Probability And Statistics Formulas

Taszid Chowdhury

Stockton University

December 112024

Professor Byron

TABLE OF CONTENTS

* [Chapter 1](#_Chapter_1:_What): What is Statistics?
  + [Definition 1.1](#Def_1_1) – Mean
  + [Definition 1.2](#Def_1_2) – Variance
  + [Definition 1.3](#Def_1_3) – Standard Deviation
* [Chapter 2](#_Chapter_2:_Probability): Probability
  + [Definition 2.1](#Def_2_1) – Experiment
  + [Definition 2.2](#Def_2_2) – Simple Event
  + [Definition 2.3](#Def_2_3) – Sample Space
  + [Definition 2.4](#Def_2_4) – Discrete Sample Space
  + [Definition 2.5](#Def_2_5) – Event
  + [Definition 2.6](#Def_2_6) – Probability
  + [Theorem 2.1](#Thm_2_1) – *rule*
  + [Definition 2.7](#Def_2_7) – Permutation
  + [Theorem 2.2](#Thm_2_2) – Permutation Theorem
  + [Theorem 2.3](#Thm_2_3) – Partitioning a set of objects into nonoverlapping groups
  + [Definition 2.8](#Def_2_8) – Combinations
  + [Theorem 2.4](#Thm_2_4) – Combinations Theorem
  + [Definition 2.9](#Def_2_9) – Conditional Probability
  + [Definition 2.10](#Def_2_10) – Independent Events
  + [Theorem 2.5](#Thm_2_5) – The Multiplicative Law of Probability
  + [Theorem 2.6](#Thm_2_6) – The Additive Law of Probability
  + [Theorem 2.7](#Thm_2_7) – Probability of the Complement of an Event
  + [Definition 2.11](#Def_2_11) – Total Probability
  + [Theorem 2.8](#Thm_2_8) – Total Probability Theorem
  + [Theorem 2.9](#Thm_2_9) – Bayes’ Rule
  + [Definition 2.12](#Def_2_12) – Random Variable
  + [Definition 2.13](#Def_2_13) – Random Sample
* [Chapter 3](#_Chapter_3:_Discrete): Discrete Random Variables and Their Probability Distributions
  + [Definition 3.1](#Def_3_1) – Discrete Random Variable
  + [Definition 3.2](#Def_3_2) – Probability of a Random Variable
  + [Definition 3.3](#Def_3_3) – Probability Distribution of a Random Variable
  + [Theorem 3.1](#Thm_3_1) – Properties of a Discrete Probability Distribution
  + [Definition 3.4](#Def_3_4) – Expected Value of a Discrete Random Variable
  + [Theorem 3.2](#Thm_3_2) – Expected Value of a Function of a Discrete Random Variable
  + [Definition 3.5](#Def_3_5) – Variance of a Discrete Random Variable
  + [Theorem 3.3](#Thm_3_3) – Expected Value of a Constant
  + [Theorem 3.4](#Thm_3_4) – Expected Value of a Constant Times a Function of a Random Variable
  + [Theorem 3.5](#Thm_3_5) – Expected Value of a Sum of Functions of a Random Variable
  + [Theorem 3.6](#Thm_3_6) – Expected Value of a Discrete Random Variable Theorem
  + [Definition 3.6](#Def_3_6) – Binomial Experiment Properties
  + [Definition 3.7](#Def_3_7) – Binomial Distribution
  + [Theorem 3.7](#Thm_3_7) – Expected and Variance of a Binomial Distribution
  + [Definition 3.8](#Def_3_8) – Geometric Distribution
  + [Theorem 3.8](#Thm_3_8) – Expected and Variance of a Geometric Distribution
  + [Definition 3.10](#Def_3_10) – Hypergeometric Distribution
  + [Theorem 3.10](#Thm_3_10) – Expected and Variance of a Hypergeometric Distribution
  + [Definition 3.11](#Def_3_11) – Poisson Probability Distribution
  + [Theorem 3.11](#Thm_3_11) – Expected and Variance of a Hypergeometric Distribution
  + [Theorem 3.14](#Thm_3_14) – Tchebysheff’s Theorem
* [Chapter 4](#_Chapter_4:_Continuous): Continuous Variables and Their Probability Continuous Distributions
  + [Definition 4.1](#Def_4_1) – Distribution Function of a Continuous Random Variable
  + [Theorem 4.1](#Thm_4_1) – Properties of a Distribution Function
  + [Definition 4.2](#Def_4_2) – Continuous Distribution Function Definition
  + [Definition 4.3](#Def_4_3) – Probability Density Function
  + [Theorem 4.2](#Thm_4_2) – Properties of a Density Function
  + [Definition 4.4](#Def_4_4) – *p*th quantile of
  + [Theorem 4.3](#Thm_4_3) – Integral of a Density Function
  + [Definition 4.5](#Def_4_5) – Expected Value of a Continuous Random Variable
  + [Theorem 4.4](#Thm_4_4) – Expected Value of a Function of a Continuous Random Variable
  + [Theorem 4.5](#Thm_4_5) – Expected Values of Three Important Functions of a Continuous Random Variable
  + [Definition 4.6](#Def_4_6) – Continuous Uniform Probability Distribution
  + [Definition 4.7](#Def_4_7) – Parameters of a Density Function
  + [Theorem 4.6](#Thm_4_6) – Expected and Variance of a Uniformly Distributed Random Variable
  + [Definition 4.8](#Def_4_8) – Normal Probability Distribution
  + [Theorem 4.7](#Thm_4_7) – Expected and Variance of a Normal Probability Distribution
  + [Definition 4.9](#Def_4_9) – Gamma Distribution
  + [Theorem 4.8](#Thm_4_8) – Expected and Variance of a Gamma Distribution
  + [Definition 4.10](#Def_4_10) – Chi-Square Distribution and Degrees of Freedom
  + [Theorem 4.9](#Thm_4_9) – Expected and Variance of a Chi-Square Random Variable with Degrees of Freedom
  + [Definition 4.11](#Def_4_11) – Exponential Distribution
  + [Theorem 4.10](#Thm_4_10) – Expected and Variance of an Exponential Distribution
  + [Definition 4.12](#Def_4_12) – Beta Probability Distribution
  + [Theorem 4.11](#Thm_4_11) – Expected and Variance of a Beta Distribution
  + [Definition 4.13](#Def_4_13) – Expected of a Continuous Random Variable with th Moment About the Origin
  + [Definition 4.14](#Def_4_14) – Moment-Generating Function of a Continuous Random Variable
  + [Theorem 4.12](#Thm_4_12) – Finding the Moments of a Function of a Random Variable
  + [Theorem 4.13](#Thm_4_13) – Tchebysheff’s Theorem
  + [Definition 4.15](#Def_4_15) - Expectations of Random Variables with Mixed Distributions
* [Chapter 5](#_Chapter_5:_Multivariate): Multivariate Probability Distributions
  + [Definition 5.1](#Def_5_1) – Bivariate Probability Function
  + [Theorem 5.1](#Thm_5_1) – Properties of Bivariate Functions
  + [Definition 5.2](#Def_5_2) – Bivariate Probability Distribution Function
  + [Definition 5.3](#Def_5_3) – Bivariate Probability Density Function
  + [Theorem 5.2](#Thm_5_2) – Properties of Bivariate Probability Density Function
  + [Theorem 5.3](#Thm_5_3) – More Properties of Bivariate Probability Density Function

