Notes: January 16, 2013 Probability Cont.

Review of z-scores: a z-score is the standardized distance of an observation from the mean. A z-score of 1 represents an observation that is 1 standard deviation away from the mean.

Homework format:

11 pt font

Labeled by lab (ta name or section number)

Late assignments get a zero

Separate by sections, staple (a must) separately, and provide cover sheets for each

Randomness:

Simple random sample: for n subjects from a population, where each possible sample of that size has the same probability of being selected.

Stratified random sample: random sampling by dividing into groups. Select random cities, in cities, select random neighborhoods, in neighborhoods, select random streets etc.

Systematic random sample: where if $n = sample$ and $N = population$ then $k = N/n$ {need more info}

i.i.d (independently and identically distributed)

Sampling error: is inherent error when a statistic is based on sample estimates or predicts the value of a population parameter.

The sampling error decreases in the square root of the sample, thus, a larger sample provides a smaller sample error. Larger samples also provide more normalized data, though it can also introduce a larger range of observations, which can include outliers.

Anything other than simple random sampling creates bias, however small.

Bias

Bias: When the statistics systematically give the wrong answers Example: which university is best in Utah, conducted in the Cougareat.

Missing data: is a big problem with samples, though complete data that has some observations randomly lost is not a problem because the data is still random. The bigger problem is having ,say, people answer a survey and some questions are not answered though others are. For this class, just ignore it.

Sampling distributions:

Sampling distributions: Is a probability distribution that determines probabilities of the possible values of a sample statistic (nobody does this)

Note: the sampling distribution tells us the relative frequency, or probability, of different values of the statistic. That is, in repeated applications of random, we would obtain different values of the statistic.

Thus, there are two types of variation:

1. The variance of sample data, or Y
2. The variance of summary statistics, the mean of Y