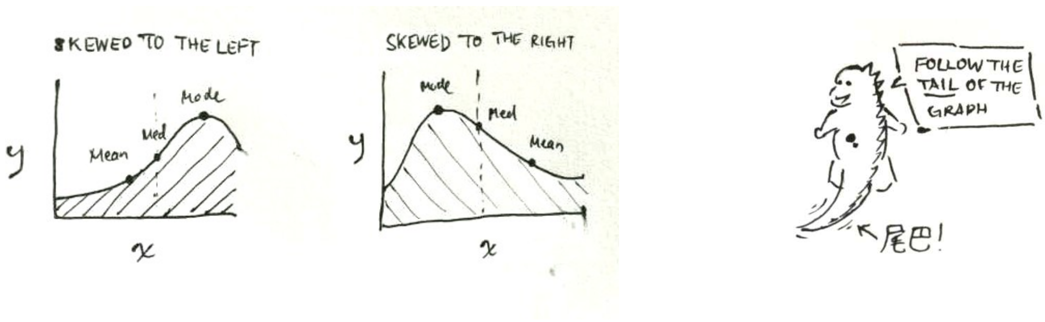
# Definitions and Terms

## A. Statistical Terminology

| | |
|---|---|
| Qualitative | Not naturally numerical; more of classification. Has no real meaning. |
| Quantitative | Naturally measured on a numerical scale; Has real meaning. |
| Interval Data (Quan) | The origin is artificial. (Dates, temperature, etc.) |
| Ratio (Quan) | The origin is natural. Can often be determined by the presence of a common absolute zero. (ex. Weight, Time) |
| Discrete (Quan) | Can attain values on a countable set of numbers (ex. Number of people can only be 1, 2, 3… You cannot have 2.5 or 3.4 people.). Can be infinite or finite. |
| Continuous (Quan) | Assumes values in a given finite/infinite set of data. (ex. A person’s weight can be 101, 101.1, 101.12…) |
| Nominal (Qual) | A descriptive category, with no natural numerical value with respect to magnitude. Cannot be reasonably ranked. (ex. sex, nationality) |
| Ordinal (Qual) | Can be ordered and counted, but not reasonably measured. (ex. swimming level, grades) |
| Univariate | Measures only one characteristic. (ex. age or salary only.) |
| Multivariate | More than one characteristic measured. |
| Histogram | Represents frequency of data |
| Scatterplot | Used to determine trends and relations |
| Time Series | Horizontal or x axis measures time. |
| Sample Mean | Just the average. |
| Median (m) | Middle value when arranged in order. (Ex. 2 in the set {1, 2, 3}) |
| Mode | Most frequently appearing value |
| Skewness | If skewed to the right, the median is less than the mean. If to the left, mean is less than the median. See Figure 1. |
| Symmetry | If a graph is symmetric, mean=median=mode |
| Sample Range | Maximum Value - Minimum Value |
| Upper Quartile | Upper 25%, or 75th percentile |
| Lower Quartile | Lower 25th percent, or 25th percentile. |
| Inter-Quartile Range | Distance between upper and lower quartiles. |
| Variance | Square of the Standard Deviation |
| Standard Deviation | Square root of the variance. Measure of deviation. |

**Figure 1, Skewness:**




## B. Variables

| | |
|---|---|
| X (with line on top) | Average |
| μ | Real Population Mean. Estimated by X. |
| $\delta^2$ | Population Variance |
| δ | Population Standard Deviation |
| m | Median |
| $Q_U$ | Upper Quartile |
| $Q_L$ | Lower Quartile |
| S | Universal Set |
| ∅ | Empty Set |

## C. Measures of Deviation

How to solve for Standard Deviation.

Given:
Daniel's Grades: 80, 90, 100   X=90

What is the Standard Deviation for Daniel? The Variance?

# Probability

*"If the census is not made, then the parameter's value will never be known to us. They can only be estimated by sample observations."*

### A. Axioms of Probability

P(S)=1
If A & B are disjoint, then: **P(A∪B)=P(A)+P(B)**

### B. Theorems

P(A')=1-P(A)
For any two events A & B, **P(A∪B)=P(A)+P(B)-P(A∩B)**
If A⊂B (A is a subset of B), then P(A) ≤ P(B)
Assume n(S) = k, where n(s) represents the possible outcomes in S; n(A) in A. If the probability of each single outcome of S is equal, than S is called a finite sample space with equal possibilities.