# Chapter 1: What Is Statistics?

Definition 1.1)

The *mean* of a sample of n measured responses

The corresponding population is denoted *µ*.

Definition 1.2)

The *variance* of a sample of measurements

The corresponding *population* is variance is denoted by the symbol .

Definition 1.3)

The *standard deviation* of a sample of measurements

The corresponding *population* standard deviation is denoted by .

# Chapter 2: Probability

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*.

A simple event *E* with the subscript *x* where *x* will start at 1 and for every new simple event will be incremented by one.

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*) ≥ 0.

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

Axiom 3: If *A1, A2, A3*, … form a sequence of pairwise mutually exclusive events in *S* (that is, *Ai* ∩ *Aj =* ∅ if i = j), then

Theorem 2.1)

With *m* elements and *n* elements b1, b­2, …, bn, it is possible to form pairs containing one element from each group.

Definition 2.7)

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

Theorem 2.2)

Proof: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 *r*th in the (*n – r +* 1­) ways. Hence, the total number of distinct arrangements is

.

Expressed in terms of factorials,

Where *n*! = and 0! = 1.

Theorem 2.3)

The number of ways of partitioning *n* distinct objects into *k* distinct groups containing objects, respectively, where each object appears in exactly one group and , is

Proof: N is the number of distinct arrangements of *n* objects in a row for a case in which rearrangement of the objects within a group does not count. For example, the letters *a* to *l* are arranged in three groups, where , and :

Is one such arrangement.

The number of distinct arrangements of the n objects, assuming all objects are distinct, is (from Theorem 2.2). Then equals the number of ways of partitioning the *n* objects into *k* groups (ignoring order within groups) multiplied by the number of ways of ordering elements withing each group. This application of the extended *mn* rule gives

where is the number of distinct arrangements of the objects in the group *i*.

