03/26/2013

Probability

Mutually Exclusive: two events _____ happen at some point.

Addition Rule: P(A) + P(B) (Probability of A + Probability of B)

General Addition Rule: P(A) + P(B) – P(A&B)

Independence: One event does not affect the outcome of the next

General Multiplication Rule: P(A) x P(B|A) (Probability of B given A)

- P(B|A)= P(A&B)/P(A)

How to check for independence:

- P(A) x P(B) = P(A&B)
- If equal, than A&B are independent

Compliment Rule

- P(A^c) = 1 – P(A)

Statistics and Variation 03/26/2013

Intro to Book:

- Statistics- a way of reasoning, along with a collection of tools and methods, designed to help us understand the world. (quantities calculated from data)
- Data- values along with their context (singular form= Datum)

Data 03/26/2013

Data

- Answering Who, What, When, Where, Why and How can provide a context for data values
  - If you cant answer who and what then you don't have data
  - Add the context of Who and What and organize the values into a data table
- The rows of a data table correspond to individual cases about whom (or about which- if they're not people) we record some characteristics
- Case- an individual about whom or which we have data.
- Respondents – individuals who answer a survey
- Subjects or Participants- people on whom we experiment
- Experimental Units- animals, plants, websites, and other inanimate subjects that are tested on
- Records- in a database rows are called records ex. Purchase records
- Variables- the characteristics about each individual case (What has been measured)
- Relational Database- two or more separate data tables linked together so that information can be merged across them.
- Categorical Variable- variable names that categorize and answer questions about how cases fall into those categories.
- Quantitative Variable- when a variable measured numerical values with units and the variable tells us about the quantity of what is measured.
- Units- tell us how each value has been measured
- There are exactly as many categories as individuals and only one individual in each category. This is called the Identifier variable.
- Nominal Variables- Categorical variables used only to name categories
- Ordinal Values- the numeric location of an item within an order
- Time Series- when you have the same variable measured at regular intervals over time. (popular in business)
- Ratio Level- Similar to interval, except there is a true zero, or starting point, and the ratios of data values have meaning.
- Cross-Sectional Data- where several variables are measures at the same time point