Probability

- Two events are **mutually exclusive** if $\boxed{P(A \cap B) = 0}$
- Two events are **independent** if $\boxed{P(A \cap B) = P(A)P(B)}$
  Alternatively, $P(A \mid B) = P(A)$ if A, B independent.

DRV

- $\boxed{\text{Var}(X) = E(X^2) - \Big(E(X)\Big)^2}$

Binomial

- (the number of trials, n, is fixed)
- (each trial results in only two outcomes, "success" or "failure")
- The probability of success, p, is **the same** for each trial
- Each trial is **independent** from each other

Normal

- $E(aX - b) = aE(X) - b$
- $\text{Var}(aX - b) = a^2\text{Var}(X)$
- $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$
- For $X_1 + X_2$, $2X - 3Y$, $\overline{X}$ to be normally distributed, X_1, X_2, Y must be **independent**
- Characteristics of a normal distribution:
  - Almost all observations lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.
  - Symmetrical about μ.
  - Has one mode at μ.

Sampling

- The **population** includes all members of a study
- A **sample** is a subset of the population
- A sample if **random** if each member of the population has the same probability of being selected into the sample.
- μ: **population mean**
- $\overline{x}$: **sample mean** (also used as the unbiased estimate of population mean)
- σ: **population standard deviation**
- σ^2: **population variance**
- s^2: **unbiased estimate of population variance**
- $s^2 = \dfrac{n}{n-1}$(sample variance)

Hypothesis Testing

- A **null hypothesis**, H_0, is an assertion about a population that we would like to test
- An **alternative hypothesis**, H_1, is a contrasting hypothesis that will be tested against H_0
- $\overline{x}$ and z are our **test statistic**
- The **critical region** is the set of values for the test statistic such that H_0 is rejected
- *** The **critical value** is the cut-off value for the critical region
- The **level of significance** is the probability that our hypothesis test concludes that $\boxed{H_1 \text{ is true}}$ when in fact $\boxed{H_0 \text{ is true}}$
- The **p-value** is the
  - probability that, under H_0, we obtain a test statistic less/more extreme/more than the value obtained from our sample
  - the least level of significance that will result in the rejection of H_1
- If n is large, by the Central Limit Theorem, the sample mean $\overline{X}$ is normally distributed approximately so it is **not** necessary to assume anything about the distribution of X.
- If n is small, it is necessary to assume that X is normally distributed.