Solving for *N*, we have

Definition 2.8)

The number of *combinations* of *n* objects take *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 or

Theorem 2.4)

The number of unordered subsets of size *r* chosen (without replacement) from *n* available objects is

.

Proof: The selection of *r* objects from a total of *n* is equivalent to partitioning the *n* objects from *k =* 2 groups, the *r* selected, and the (n – r) remaining. This is a special case of the general partitioning problem dealt with in Theorem 2.3. In the present case, k = 2, , and and therefore,

Definition 2.9)

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

Providing [The symbol 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:

,

,

.

Otherwise, the events are said to be *dependent*.

Theorem 2.5)

The Multiplicative Law of Probability The probability of the intersection of two events *A* and *B* is

.

If *A* and *B* are independent, then

.

Proof: The multiplicative law follows directly from Definition 2.9, the definition of conditional probability.

Theorem 2.7)

If *A* is an event, then

.

Proof: Observe that . Because and are mutually exclusive events, it follows that . Therefore, and the result follows.

Definition 2.11)

For some positive integer *k*, let the sets such that

1. .

2. , for .

Then the collection of sets {} is said to be a *partition* of S.

Theorem 2.8)

Assume that {} is a partition of *S* (see Definition 2.11) such that , for

. Then for any event *A*

.

Proof: Any subset *A* of *S* can be written as

= (.

Notice that, because {} is a partition of S, if ,

and that and are mutually exclusive events. Thus,

.

Theorem 2.9)

Bayes’ Rule: Assume that {} is a partition of *S* (see Definition 2.11) such that , for . Then

Proof: The proof the follows directly from the definition of conditional probability and the law of total probability. Note that

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 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.*

Chapter 3: Discrete Random Variables and Their Probability Distributions

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 of *Y* takes on the value *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 by

Definition 3.3)

The *probability distribution* for a discrete variable *Y* can be represented by a formula, a table, or a graph that provides for all *y*.

Theorem 3.1)

For any discrete probability, the following must be true:

1. for all *y*.

2. , 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 . Then the *expected value* of *Y*, , is defined to be

Theorem 3.2)

Let *Y* be a discrete random variable with probability function and be a real-valued function of . Then the expected value of is given by

Proof: We prove the result in the case where the random variable takes on the finite number of values . Because the function may not be one to-one, suppose that is a random variable such that for

Thus, by definition 3.4,

Definition 3.5)

If is a random variable with mean , the variance of a random variable is defined to be the expected value of . That is,

.

The *standard deviation* of is the positive square root of .

Theorem 3.3)

Let be a discrete random variable with probability function and be a constant. Then .

Proof: Consider the function . By Theorem 3.2,

But (Theorem 3.1) and, hence,

Theorem 3.4)

Let be a discrete random variable with probability function be a function of , and be a constant. Then

Proof: By Theorem 3.2

Theorem 3.5)

Let be a discrete random variable with probability function and be a function of . Then

Proof: We will demonstrate the proof only for the case , but analogous steps will hold for any finite . By Theorem 3.2,

Theorem 3.6)

Let be a discrete random variable with probability function and mean ; then

Proof:

(by Theorem 3.5)

Noting that is a constant and applying Theorems 3.4 and 3.3 to the second and third terms, respectively, we have

But and, therefore,

Definition 3.6)

A *binomial experiment* possesses the following properties:

1. The experiment consists of a fixed number, , of identical tries.

2. Each trial results in one of two outcomes: success, , or failure, .

3. The probability of success on a single trial is equal to some value and remains the same form trial to trial. The probability of a figure is equal to .

4. The trials are independent.

5. The random variable of interest is , the number of successes observed during the trials.

Definition 3.7)

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

WHEN TO USE: Binomial distribution should be used when there are exactly two mutually exclusive outcomes. This means that there is either a success or failure and an appropriate likelihood for both. Not only this but you are either looking for the amount of either success or failures given trials.

Theorem 3.7)

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

and

Definition 3.8)

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

WHEN TO USE: Should be used when you are trying to find the likelihood of success given a limited number of trials. At the th trial is when the success occurs. The resulting probability is a success at the th run.

Theorem 3.8)

If is a random variable with a geometric distribution,

and

Definition 3.10)

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

where is an integer 0, 1, 2, …, , subject to the restrictions and

WHEN TO USE: Used when the conditional probability of success and failure are not independent. Also used when you want to determine the probability of obtaining a certain number of successes without replacement from a specific group size.

Theorem 3.10)

If is a random variable with a hypergeometric distribution,

and

Definition 3.11)

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

WHEN TO USE: The Poisson distribution should be used when you are looking for the probability distribution of the number of rare events that occur in space, time, volume, or any other dimension where is the average value of .

