Random variables

- **random variable**
  variable whose value cannot be predicted with certainty before it is measured

- **probability model**
  method for assigning probabilities to the various possible outcomes of a random event, like the measure of a random variable

When a random variable has *only finitely many possible values* (the variable is **discrete**), assign probabilities to each value; then the probability of any event consisting of a collection of values of the variable is the sum of the probabilities for the individual measurements.

When *events concerning a random variable consist of interval ranges of values* (the variable is **continuous**), as with a variable described by the normal model, the probability of any event is the area under the model curve over that interval range.
Discrete random variables

• **expected value**
  if $X$ is a quantitative discrete random variable, then its expected value is its ideal mean value $\mu$ and can be computed from its probability model as the sum of the products of the values of $X$ with their associated probabilities

$$\mu = E(X) = \sum x \cdot P(X = x)$$

• **variance (of a random variable)**
  if $X$ is a quantitative discrete random variable, then its variance is *the expected value of the squared deviations* of the values of $X$; it can be computed from its probability model:

$$\sigma^2 = Var(X) = \sum (x - \mu)^2 \cdot P(X = x)$$

• **standard deviation (of a random variable)**
  if $X$ is a quantitative discrete random variable, then its variance is *the expected value of the squared deviations* of the values of $X$; it can be computed from its probability model:

$$\sigma = SD(X) = \sqrt{\sum (x - \mu)^2 \cdot P(X = x)}$$
Rescaling random variables

- **Shifting** values of a random variable \( X \) by adding or subtracting a fixed value \( c \) should shift the expected value by the same amount, but leave the variance unchanged (the center of the distribution of data produced by sampling \( X \) would shift by the amount \( c \) but its spread would be unaffected):

\[
E(X \pm c) = E(X) \pm c, \quad \text{Var}(X \pm c) = \text{Var}(X)
\]

- **Scaling** values of a random variable \( X \) by multiplication by a fixed ratio \( a \) should scale the expected value by the factor of \( a \), and scale the variance by a factor \( a^2 \) (data produced by \( X \) would scale by \( a \) while squared deviations from their mean would scale by the square of \( a \)):

\[
E(aX) = a \cdot E(X), \quad \text{Var}(aX) = a^2 \cdot \text{Var}(X)
\]

- **Adding** or **subtracting** values of two random variables \( X \) and \( Y \) produces a third random variable \( X \pm Y \) whose expected value is the sum or difference of the expected values of the component variables and whose variance is the sum of the variances of the component variables:

\[
E(X\pm Y) = E(X) \pm E(Y), \quad \text{Var}(X\pm Y) = \text{Var}(X) + \text{Var}(Y)
\]
Continuous random variables

• **expected value**
  if $X$ is a continuous random variable, then its expected value is its ideal mean value $\mu$, and can be identified from its probability model as the point on the scale where the ideal distribution curve “balances its weight”; for a normal model, it is the position of its central peak, or mode

• **variance**
  if $X$ is a continuous random variable, then its variance $\sigma^2$ is the expected value of the squared deviations of the values of $X$; it can be computed from its probability model using ideas from calculus (which we will not treat here)

• **standard deviation (of a random variable)**
  if $X$ is a continuous random variable, then its standard deviation $\sigma$ is the square root its variance; for a normal model, it is also the distance from $\mu$ to the position of the inflection point along either side of the model curve

All the rules for shifting, scaling, adding or subtracting discrete random variables apply in the continuous case as well.