**Formula Sheet**

**Definition 1.2**

The variance of a sample of measurements y1, y2,..., yn is the sum of the square of the differences between the measurements and their mean, divided by n − 1. Symbolically, the sample variance is

**Definition 1.3**

The standard deviation of a sample of measurements is the positive square root

of the variance; that is,

**Chapter 2**

**Definitions Chapter 2**

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

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.

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.

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

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

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:**

**Axiom 2:**

**Axiom 3: If A1, A2, A3, . . . form a sequence of pairwise mutually exclusive events in S (that is, if ) then**

**Sample Point Method**

The sample-point method is outlined in Section 2.4. The following steps are used to find the probability of an event:

1. Define the experiment and clearly determine how to describe one simple event.

List the simple events associated with the experiment and test each to make certain that it cannot be decomposed. This defines the sample space S.

Assign reasonable probabilities to the sample points in S, making certain

that P(Ei)≥0 and   
4. Define the event of interest, A, as a specific collection of sample points.

(A sample point is in A if A occurs when the sample point occurs. Test all

sample points in S to identify those in A.)  
5. Find P(A) by summing the probabilities of the sample points in A.

**Theorem 2.1**

With *m* elements *a*1, *a*2,..., *am* and *n* elements *b*1, *b*2,..., *bn*, it is possible to form *mn* = *m* × *n* pairs containing one element from each group.

Verification of the theorem can be seen by observing the rectangular table in Figure2.9.Thereisonesquareinthetableforeach*ai*,*bj* pairandhenceatotal of *m* × *n* squares.

**Definition 2.7**

An ordered arrangement of *r* distinct objects is called a *permutation*. The num- ber of ways of ordering *n* distinct objects taken *r* at a time will be designated by the symbol *Prn*.

**Theorem 2.2**

*Prn* =*n*(*n*−1)(*n*−2)···(*n*−*r*+1)= *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,thethirdin(*n*−2),andthe*r*thin(*n*−*r* +1)ways.Hence, the total number of distinct arrangements is

*Prn* =*n*(*n*−1)(*n*−2)···(*n*−*r*+1). Expressed in terms of factorials,

*Prn* =*n*(*n*−1)(*n*−2)···(*n*−*r*+1)(*n*−*r*)!= (*n* − *r* )!

where *n*! = *n*(*n* − 1)···(2)(1) and 0! = 1.

*n*!  
(*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) =

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 ∩ B) = P(A)P(B). 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

P(A ∩ B) = P(A)P(B|A)

= P(B)P(A|B).

If A and B are independent, then  
P(A ∩ B) = P(A)P(B).

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

**Theorem 2.6**

The Additive Law of Probability The probability of the union of two events A and B is

P(A ∪ B) = P(A) + P(B) − P(A ∩ B).

If A and B are mutually exclusive events, P(A ∩ B) = 0 and

P(A ∪ B) = P(A) + P(B).

Theorem 2.7

If A is an event, then

P(A) = 1 − P(A).

Observe that S = A ∪ A. Because A and A are mutually exclusive events, it follows that P(S) = P(A) + P(A). Therefore, P(A) + P(A) = 1 and the result follows.

**Definition 2.11**

For some positive integer k, let the sets B1, B2, . . . , Bk be such that 1. S=B1∪B2∪···∪Bk.

2. Bi∩Bj=∅,fori≠ j.  
Then the collection of sets {B1, B2,..., Bk} is said to be a partition of S

**Theorem 2.8**

Assume that {*B*1, *B*2, . . . , *Bk* } is a partition of *S* (see Definition 2.11) such that *P*(*Bi*) > 0, for *i* = 1,2,...,*k*. Then for any event *A*

*k i*=1

Any subset *A* of *S* can be written as  
*A* = *A* ∩ *S* = *A* ∩ ( *B*1 ∪ *B*2 ∪ · · · ∪ *Bk* )

= ( *A* ∩ *B*1 ) ∪ ( *A* ∩ *B*2 ) ∪ · · · ∪ ( *A* ∩ *Bk* ). Notice that, because {*B*1, *B*2,···, *Bk*} is a partition of *S*, if *i* ≠ *j*, (*A*∩*Bi*)∩(*A*∩*Bj*)= *A*∩(*Bi* ∩*Bj*)= *A*∩∅=∅

and that (*A* ∩ *Bi*) and (*A* ∩ *Bj*) are mutually exclusive events. Thus, *P*(*A*)= *P*(*A*∩*B*1)+*P*(*A*∩*B*2)+···+*P*(*A*∩*Bk*)

= *P*(*A*|*B*1)*P*(*B*1) + *P*(*A*|*B*2)*P*(*B*2) + ··· + *P*(*A*|*Bk*)*P*(*Bk*)

*k*= *P*(*A*|*Bi*)*P*(*Bi*).

