Linear Algebra and Probability

The Extremely Abridged Version for Data Analytics

Matrices and Vectors

Matrix: A rectangular array of numbers, e.g., $A \in \mathbb{R}^{m \times n}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Vector: A matrix consisting of only one column (default) or one row, e.g., $x \in \mathbb{R}^n$

Matrix Math

If $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $C = AB$, then $C \in \mathbb{R}^{m \times p}$:

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Special cases: Matrix-vector product, inner product of two vectors. e.g., with $x, y \in \mathbb{R}^n$:

$$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$

Associative: $(AB)C = A(BC)$
Distributive: $A(B + C) = AB + AC$
Non-commutative: $AB \neq BA$

Operators and Properties

Transpose: $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$: $(A^T)_{ij} = A_{ji}$

Properties:

$(A^T)^T = A$
$(AB)^T = B^T A^T$
$(A + B)^T = A^T + B^T$

Trace: $A \in \mathbb{R}^{n \times n}$, then: $tr(A) = \sum_{i=1}^{n} A_{ii}$

Properties:

$tr(A) = tr(A^T)$
$tr(A + B) = tr(A) + tr(B)$
$tr(\lambda A) = \lambda tr(A)$
If AB is a square matrix, $tr(AB) = tr(BA)$

Special Matrices

Identity matrix: $I = I_n \in \mathbb{R}^{n\times n}$:

$$I_{ij} = \begin{cases} 1 & \text{i=j,} \\ 0 & \text{otherwise.} \end{cases}$$

$\forall A \in \mathbb{R}^{m\times n}$: $AI_n = I_m A = A$

Diagonal matrix: $D = diag(d_1, d_2, \ldots, d_n)$:

$$D_{ij} = \begin{cases} d_i & \text{j=i,} \\ 0 & \text{otherwise.} \end{cases}$$

Symmetric matrices: $A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^T$.

Orthogonal matrices: $U \in \mathbb{R}^{n\times n}$ is orthogonal if
$UU^T = I = U^T U$

Linear Independence and Rank

A set of vectors $\{x_1, \ldots, x_n\}$ is linearly independent if $\nexists\{\alpha_1, \ldots, \alpha_n\}$: $\sum_{i=1}^{n} \alpha_i x_i = 0$

Rank: $A \in \mathbb{R}^{m \times n}$, then $rank(A)$ is the maximum number of linearly independent columns (or equivalently, rows)

Properties:

$rank(A) \leq \min\{m, n\}$
$rank(A) = rank(A^T)$
$rank(AB) \leq \min\{rank(A), rank(B)\}$
$rank(A + B) \leq rank(A) + rank(B)$

Matrix Inversion

If $A \in \mathbb{R}^{n \times n}$, $rank(A) = n$, then the inverse of A, denoted A^{-1} is the matrix that: $AA^{-1} = A^{-1}A = I$

Properties:

$$(A^{-1})^{-1} = A$$
$$(AB)^{-1} = B^{-1}A^{-1}$$
$$(A^{-1})^{T} = (A^{T})^{-1}$$

Really Useful Things Matrix Operations Can Do!

- Rotate a vector
- Solve linear systems of equations

A linear system might be described by the following equations:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

These equations could be written in matrix form as:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

The matrix equation could be written as: **Ax = b**

rotation matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}$$

NOTE: this is a 2 by 1 column vector, not a 2 by 2 matrix.

Probability

For events A, B:

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}$$

A, B independent if $P(A|B) = P(A)$ or equivalently: $P(A \bigcap B) = P(A)P(B)$

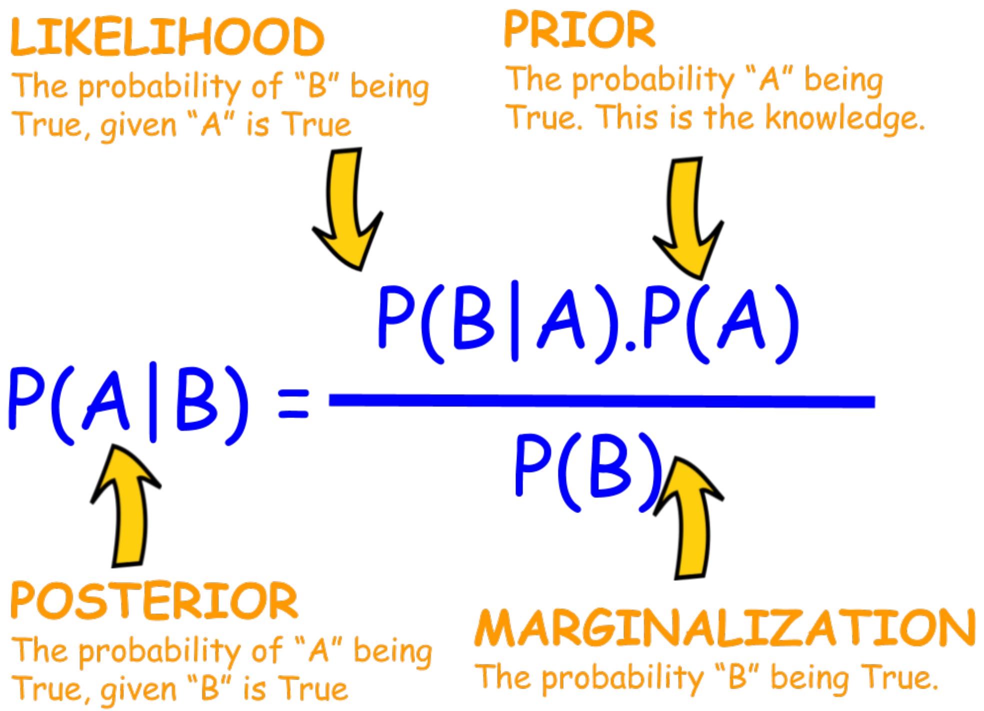
Sample Space Ω: Set of all possible outcomes

Event Space $\mathcal{F}$: A family of subsets of Ω

Probability Measure: Function $P : \mathcal{F} \to \mathbb{R}$ with properties:

1. $P(A) \geq 0 \quad (\forall A \in \mathcal{F})$
2. $P(\Omega) = 1$
3. A_i's disjoint, then $P(\bigcup_i A_i) = \sum_i P(A_i)$

Bayes Rule



Random Variables

A random variable X is a function $X : \Omega \rightarrow \mathbb{R}$
Example: Number of heads in 20 tosses of a coin

Probabilities of events associated with random variables defined based on the original probability function. e.g., $P(X = k) = P(\{\omega \in \Omega | X(\omega) = k\})$

Cumulative Distribution Function (CDF) $F_X : \mathbb{R} \rightarrow [0, 1]$:
$F_X(x) = P(X \leq x)$

Probability Mass Function (pmf): X discrete then
$p_X(x) = P(X = x)$

pmf is also known as, a discrete probability density function

CDF:

$0 \leq F_X(x) \leq 1$
F_X monotone increasing, with $lim_{x \rightarrow -\infty} F_X(x) = 0$,
$lim_{x \rightarrow \infty} F_X(x) = 1$

pmf:

$0 \leq p_X(x) \leq 1$
$\sum_x p_X(x) = 1$
$\sum_{x \in A} p_X(x) = p_X(A)$

Common Random Variables

$X \sim \mathit{Bernoulli}(p)$ $(0 \leq p \leq 1)$:

