Summary Statistics

- The sample mean : $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
- The sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
or $s^2 = \frac{1}{n-1}\left[(\sum_{i=1}^{n} y_i^2) - n(\bar{y}^2)\right]$
- The sample standard deviation: $s = \sqrt{s^2}$.
- The k^{th} percentile: Find the $(\frac{k}{100})(n+1)$ location, then use the "rounding rule".
- The median ($\tilde{y}$) is the 50^{th} percentile.
- The five number summary is: Min, Q_1,$\tilde{y}$, Q_3, Max.
- The lower cutoff is: $Q_1 - 1.5(Q_3 - Q_1)$
- The upper cutoff is: $Q_3 + 1.5(Q_3 - Q_1)$

Probability Theory

Consider events A and B.

- Rule 1: $0 \leq Pr\{A\} \leq 1$.
- Rule 2: If there are k events $A_1, A_2, \ldots, A_k$ that make up all possible events, then $\sum_{i=1}^{k} Pr\{A_i\} = 1$
- Rule 3: The probability that A does not occur is: $P\{A^C\} = 1 - Pr\{A\}$
- Rule 4: For any two events A and B, the probability of "A occurs or B occurs or both occur" is:
$P\{A \text{ or } B\} = Pr\{A\} + Pr\{B\} - Pr\{A \text{ and } B\}$
- Rule 5: If A and B are mutually exclusive (or disjoint), then $Pr\{A \text{ and } B\} = 0$
- Rule 6: The conditional probability of A given B has occurred is:
$Pr\{A|B\} = \frac{Pr\{A \text{ and } B\}}{Pr\{B\}}$
- Rule 7: $Pr\{A \text{ and } B\} = Pr\{A|B\}Pr\{B\}$
- Rule 8: $Pr\{A \text{ and } B^c\} = Pr\{A\} - Pr\{A \text{ and } B\}$
- Rule 9: $Pr\{A^C|B\} = 1 - Pr\{A|B\}$
- Rule 10: If an event A is split by multiple events $B_1, B_2, \ldots, B_k$, then the following is true: $Pr\{A\} = Pr\{A \text{ and } B_1\} + Pr\{A \text{ and } B_2\} + \cdots + Pr\{A \text{ and } B_k\}$

  For two events A and B:
  $Pr\{A\} = Pr\{A \text{ and } B\} + Pr\{A \text{ and } B^C\}$
- For two events A and B which are independent, both of the following properties hold true:
  1. $Pr\{A \text{ and } B\} = Pr\{A\}Pr\{B\}$
  2. $Pr\{A|B\} = Pr\{A\}$

Discrete Random Variables

- The mean of a discrete random variable is:
$\mu_Y = \sum_{y_i} y_i Pr\{Y = y_i\}$
- The variance of a discrete random variable is:
$\sigma_Y^2 = \sum_{y_i}(y_i - \mu_Y)^2 Pr\{Y = y_i\}$
$= \left(\sum_{y_i} y_i^2 Pr\{Y = y_i\}\right) - (\mu_Y)^2$
- The standard deviation of a discrete random variable is:
$\sigma_Y = \sqrt{\sigma_Y^2}$

Linear Combinations of R.V.s

For any random variable X with mean μ_X and standard deviation σ_X, if $Y = a + bX$ (where a, b are constants) then

- $\mu_Y = a + b\mu_X$
- $\sigma_Y^2 = b^2\sigma_X^2$

Binomial Random Variables

If Y is a binomial random variable;

- $Pr\{Y = j\} = \binom{n}{j} p^j (1-p)^{n-j}$
where $\binom{n}{j} = \frac{n!}{j!(n-j)!}$
- $\mu_Y = np$
- $\sigma_Y^2 = np(1-p)$

Normal Random Variables

If Y is a normal random variable with mean μ_Y, standard deviation σ_Y (i.e $Y \sim N(\mu_Y, \sigma_Y)$) then;

- $Z = \frac{Y - \mu_Y}{\sigma_Y}$ is standard normal, i.e. $Z \sim N(0,1)$.
- $Pr\{Z > a\} = 1 - Pr\{Z < a\}$
for some constant a.
- $Pr\{a < Z < b\} = Pr\{Z < b\} - Pr\{Z < a\}$
for some constants a and b.
- The k^{th} percentile of Y is :
$Y^{(k)} = \mu_Y + Z^{(k)}\sigma_Y$
where $Z^{(k)}$ is the k^{th} percentile of a Z.