Unit 1 – Review of PubHlth 540, *Introductory Biostatistics*

Practice Problems Week #1

Due: Monday February 2, 2009

1. Recall that variables can be of different types. We learned in introductory biostatistics that appropriate methods of summarization depend on the variable type. And we noticed that, sometimes, a method of summarization is ***not*** appropriate. For example, it is not appropriate to construct a cumulative frequency graph summary of nominal data.

   Using whatever sources you have for introductory biostatistics, complete the following table.

	Random Variable			
	Discrete		**Continuous**	
	Nominal	**Ordinal**	**Interval**	**Ratio**
Descriptive Methods				
Numerical Summaries				

2. Try your hand at the following probability exercises.

   a) Divide a line segment into three parts such that one portion is half the length of original line and the other two portions are each one quarter then length of the original line. Choose a point at random. What is the probability that this apoint is in the ½ length portion?

   b) If there is a 14% chance that any person selected at random was born on a Monday, what is the probability that, of any seven people selected at random, exactly one was born on a Monday?

   c) What are the odds of getting exactly one pair in five card stud poker using a 52 card deck?

   d) Suppose a quiz contains 20 true/false questions. You know the correct answer to the first 10 questions. You have no idea of the correct answer to questions 11 through 20 and decide to answer each using the coin toss method. Calculate the probability of obtaining a total quiz score of at least 85%.

3. Refresh your memory of the elements of a confidence interval. There are three: (1) point estimate; (2) standard error of the point estimate; and (3) confidence coefficient. Complete the following summary table

Normal Distribution: Confidence Interval for [$\mu_1 - \mu_2$] (Two Independent Groups) CI = [point estimate] $\pm$ (conf.coeff)SE[point estimate]			
	σ_X^2 and σ_Y^2 are both known	σ_X^2 and σ_Y^2 are both NOT known but are assumed EQUAL	σ_X^2 and σ_Y^2 are both NOT known and NOT Equal
Estimate			
SE to use			
Confidence Coefficient Use Percentiles from			
Degrees freedom			

4. See if you can recall some of the important concepts that are discussed in introductory biostatistics.

   (a) What is a sampling distribution?

   (b) What does the central limit theorem tell us? Why is it so useful to us? What is a z-score? What is a t-score?

   (c) In a sentence or two, explain the meaning of a 95% confidence interval for a population mean that has lower limit 35.6 and upper limit 52.8

   (d) Define p-value. Interpret $p<.05$ and $p<.01$. Given identical study conditions, which gives stronger evidence against the null hypothesis?

   (e) Suppose a two sided hypothesis test of treatment benefit in a randomized controlled trial of placebo versus active treatment yields a p-value of 0.045. What are the possible explanations for this result?