Definition 1.1

The *mean* of a sample of n measured responses $y_1, y_2, \ldots, y_n$ is given by

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \,.$$

The corresponding population mean is denoted μ .

Definition 1.2

The *variance* of a sample of measurements $y_1, y_2, \ldots, y_n$ is the sum of the square of the differences between the measurements and their mean, divided by $n - 1$. Symbolically, the sample variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \overline{y})^2 .$$

The corresponding population variance is denoted by the symbol σ^2.

Definition 1.3

The *standard deviation* of a sample of measurements is the positive square root of the variance; that is,

$$s = \sqrt{s^2}.$$

The corresponding population standard deviation is denoted by $\sigma = \sqrt{\sigma^2}$.

Definition 2.1

An experiment is the process by which an observation is made.

Definition 2.2

A *simple event* is an event that cannot be decomposed. Each simple event corresponds to one and only one *sample point*. The letter E with a subscript will be used to denote a simple event or the corresponding sample point.

Definition 2.3

The *sample space* associated with an experiment is the set consisting of all possible sample points. A sample space will be denoted by S.

Definition 2.4

A *discrete sample space* is one that contains either a finite or a countable number of distinct sample points.

Definition 2.5

An *event* in a discrete sample space S is a collection of sample points—that is, any subset of S.

Definition 2.6

Suppose S is a sample space associated with an experiment. To every event A in S (A is a subset of S), we assign a number, $P(A)$, called the probability of S, so that the following axioms hold:

Axiom 1: $P(A) \geq 0$.

Axiom 2: $P(S) = 1$.

Axiom 3: If $A_1, A_2, A_3, \ldots$ form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = \sum_{i=1}^{\infty} P(A_i).$$

Definition 2.7

An ordered arrangement of r distinct objects is called a *permutation*. The number of ways of ordering n distinct objects taken r at a time will be designated by the symbol P_r^n.

Definition 2.8

The number of *combinations* of n objects taken r at a time is the number of subsets, each of size r, that can be formed from the n objects. This number will be denoted by $C_r^n \; or \; \binom{n}{r}$.

Definition 2.9

The *conditional probability* of an event A, given that an event B has occurred, is equal to

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

provided $P(B) > 0$. [The symbol $P(A|B)$ is read "probability of A given B."]

Definition 2.10

Two events A and B are said to be *independent* if any one of the following holds:

$$P(A|B) = P(A),$$

$$P(B|A) = P(B),$$

$$P(A \cap B) = P(A)P(B).$$

Otherwise, the events are said to be dependent.

Definition 2.11

For some positive integer k, let the sets $B_1, B_2 \ldots, B_k$ be such that

1. $S = B_1 \cup B_2 \cup \ldots \cup B_k$.
2. $B_i \cap B_j = \emptyset, \; for \; i \neq j$

Then the collection of sets $\{B_1, B_2, \ldots, B_k\}$ is said to be a *partition* of S.

Definition 2.12

A *random variable* is a real-valued function for which the domain is a sample space.

Definition 2.13

Let N and n represent the numbers of elements in the population and sample, respectively. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an equal probability of being selected, the sampling is said to be random, and the result is said to be a *random sample*.

Definition 3.1

A random variable Y is said to be *discrete* if it can assume only a finite or countably infinite number of distinct values.

Definition 3.2

The probability that Y takes on the value y, $P(Y = y)$, is defined as the *sum of the probabilities of all sample points* in S that are assigned the value y. We will sometimes denote $P(Y = y)$ by $p(y)$.

Definition 3.3

The *probability distribution* for a discrete variable Y can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.

Definition 3.4

Let Y be a discrete random variable with the probability function $p(y)$. Then the expected value of Y, $E(Y)$, is defined to be

$$E(Y) = \sum_y yp(y).$$

Definition 3.5

If Y is a random variable with mean $E(Y) = \mu$, the variance of a random variable Y is defined to be the expected value of $(Y - \mu)^2$. That is,

$$V(Y) = E[(Y - \mu)^2].$$

The *standard deviation* of Y is the positive square root of $V(Y)$.

Definition 3.6

A binomial experiment possesses the following properties:

1. The experiment consists of a fixed number, n, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is Y, the number of successes observed during the n trials.

