meteorite impacts in the state of Colorado? Explain.** |
|  |

**#13) How do you randomly sample the earth’s surface?**

|  |
| --- |
| 1. **Suppose the team that is doing this work prepares the following “Random Sample” of the earth’s surface according to the following rule: We generate a Random Uniform (-90, +90) distribution of ‘Y’ or ‘Latitude’ Values and pair them randomly with a Random Uniform (-180, +180) set of ‘Longitude’ Values. This set of ‘Y’ (aka Latitude) and ‘X’ (aka Longitude) pairs will randomly sample the surface of the earth (assuming it’s a sphere which we know it is not but who cares?). Will this approach randomly sample the surface of the earth – assuming it is a sphere? Explain.** |
|  |
| 1. **Only if you think the approach presented above is NOT random – Propose a better method of randomly sampling the surface of the earth using simple and reasonable probability density functions. Explain.** |
|  |

**#14) Write a 4 to 7 sentence summary of Chapter 4: *Much ado about practically nothing***

|  |
| --- |
|  |

**#15) Write a 4 to 7 sentence summary of Chapter 5: *The Gee-Whiz Graph***

|  |
| --- |
|  |

**#17) Write a 4 to 7 sentence summary of Chapter 6: *The one-dimensional picture***

|  |
| --- |
|  |

**# 18) Suppose that just the other day the Dow Jones Industrial Average dropped by 300 points. Find the last month’s worth of Dow Jones Industrial Averages and plot them. Now, Prepare two graphs: 1) A graph where you are trying to scare people that you are showing the graph to that the Dow Jones is Crashing (e.g. the sky is falling); and 2) Where you are trying to assuage the fears of your investors and ‘minimize’ this 300 point drop in the Dow Jones Industrial Average. Explain.**

|  |
| --- |
|  |