$$p_X(x) = \begin{cases} p & \text{x=1,} \\ 1-p & \text{x=0.} \end{cases}$$

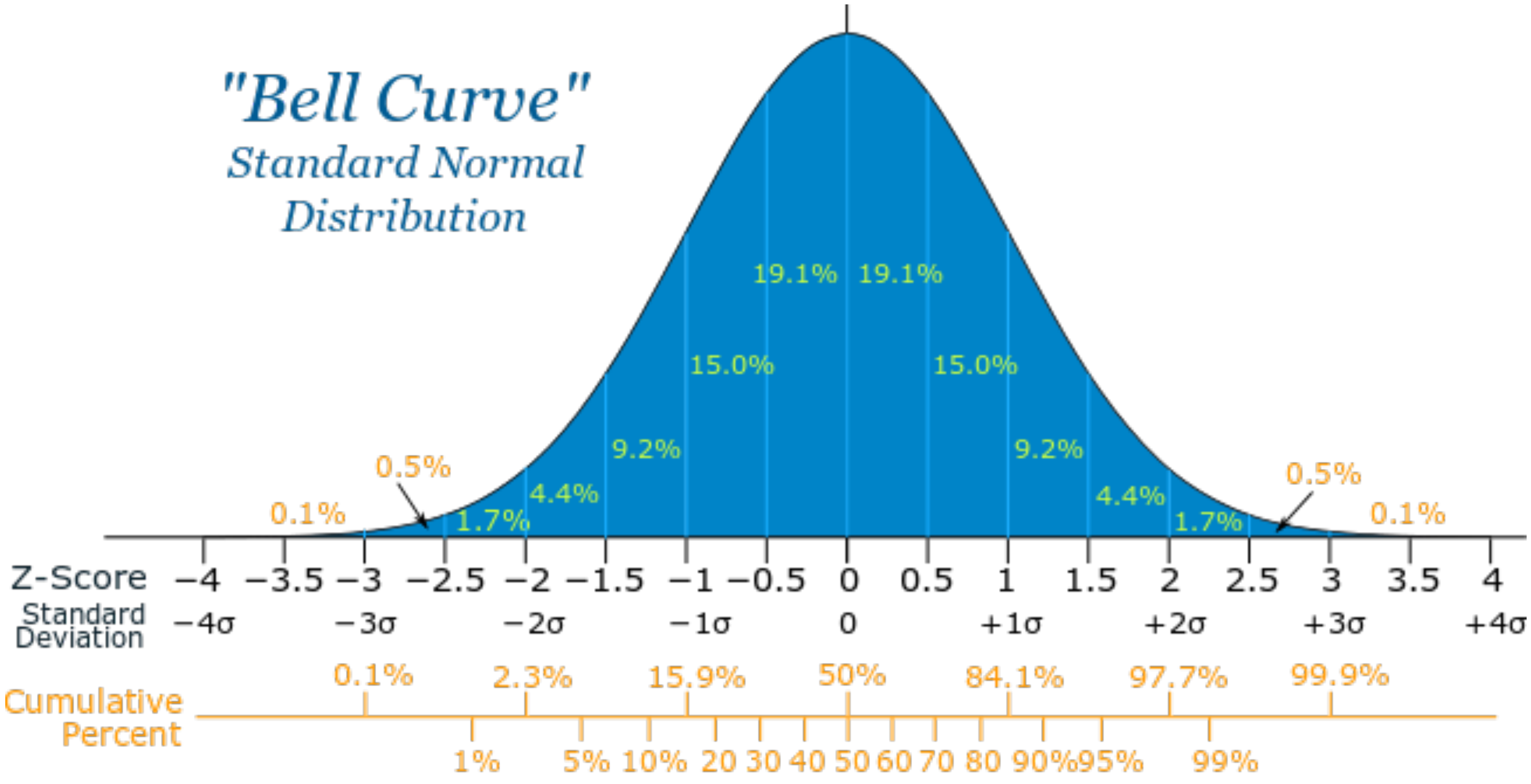
$X \sim \mathit{Geometric}(p)$ $(0 \leq p \leq 1)$: $p_X(x) = p(1-p)^{x-1}$
$X \sim \mathit{Uniform}(a, b)$ $(a < b)$:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

$X \sim \mathit{Normal}(\mu, \sigma^2)$:

