Formula "Card" For Basic Biostat 9/26/2007 Draft

Chapter 1: Measurement. The main measurement scales are quantitative, ordinal, and categorical. Make sure students understand that the validity and usefulness of a study is contingent on the *quality* of its measurements (GIGO principle).

Chapter 2: Types of Studies. Studies can be classified as surveys or comparative studies. Survey use probability samples to infer population characteristics. The SRS is the most fundamental type of probability sample. Comparative studies may be experimental or nonexperimental (observational). Experimentation usually requires the random assignment of the intervention. Use the Table A or a random number generator to help select SRSs (surveys) and to randomize subjects to treatment groups (experimental studies).

Chapter 3: Frequency Distributions. Explore the shape, center, and spread of the distribution. In addition, identify outliers. Useful exploratory techniques for quantitative variables include stemplots, histograms, and boxplots. (Boxplots are introduced in Chapter 4). Useful exploratory techniques for categorical variables include frequency tables, pie charts, and bar charts.

Chapter 4: Summary Statistics. Use the mean and standard deviation as the primary summary statistics for symmetrical distributions.

$$\bar{x} = \frac{1}{n}\sum x_i \qquad\qquad s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

For asymmetrical and oddly shaped distributions, use the median and IQR (or five-point summary) as the primary summary statistics. The median is the value with a depth of $\frac{n+1}{2}$. Average the two middle values when *n* is odd. Tukey's hinges can be used to identify Q1 and Q3. When you split the data in half, the median is included in both the low group and high group when *n* is odd. Note 1: It is advisable to *check* calculations with a calculator or computer application. Note 2: Supplement summary statistics with exploratory plots when possible.

Chapters 5: Probability Concepts. The four introductory properties of probability are: (1) $0 \le \Pr(A) \le 1$; (2) $\Pr(S) = 1$; (3) $\Pr(\bar{A}) = 1 - \Pr(A)$; and (4) Pr(A or B) = Pr(A) + Pr(B) for disjoint events. (Additional properties of probabilities are described on pp. 105 – 112). Use *pmf*s to determine probabilities for discrete random variables. Use *pdf*s to determine probabilities for continuous random variables. In both instances, the area under the curve (AUC) corresponds to probability. Advanced: For discrete random variables $\mu = \sum x \cdot \Pr(X = x)$ and $\sigma^2 = \sum (x-\mu)^2 \cdot \Pr(X = x)$.

Chapter 6: Binomial Distributions. Binomials are based on *n* independent Bernoulli trials, each with probability of success *p*. The binomial formula is $\Pr(X = x) = {}_nC_x p^x q^{n-x}$ where ${}_nC_x = \frac{n!}{x!(n-x)!}$.

The shortcut formulas for the mean and standard deviation are $\mu = np$ and $\sigma = \sqrt{npq}$ where $q = 1 - p$.

Chapter 7: Normal Distributions Draw the curve and use the 68-95-99 rule and fact that the AUC = 1 to approximate probabilities when possible. When that is not possible, (1) State the problem (2) standardize the values via $z = \frac{x-\mu}{\sigma}$ (3) Sketch and label the curve (4) Use Table B. To find a value that corresponds to the Normal probability (1) State the problem (2) Use Table B to look up the z-score (3) Sketch and label the curve (4) Unstandardize: $x = \mu + z_p\sigma$