*STAT1005 Week 2 Cheat Sheet*

***Counting Methods***

**Product (Multiplication) Rule**

- Counts how many outcomes from combining two sets (How many combinations of left shoe with rating 1-5 and right shoe with rating 1-5 are there in a pair of shoes)
  - If set A has $n_A$ outcomes and set B has $n_B$ outcomes, then the number of total outcomes of possible combinations is $n_A \times n_B$

**Definitions**

- n - How many options in a set there are to choose from (eg. rolling a dice a certain number of times, n will be what you might roll, 1, 2, ..., 6)
- k - How many times you will repeat choosing from the set
- Ordered - Means that the order of the outcomes matters in whether an outcome is different or not (eg. DOG and GOD, despite containing the same letters, are not the same outcome, because the letter are not in the same order, the order the letters came out in makes them different outcomes.)
- Unordered - Means that the order of the outcomes is irrelevant and different orders are still considered the same outcome (eg. DOG and GOD, they are considered the same outcome, because despite the different order in each, it's the same three letters and the order doesn't make a difference.) Look for DISTINCT.
- With replacement - Means that even if an option in a set has been chosen in a previous choice, it is still equally likely to be chosen again (Each time you roll a dice, you could land on any of the 6 sides.)
- Without replacement - Means that once an options has been chosen, it cannot be chosen again in future (Once someone has had a birthday, they cannot have another birthday that year).
- Permutations - Is how many ORDERED outcomes there are WITHOUT REPLACEMENT
- Combinations - Is how many UNORDERED outcomes there are WITHOUT REPLACEMENT

**General Product Rule**

- Counts how many ORDERED outcomes there are in a set.
  - $n_1 \times n_2 \times n_3 \ldots \times n_k$
  - If there IS replacement, then there are $n^k$ possible outcomes
  - If there is NOT/it is WITHOUT replacement, then there are n x (n-1) x (n-2) ... x (n-k+1) outcomes
    - This is considered PERMUTATIONS and can be simplified to $P_{k,n} = \frac{n!}{(n-k)!}$
      - For any positive integer n and 0! = 1

**Combinations**

- A continuation of permutations but UNORDERED
  - This can be written as $\binom{n}{k}$ or ${}_nC_k$
  - $\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{(n-k)!k!}$
    - Where $\binom{n}{n} = 1$ and $\binom{n}{0} = 1$, because there is only one way to choose a set of all n elements or of no elements, and $\binom{n}{1} = n$ since there are n subsets of size 1

***Non-Naive Definition of Probability***

- The probability of an event P(A) is a real number between 0 and 1.
- The function P must satisfy:
  - For any event A, $P(A) \geq 0$
  - $P(S) = 1$; $P(\emptyset) = 0$
  - If $A_1, A_2, A_3, \ldots$ is an infinite collection of disjoint events ($A_i \cap A_j = \emptyset$) for $i \neq j$, then
    - $P(A_1 \cup A_2 \cup A_3 \ldots) = \sum_{i=1}^{\infty} P(A_i)$

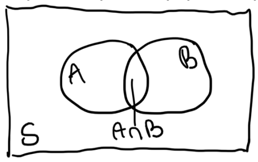
***Interpretations of Probability***

- **Frequentist**
  - Probability represents a long-run frequency over a large number of repetitions of an experiment
- **Bayesian**
  - Probability represents a degree of belief about the event in question, so we can assign probabilities to hypotheses, even if it isn't possible to repeat the same event over and over again

***Additional Properties of Probability***

- $P(A^c) = 1 - P(A)$
- If $A \subseteq B$, then $P(A) \leq P(B)$, $A \leq B$ - means if A is contained in B


- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

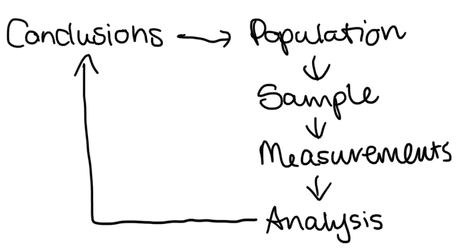


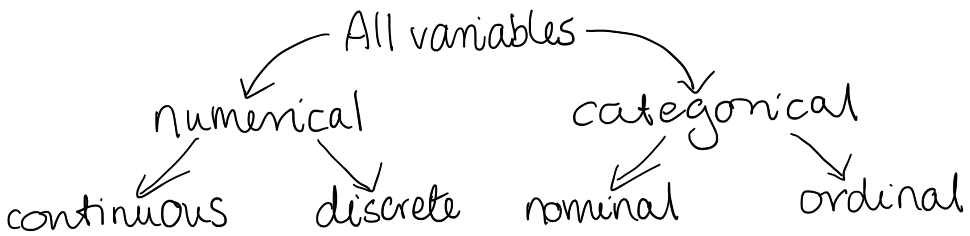

- **Property 1**
  - Since A and $A^c$ are disjoint and their union is S,
    - $P(S) = P(A \cup A^c) = P(A) + P(A^c)$
    - $P(S) = 1$, therefore $P(A) + P(A^c) = 1$
- **Property 2**
  - If $A \subseteq B$, then B can be written as A and $B \cap A^c$
    - Since A and $B \cap A^c$ are disjoint,
      - $P(B) = P(A \cup (B \cap A^c) = P(A) + P(B \cap A^c)$
    - Since probability is non-negative, $P(B \cap A^c) \geq 0$ and therefore $P(B) > P(A)$
- **Property 3**
  - $P(A \cup B) = P(A \cup (B \cap A^c))$
    $= P(A) + P(B \cap A^c)$
  - It follows that
    - $P(B \cap A^c) = P(B) - P(A \cap B)$
    - $P(A \cap B) + P(B \cap A^c) = P(B)$
    - $P(B \cap A^c) = P(B) - P(A \cap B)$
  - **DeMorgan's Laws**
    - $(A \cap B)^c = A^c \cup B^c$
    - $(A \cup B)^c = A^c \cap B^c$

***Different Types of Studies***

- **Observational**
  - Observes individuals or individual experimental units and measures variables of interests, but does not attempt to influence the responses
- **Experimental**
  - Deliberately imposes some treatment on individuals or experimental units to measure responses
  - Purpose is to study whether the treatment causes a change in the response

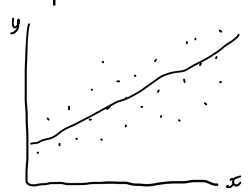
*When we want to understand CAUSE AND EFFECT, experiments are the only source os full convincing data*



- **Definitions**
  - Population - Entire group of individuals or experimental units about which we want information, or want to draw conclusions
  - Sample - A part of the population from which we actually collect information - variables
  - Statistical Inference - The process of using information from a sample to draw conclusions about the population
- **Types of Sampling**
  - Simple random sampling (SRS)
    - Best way to avoid bias
    - Each individual has an equal chance of being selected (usually done by a computer)
    - Why random sampling?
      - The eliminate bias in selecting samples
      - The laws of probability allow trustworthy inference about the population
        - Results with random sampling will come with a margin of error that sets bounds on the size of the likely uncertainty
        - Bigger samples give better information about the population
  - Stratified random sampling
    - First classify the population into groups of similar individuals called strata
    - Then choose a separate SRS in each stratum and combine them to form the full sample
- **Types of variables**
  - 

  - Quantitative/Numerical - variables have numerical values that measure a characteristic of each unit
    - Continuous: Can have infinite decimal places (time, distance, etc)
    - Discrete: Distinct, absolute numbers (number of cyclones, number of people, etc)
  - Categorical - variables place units into one of two or more categories
    - Nominal (unordered categories): Each category is equal but different (State of residence, gender)
    - Ordinal (ordered categories): Values placed on categories that are meaningful in relation to each other (Strongly agree, ..., strongly disagree)

***Data Analysis***

- Looking at the data in different way using graphical and numerical summaries
  - Helps to understand patterns, trends, unusual observations
  - Exploratory, because you're not making any formal conclusions
- Numerical summaries
  - Means, medians, ranges, ...
- Graphical summaries
  - Histograms, scatterplots, boxplots, ...
- Scatterplot
  - 

- Boxplot
  - 

- Sample mean -
  - $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$
- Sample standard deviation -
  - $$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}}$$
- Sample minimum, maximum
- Sample median
- Sample quartiles and interquartile range