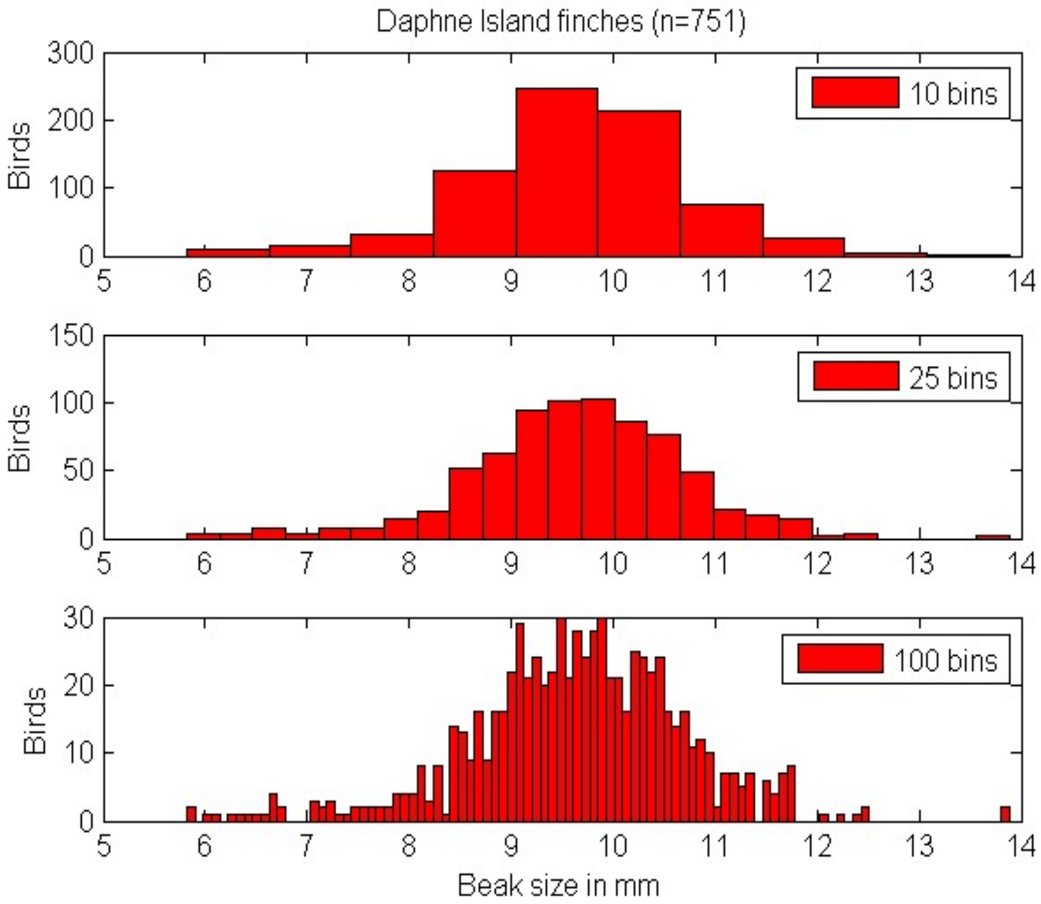
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Gaussian (aka Normal) Distribution