Theorem 3.11)

If is a random variable possessing a Poisson distribution with parameter , then

and

Theorem 3.14)

Tchebysheff’s Theorem Let be a random variable with mean and finite variance . Then, for any constant

or

WHEN TO USE: Can be used for any probability distribution where you are looking to determine a lower bound for the probability that the random variable of interest falls in an interval .

In other words, it is used to find the probability of an event occurring within a certain number of standard deviations from the average.

# Chapter 4: Continuous Variables and

Definition 4.1)

Let denote any random variable. The *distribution function* of , denoted by , is such that

for

Theorem 4.1) Properties of a Distribution Function

If is a distribution function, then

1. .

2. .

3. is a nondecreasing function . [If and are *Ant* values such that , then

Definition 4.2)

A random variable with distribution function is said to be continuous if is continuous, for .

Definition 4.3)

Letbe the distribution function for a continuous random variable . Then , is given by

Wherever the derivative exists, is called the *probability density function* for the random variable.

Theorem 4.2) Properties of a Density Function

If is a density function for a continuous random variable, then

1. for all

2.

Definition 4.4)

Let denote any random variable, if the *p*th *quantile* of , denoted by , is the smallest value such that If is continuous, is the smallest value such that . Some prefer to call the 100 *p*th *percentile* of Y.

Theorem 4.3)

If the random variable has density function and , then the probability that falls in the interval is

­

Definition 4.5)

The expected value of a continuous random variable is

Provided that the integral exists.

Theorem 4.4)

Let be a function of ; then the expected value of is given by

Provided that the integral exists.

Theorem 4.5)

Let c be a constant that let be functions of a continuous function variable . Then the following results hold:

1.

2. .

3. .

Definition 4.6)

If , a random variable is said to have a continuous *uniform probability distribution* on the interval if and only if the density function of is

Definition 4.7)

The constants that determine the specific form of a density function are called *parameters* of the density function

Theorem 4.6)

If and is a random variable uniformly distributed on the interval , then

and .

Proof

By Definition 4.5,

Definition 4.8)

A random variable is said to have a *normal probability distribution* if and only if, for and the density function of is

Theorem 4.7)

If is a normally distributed random variable with parameters and , then

and

Definition 4.9)

A random variable is said to have a *gamma distribution with parameters* and if and only if the density function of is

where

Theorem 4.8)

If has a gamma distribution with parameters and , then

= and .

Definition 4.10)

Let be a positive integer. A random variable is said to have a *chi-square distribution* *with v degrees of freedom* if and only if is a gamma-distributed random variable with parameters .

Theorem 4.9)

If is a chi-square random variable with degrees of freedom, then

and .

Definition 4.11)

A random variable is said to have an *exponential distribution with parameter*  if and only if the density function of is

Theorem 4.10)

If is an exponential random variable with parameter then

and .

Definition 4.12)

A random variable is said to have a *beta probability distribution with parameter* and if and only if the density function of is

where

Theorem 4.11)

If is a beta-distributed random variable with parameters and then

|  |  |  |
| --- | --- | --- |
|  | and |  |

Definition 4.13)

If is a continuous random variable, then the th *moment about the origin* is given by

The ­th *moment about the mean,* or the th *central moment,* is given by

Definition 4.14)

If is a continuous random variable, then the *moment-generating function of* is given by

The moment-generating function is said to exist if there exists a constant such that is finite for .

Theorem 4.12)

Let be a random variable with density function and be a function of . Then the moment-generating function for is

Theorem 4.13) Tchebysheff’s Theorem

Let be a random variable with finite mean and variance . Then, for any ,

|  |  |  |
| --- | --- | --- |
|  | or |  |

Definition 4.15)

Let have the mixed distribution function

and suppose that is a discrete random variable with distribution function and that is a continuous random variable with distribution function . Let g(Y) denote a function of . Then

# Chapter 5: Multivariate Probability Distributions

Definition 5.1)

Let and be discrete random variables. The *join* (or bivariate) *probability function* for and is given by

Theorem 5.1)

If and be discrete random variables with joint probability function , then

1. for all .
2. where the sum is over all values ( that are assigned nonzero probabilities.

Definition 5.2)

For any random variables and , the joint (bivariate) distribution function is

Definition 5.3)

Let and be continuous random variables with join distribution function . If there exists a nonnegative function such that

for all , then and are said to be *jointly continuous random variables*. The function is called the *joint probability density function.*

Theorem 5.2)

If and are random variables with join distribution function , then

1. If and , then

Theorem 5.3)

If and are jointly continuous random variables with a join density function given by then

1. for all .