### C. Combinatorics I

1. Rule of Products: If you have something that can be done in x stages and can be done in z different ways in each step, & m in the following stages, the whole thing can be done in mz ways.
2. Given n distinct objects, permutations can either be with or without repetitions. Order is important.
3. The number of permutations of n distinct objects without repetition is **n!**
4. The number of permutations of n distinct objects with repetitions is **$n^r$.**
5. The number of permutations of r out of n distinct items with repetitions is **$n^r$**.
6. The number of permutations of r items out of n distinct items without repetition is **n!/(n-r)!**

So, in general,
Combinations: Order is not important (Ex. Creating committees.)
Permutations: Order is important (Ex. Making passwords.)

Combination and Permutation Notation

**nPr = n!/(n-r)!** - order matters
**nCr = (n | r) = n!/r!(n-r)!** - order does not matter, no repetition
nCr (repetition allowed) =**(n+r-1)/r!(n-1)!**

Example: How many different ways can you arrange 3 pool balls out of a set of 16?

Permutations Solution: 16!/13!

Combinations Solution: 16!/3!(13!) *Why*? Because we only care about which balls are selected, so we take out the possible switching around of the order of the three balls by dividing by 3!.

Example 2: If you had 5 ice cream flavours, and you could choose 3, while allowing for the repeating of flavours, how many possible combinations are there?

7!/3!(4!)
=35

# Probability II

## Conditional Probability

Notation: **P(B|A) = P(B will occur given A occurred)**

Formula: **P(B|A) = (Relative Area of B out of A)/(Area of A) = P(A∩B)/P(A)**

## Independence and Dependence

Given events A & B, B is said to be independent of the event A if & only if
**P(B|A)=P(B), [P(A)>0] = P(A∩B)/P(A)**

Alternatively,
**P(A∩B) = P(A) P(B)**

## SUMMARY OF FORMULAE:

| | |
|---|---|
| Standard Deviation | **SQRT[(SUM(N**<br>where $N_i$<br>is the number of items in the sample set. See Figure 2. |
| Probability of an Event A | P(A)=n(A)/n(S) |
| Combined Events | P(A∪B) = P(A)+P(B)-P(A∩B) |
| Mutually Exclusive Events | P(A∪B) = P(A) + P(B) |
| Independence Test I | P(A∩B) = P(A)P(B) |
| Independence Test II | P(B), [P(A)>0] = P(A∩B)/P(A) |
| Conditional Probability | P(A∩B)=P(A) P(B\|A) |

**Figure 2:**

$$SD=\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

# Sample Solutions

***Sample 1: Simple Combinatorics***

How many possible passwords can you make out of four numbers, if you are allowed to repeat? What if you cannot repeat?

a.) If you can repeat: 10*10*10*10
b.) If you cannot repeat: 10*9*8*7

***Sample 2: Intermediate Combinatorics***

Given three fair coin tosses, what is the probability that you will get two tails and one head?

3/8. There are only three possibilities within the constraints, namely HTT, TTH and HTH.
(H=Heads, T=Tails)

***Sample 3: Conditional Probability and Independence***

For two events A & B, P(A)=0.4, P(B)=0.2, and P(A ∩ B)=0.1.

a.) Find P(A|B): Possibility of A, given B= 0.5
b.) Find P(B|A): Possibility of B, given A=0.25
c.) Are A & B independent? No, they are dependent, as they do not fulfil a requisite formula that indicates independence.

***Sample 4: Loss-Profit Analysis***

There is a lottery for which the prize is $7,000,000 and the chance of winning is 1/23,000,000 for every $1 ticket.

If there is a 1 in 23,000,000 chance of winning the lottery, one will need to purchase $23,000,000 worth of tickets (or simply 23,000,000 tickets) to be assured of a win. Thus, given an assured win, the net loss is $23,000,000-$7,000,000, or $16,000,000. Therefore, dividing the net losses per game (or $16,000,000/23,000,000), you get a net loss of around 70 cents for every dollar you spend.