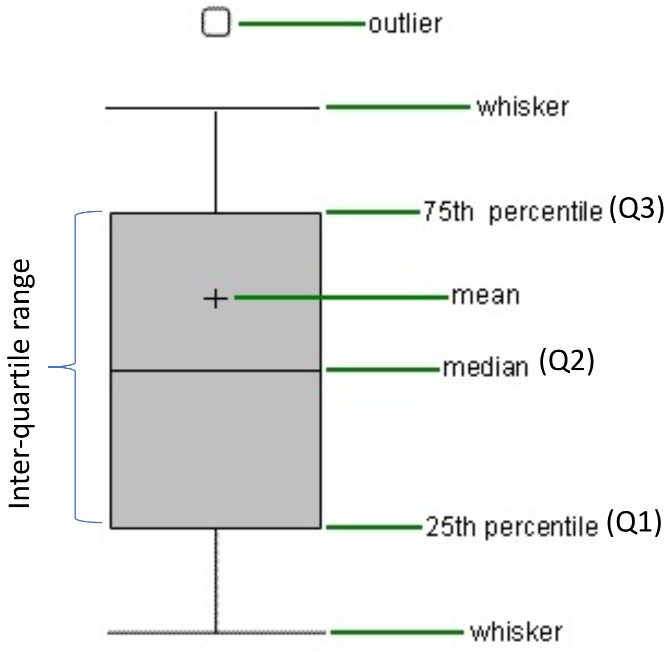


Univariate Analysis – Continuous Variables

- Distributions



- Box and Whisker Plots & Quartiles

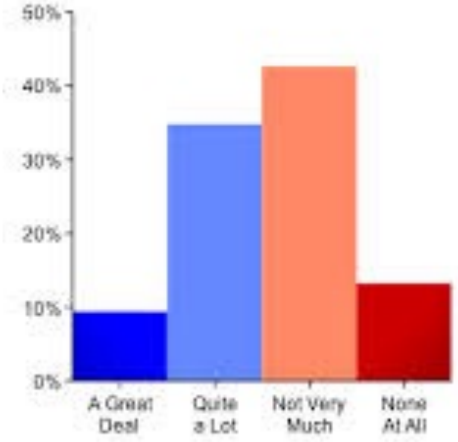


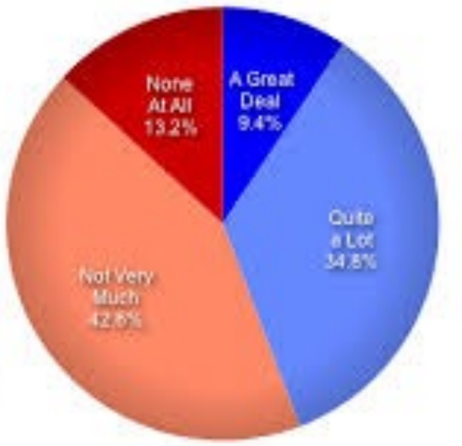
Outlier = larger than Q3 or smaller than Q1 by at least 1.5 times the inter-quartile range.

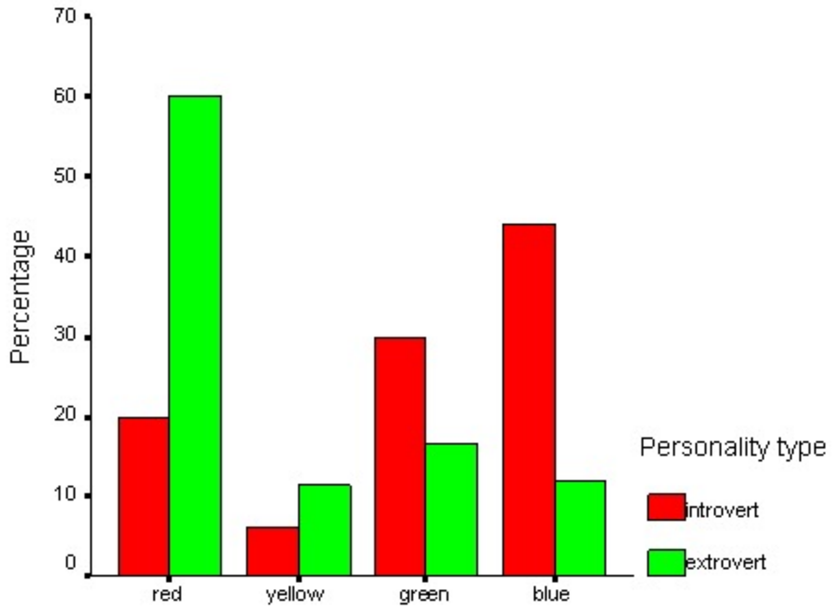
Univariate Analysis – Categorical Variables

- Frequency Table, bar chart, pie chart

Degree	Frequency	Percentage
High School	2	5.0
Bachelor's	7	17.5
MBA	20	50.0
Master's	3	7.5
Law	4	10.0
PhD	4	10.0
	40	