Definition 3.7

A random variable Y is said to have a binomial distribution based on n trials with success probability p if and only if

$$p(y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0, 1, 2, \ldots, n \quad and \quad 0 \leq p \leq 1.$$

Definition 3.8

A random variable Y is said to have a *geometric probability distribution* if and only if

$$p(y) = q^{y-1}p, \quad y = 1, 2, 3, \ldots, \quad 0 \leq p \leq 1.$$

Definition 3.9

A random variable Y is said to have a *negative binomial probability distribution* if and only if

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \quad y = r, r + 1, r + 2, \ldots, 0 \leq p \leq 1.$$

Definition 3.10

A random variable Y is said to have a *hypergeometric probability distribution* if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer $0, 1, 2, \ldots, n$, subject to the restrictions $y \leq r$ and $n - y \leq N - r$.

Permutation

$$P^n_r = \frac{n!}{(n-r)!}$$

Combination

$$C^n_r = \frac{n!}{r!(n-r)!}$$

Bayes Theorem

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

Poisson Distribution

$$P(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

Tchebysheff's Theorem

Let *Y* be a random variable with mean *μ* and finite variance σ^2. Then, for any constant k >0,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad or \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Definition 4.1

Let *Y* denote any random variable. The *distribution function* of *Y*, denoted by *F(y)*, is such that $F(y) = P(Y \leq y)\ for\ -\infty < y < \infty$.

Theorem 4.4

Sometimes you want the expected of a function instead of the random variable.

$$E(g(Y)) = \int_{-\infty}^{\infty} g(y)f(y)dy$$

As long as the integral exists.

Theorem 4.5

In general

Expected of constant is a constant

$$E(c) = c$$

You can factor out constants

$$E[cg(Y)] = cE[g(Y)]$$

You can sum up all values then average, or average each value then sum

$$E[g1(y) + g2(y + \ldots] = E[g1(y)] + E[g2(y)] + \ldots$$

Expected Value

The expected value of a uniform distribution is:

$$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2}$$

Gamma Distribution

A random variable *Y* is said to have a *gamma distribution* with parameters $\alpha > 0$ and $\beta > 0$ if and only if the density function of *Y* is

$$f(y) = \begin{cases} \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta \alpha \Gamma(\alpha)} & 0 \leq y < \infty \\ 0, & elsewhere, \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Expected and Variance of Gamma Distribution

If *Y* has a gamma distribution with parameters α and β, then

$$\mu = E(Y) = \alpha\beta \quad and \quad \sigma^2 = V(Y) = \alpha\beta^2.$$

Exponential Distribution of the Gamma Function

A random variable *Y* is said to have an *exponential distribution* with parameter $\beta > 0$ if and only if the density function of *Y* is

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-\frac{y}{\beta}} & 9 \le y < \infty, \\ 0, & elsewhere. \end{cases}$$

Exponential Distribution is if Alpha = 1

If Y is an exponential random variable with parameter β, then

$$\mu = E(Y) = \beta \quad and \quad \sigma^2 = V(Y) = \beta^2.$$

The proof follows directly from Theorem 4.8 with $\alpha = 1$.

Definition 5.2

For any random variables Y_1 and Y_2 the joint (bivariate) distribution function $F(y_1, y_2)$ is

$$F(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2, \quad -\infty < y_1 < \infty, \ -\infty < y_2 < \infty.$$

Definition 5.4

a Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. Then the *marginal probability functions* of Y_1 and Y_2, respectively, are given by

$$p_1(y_1) = \sum_{all\, y_2} p(y_1, y_2) \quad and \quad p_2(y_2) = \sum_{all\, y_1} p(y_1, y_2).$$

b Let Y_1 and Y_2 be jointly continuous random variables with joint density function $f(y_1, y_2)$. Then the *marginal density functions* of Y_1 and Y_2, respectively, are given by

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \quad and \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

Definition 5.5

If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then the *conditional discrete probability function* of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1=y_1|Y_2=y_2)}{P(Y_2=y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)},$$

Provided that $p_2(y_2) > 0$.

Definition 5.6

If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$, then the *conditional distribution function* of Y_1 given $Y_2 = y_2$ is

$$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2).$$

Definition 5.7

If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density of Y_1 given $Y_2 = y_2$ is given by

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

and, for any y_1 such that $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is given by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}.$$

Definition 5.8

Let Y_1 have distribution function $F_1(y_1)$, Y_2 have distribution function $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be *independent* if and only if

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

for every pair of real numbers $F(y_1, y_2)$.

If Y_1 and Y_2 are not independent, they are said to be *dependent*.