**Theorem 2.9**

**Bayes’ Rule** Assume that {*B*1, *B*2, . . . , *Bk* } is a partition of *S* (see Definition 2.11) such that *P*(*Bi*) > 0, for *i* = 1,2,...,*k*. Then

*P*(*Bj*|*A*) = *P*(*A*|*Bj*)*P*(*Bj*) .

*k  
P*(*A*|*Bi*)*P*(*Bi*)

*i*=1  
The proof follows directly from the definition of conditional probability and

the law of total probability. Note that

*P*(*Bj*|*A*) = *P*(*A* ∩ *Bj*) = *P*(*A*|*Bj*)*P*(*Bj*) .

*P* ( *A* ) *k i*=1

*P*(*A*|*Bi*)*P*(*Bi*)

**Definition 2.12**

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

**Definitions 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

**Definition 3.5**

If Y is a random variable with mean E(Y ) = µ, the variance of a random variable Y is defined to be the expected value of (Y − µ)2. That is,

**Definition 3.6**

A binomial experiment possesses the following properties:

The experiment consists of a fixed number, n, of identical trials.

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

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).

The trials are independent.

The random variable of interest is Y , the number of successes observed

during the n trials.

**Definition 3.8**

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

*p*(*y*)=*qy*−1*p*, *y*=1,2,3,..., 0≤*p*≤1.

**Definition 3.9**

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

*y* − 1  
*p*(*y*)= *r*−1 *prqy*−*r*, *y*=*r*,*r*+1,*r*+2,...,0≤*p*≤1

**Chapter 4**

**Definition 4.1**

Let *Y* denote any random variable. The *distribution function* of *Y* , denoted by *F*(*y*),issuchthat *F*(*y*)= *P*(*Y* ≤ *y*)for−∞< *y* <∞.

**Theorem 4.1**

**Properties of a Distribution Function**1 If *F*(*y*) is a distribution function, then

1. *F*(−∞) ≡ lim *F*(*y*) = 0.

*y*→−∞

1. *F*(∞) ≡ lim *F*(*y*) = 1.

*y*→∞

1. *F*(*y*) is a nondecreasing function of *y*. [If *y*1 and *y*2 are *any* values such

that *y*1 < *y*2, then *F*(*y*1) ≤ *F*(*y*2).]

**Definition 4.2**

A random variable *Y* with distribution function *F*(*y*) is said to be *continuous* if *F*(*y*) is continuous, for −∞ < *y* < ∞.2

**Theorem 4.2**

**Properties of a Density Function** If *f*(*y*)is a density function for a continuous random variable, then

1. *f*(*y*)≥0for all *y*,−∞<*y*<∞.

2. #∞ *f*(*y*)*dy*=1

**Theorem 4.3**

If the random variable *Y* has density function *f* (*y*) and *a* < *b*, then the proba- bility that *Y* falls in the interval [*a*, *b*] is

**Definition 4.4**

Let *Y* denote any random variable. If 0 < *p* < 1, the *p*th *quantile* of *Y*, denoted by φ*p*, is the smallest value such that *P*(*Y* ≤ φ*q*) = *F*(φ*p*) ≥ *p*. If *Y* is continuous, φ*p* is the smallest value such that *F*(φ*p*) = *P*(*Y* ≤ φ*p*) = *p*. Some prefer to call φ*p* the 100*p*th *percentile* of *Y*.