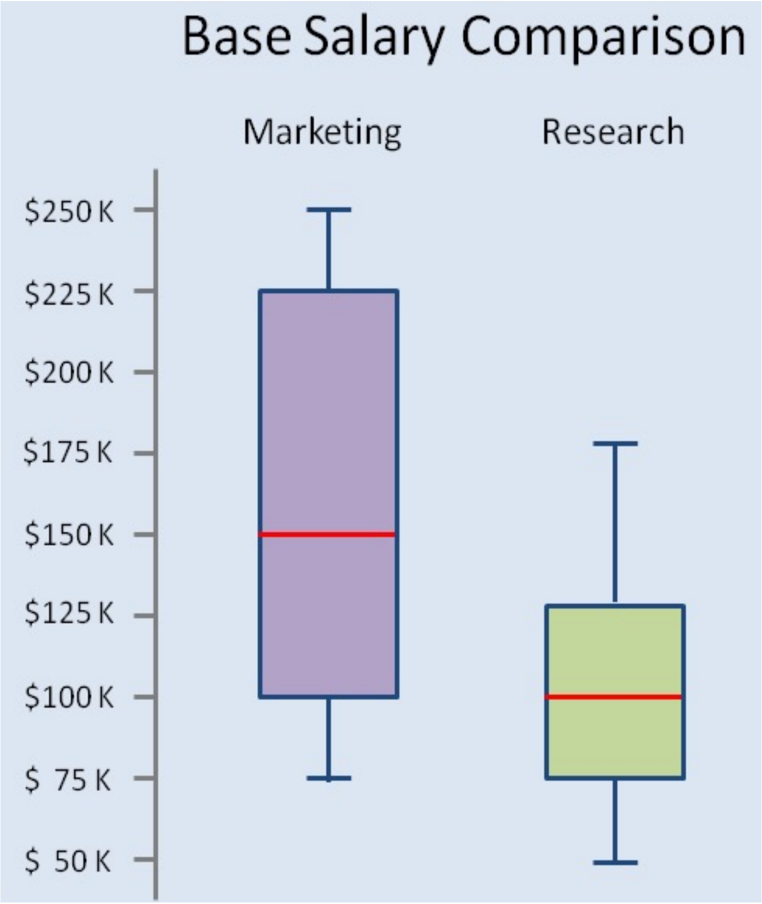






Bivariate Analysis – Continuous & Categorical

- Box and Whisker Plots



- Z-Test – is the difference in these statistically significant?

$$z = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Example: For the Marketing group of 900 people (n = 900), the mean score on the test was 9.78, with a std dev (s)= 4.05.

For the Research group of 1000 people (n= 1000), the mean score on the test was 15.10, with a std dev (s) = 4.28.

$$z = \frac{(9.78 - 15.10) - 0}{\sqrt{\frac{4.05^2}{900} + \frac{4.28^2}{1000}}} = \frac{-5.32}{\sqrt{\frac{16.40}{900} + \frac{18.32}{1000}}} = \frac{-5.32}{\sqrt{.018 + .018}} = \frac{-5.32}{.19} = -28$$

Look up z-score in a table, see if it falls below p=0.05,
if it does, reject the null hypothesis, there is a statistical difference

Set Theory

Abridged from the excellent resource at
http://www.math.clemson.edu/~mjs/courses/misc/settheory.pdf

What is a Set?

- A set is one of those fundamental mathematical ideas whose nature we understand without direct reference to other mathematical ideas. Quite simply,

A set is a collection of distinct objects.

Subsets, Equality, and Size

- Two sets S and T are equal, written $S = T$, if S and T contain exactly the same elements.

- A set S is a subset of another set T, written $S \subset T$ if every element of S is also an element of T.

- The cardinality of a set S is denoted $|S|$ and is the number of elements in the set

Intersections and Unions

- The intersection of two sets S and T, written $S \cap T$ is the set of elements common to both S and T, i.e.,
  - $S \cap T = \{x : x \in S \text{ and } x \in T\}$
- The union of two sets S and T, written $S \cup T$, is the set of elements that are in either S or T or both, i.e.,
  - $S \cup T = \{x : x \in S \text{ or } x \in T\}$

Set Differences and Complements

- The difference between S and T, written S \ T, is the set of elements in S but not also in T:
  - $S \setminus T = \{x : x \in S \text{ and } x /\in T\}$
- Relative to a universe U, the complement of S, written S 0 , is the set of all elements of the universe not contained in S, i.e.,
  - $S' = \{x : x \in U \text{ and } x \notin S\}$