Galo Villacis
Formula Sheet for Probability and Applied Statistics

Definition 1.1

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

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The corresponding population mean is denoted μ.

Standard Deviation 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

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

The corresponding population variance is denoted by the symbol σ2.

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 A, 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 …. Form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then

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

Theorem 2. 1

With m elements $a_1, a_2 ..., a_m$ and n elements $b_1, b_2,..., b_n$, it is possible to form $mn = m \times n$ pairs containing one element from each group.

Verification of the theorem can be seen by observing the rectangular table in Figure 2.9. There is one square in the table for each a_i, b_j pair and hence a total of $m \times n$ squares.

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.

Theorem 2.2 Permutations

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

We are concerned with the number of ways of filling r positions with n distinct objects. Applying the extension of the mn rule, we see that the first object can be chosen in one of n ways. After the first is chosen, the second can be chosen in(n−1) ways, the third in (n−2), and the rth in (n−r +1) ways. Hence, the total number of distinct arrangements is

$P_r^n = n(n-1)(n-2)\cdots$.

Expressed in terms of factorials,

$P_r^n = n(n-1)(n-2)\cdots(n-r+1)\frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$

Where $n! = (n-1)\cdots(2)(1)$ and $0! = 1$.

Theorem 2.3 Multinomial coefficients

$$N = \binom{n}{n_1\, n_2 \cdots n_k} = \frac{n!}{n_1!\; n_2!\; \cdots n_k!}$$

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 $\binom{n}{r}$

Theorem 2.4 Combinations

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!\,(n-r)!}$$

Definition 2.9

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

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

Definition 2.10

A and B are Independent if

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

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

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

Theorem 2.5
The Multiplicative Law of Probability:

The probability of the intersection of two events A and B is

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

If A and B are independent, then

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

Theorem 2.6
The Additive Law of probability:

The probability of the union of two events A and B is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A and B are mutually exclusive events, $P(A \cup B) = 0$

$$P(A \cup B) = P(A) + P(B)$$

Theorem 2.7

If A is an event, then

$$P(A) = 1 - P(\bar{A})$$

Theorem 2.8
Total Probability:

$$P(A) = \sum_{i=1}^{n} P(A|Bi)P(Bi)$$

Theorem 2.9
Bayes' Rule:

$$P(Bi|A) = \frac{P(A|Bj)P(Bj)}{\sum_{i=1}^{k} P(A|Bi)P(Bi)}$$

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 3.2

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 Ycan be represented by a formula, a table, or a graph that provides $p(y)\ =\ P(Y\ =\ y)$ for all $y.$

Theorem 3.1

For any discrete probability distribution, the following must be true:

1. 0≤ p(y)≤1 for all y.
2. $\sum_y p(y)$= 1, where the summation is over all values of y with nonzero probability.

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^2

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

Theorem 3.4

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y . Then the expected value of $g(Y\ )$ is given by

$$E[g(Y)] = \sum_{all\ y} g(y)p(y).$$

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 $V\ (Y\ )$

Theorem 3.3

Let Y be a discrete random variable with probability function $p(y)$ and c be a constant. Then $E(c) = c$.

Theorem 3.4

Let Y be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of Y , and c be a constant. Then

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

Theorem 3.5

Let Y be a discrete random variable with probability function $p(y)$ and $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E[g1(Y) + g2(Y) + \cdots + gk(Y)] = E[g1(Y)] + E[g2(Y)] + \cdots + E[gk(Y)].$$

Theorem 3.6

Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

$$V(Y) = \sigma 2 = E[(Y - \mu)2] = E(Y2) - \mu 2.$$

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

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

Theorem 3.7

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \; and \; \sigma 2 \; = V(Y) = npq.$$