**Theorem 4.4**

Let *g*(*Y* ) be a function of *Y* ; then the expected value of *g*(*Y* ) is given by

**Definition 4.5**

The expected value of a continuous random variable *Y* is

**Theorem 4.5**

Let *c* be a constant and let *g*(*Y*), *g*1(*Y*), *g*2(*Y*),...,*gk*(*Y*) be functions of a continuous random variable *Y* . Then the following results hold:

1. *E*(*c*) = *c*.
2. *E*[*cg*(*Y*)] = *cE*[*g*(*Y*)].
3. *E*[*g*1(*Y*)+*g*2(*Y*)+···+*gk*(*Y*)] = *E*[*g*1(*Y*)]+*E*[*g*2(*Y*)]+···+*E*[*gk*(*Y*)].

**Definition 4.6**

If θ1 < θ2, a random variable *Y* is said to have a continuous *uniform probability distribution* on the interval (θ1, θ2) if and only if the density function of *Y* is

**Definition 4.7**

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

**Chapter 5**

**Definition 5.1**

Let *Y*1 and *Y*2 be discrete random variables. The *joint* (or bivariate) *probability function* for *Y*1 and *Y*2 is given by

*p*(*y*1,*y*2)= *P*(*Y*1 = *y*1,*Y*2 = *y*2), −∞< *y*1 <∞,−∞< *y*2 <∞.

**Theorem 5.1**

If *Y*1 and *Y*2 are discrete random variables with joint probability function *p*(*y*1, *y*2), then

*p*(*y*1, *y*2) = 1,

where the sum is over all values (*y*1, *y*2) that are assigned nonzero probabilities.

1. *p*(*y*1, *y*2) ≥ 0 for all *y*1, *y*2.

**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 ≤ *y*1,*Y*2 ≤ *y*2), −∞< *y*1 <∞,−∞< *y*2 <∞.

**Theorem 5.2**

If *Y*1 and *Y*2 are random variables with joint distribution function *F* ( *y*1 , *y*2 ), then

1. *F*(−∞, −∞) = *F*(−∞, *y*2) = *F*(*y*1, −∞) = 0. 2. *F*(∞,∞)=1.  
3. If *y*1∗ ≥*y*1 and *y*2∗ ≥*y*2,then

*F*(*y*1∗, *y*2∗) − *F*(*y*1∗, *y*2) − *F*(*y*1, *y*2∗) + *F*(*y*1, *y*2) ≥ 0.

If *Y*1 and *Y*2 are jointly continuous random variables with a joint density function given by *f* (*y*1, *y*2), then

1. *f*(*y*,*y*)≥0forall*y*,*y*. # 1# 2 1 2

2. ∞ ∞ *f*(*y*1,*y*2)*dy*1*dy*2 =1.

**Definition 5.3**

Let *Y*1 and *Y*2 be continuous random variables with joint distribution function *F*(*y*1, *y*2). If there exists a nonnegative function *f* (*y*1, *y*2), such that

For all−∞< *y*1 <∞,−∞< *y*2 <∞,then *Y*1 and *Y*2 are said to be *jointly*

*continuous random variables*. The function *f* (*y*1, *y*2) is called the *joint prob- ability density function*.

**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 ≤ *y*1|*Y*2 = *y*2).

**Definition 5.7**

Let *Y*1 and *Y*2 be jointly continuous random variables with joint density *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)

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

**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 (*y*1, *y*2).  
If *Y*1 and *Y*2 are not independent, they are said to be *dependent*.

**Theorem 5.4**

If *Y*1 and *Y*2 are 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 *Y*1 and *Y*2 are independent if and only if  
*p*(*y*1, *y*2) = *p*1(*y*1)*p*2(*y*2)

for all pairs of real numbers (*y*1, *y*2).

If *Y*1 and *Y*2 are continuous random variables with joint density function *f* ( *y*1 , *y*2 ) and marginal density functions *f*1(*y*1) and *f*2(*y*2), respectively, then *Y*1 and *Y*2 are independent if and only if

*f* (*y*1, *y*2) = *f*1(*y*1) *f*2(*y*2) for all pairs of real numbers (*y*1, *y*2).

**Theorem 5.5**

Let *Y*1 and *Y*2 have a joint density *f* (*y*1, *y*2) that is positive if and only if *a*≤*y*1 ≤*b*and*c*≤*y*2 ≤*d*,forconstants*a*,*b*,*c*,and*d*;and *f*(*y*1,*y*2)=0 otherwise. Then *Y*1 and *Y*2 are independent random variables if and only if

*f* (*y*1, *y*2) = *g*(*y*1)*h*(*y*2)

where *g*(*y*1) is a nonnegative function of *y*1 alone and *h*(*y*2) is a nonnegative function of *y*2 alone.

**Theorem 5.6**

Let *c* be a constant. Then

*E*(*c*) = *c*.

**Theorem 5.7**

Let *g*(*Y*1,*Y*2) be a function of the random variables *Y*1 and *Y*2 and let *c* be a constant. Then

*E*[*cg*(*Y*1, *Y*2)] = *cE*[*g*(*Y*1, *Y*2)].

**Theorem 5.8**

Let *Y*1 and *Y*2 be random variables and *g*1(*Y*1,*Y*2),*g*2(*Y*1,*Y*2),...,*gk*(*Y*1,*Y*2) be functions of *Y*1 and *Y*2. Then

*E* [*g*1 (*Y*1 , *Y*2 ) + *g*2 (*Y*1 , *Y*2 ) + · · · + *gk* (*Y*1 , *Y*2 )]  
= *E*[*g*1(*Y*1,*Y*2)]+ *E*[*g*2(*Y*1,*Y*2)]+···+ *E*[*gk*(*Y*1,*Y*2)].

**Theorem 5.9**

Let *Y*1 and *Y*2 be independent random variables and *g*(*Y*1) and *h*(*Y*2) be functions of only *Y*1 and *Y*2, respectively. Then

*E*[*g*(*Y*1)*h*(*Y*2)] = *E*[*g*(*Y*1)]*E*[*h*(*Y*2)], provided that the expectations exist.

**Definition 5.10**

If *Y*1 and *Y*2 are random variables with means μ1 and μ2, respectively, the *covariance* of *Y*1 and *Y*2 is

Cov(*Y*1, *Y*2) = *E* [(*Y*1 − μ1)(*Y*2 − μ2)] .

**Theorem 5.10**

If *Y*1 and *Y*2 are random variables with means μ1 and μ2, respectively, then Cov(*Y*1, *Y*2) = *E* [(*Y*1 − μ1)(*Y*2 − μ2)] = *E*(*Y*1*Y*2) − *E*(*Y*1)*E*(*Y*2).

Cov(*Y*1, *Y*2) = *E* [(*Y*1 − μ1)(*Y*2 − μ2)]  
= *E*(*Y*1*Y*2 − μ1*Y*2 − μ2*Y*1 + μ1μ2).

From Theorem 5.8, the expected value of a sum is equal to the sum of the expected values; and from Theorem 5.7, the expected value of a constant times a function of random variables is the constant times the expected value. Thus,

Cov(*Y*1, *Y*2) = *E*(*Y*1*Y*2) − μ1 *E*(*Y*2) − μ2 *E*(*Y*1) + μ1μ2. Because *E*(*Y*1) = μ1 and *E*(*Y*2) = μ2, it follows that

Cov(*Y*1, *Y*2) = *E*(*Y*1*Y*2) − *E*(*Y*1)*E*(*Y*2) = *E*(*Y*1*Y*2) − μ1μ2.

**Theorem 5.11**

If *Y*1 and *Y*2 are independent random variables, then Cov(*Y*1, *Y*2) = 0.

Thus, independent random variables must be uncorrelated.

Theorem 5.10 establishes that  
Cov(*Y*1, *Y*2) = *E*(*Y*1*Y*2) − μ1μ2.

Because *Y*1 and *Y*2 are independent, Theorem 5.9 implies that *E*(*Y*1*Y*2) = *E*(*Y*1)*E*(*Y*2) = μ1μ2,

and the desired result follows immediately.

**Definition 5.11**

A *multinomial experiment* possesses the following properties:

1. The experiment consists of *n* identical trials.
2. The outcome of each trial falls into one of *k* classes or cells.
3. The probability that the outcome of a single trial falls into cell *i*, is *pi*,

*i* = 1,2,...,*k* and remains the same from trial to trial. Notice that *p*1+*p*2+*p*3+···+*pk* =1.

1. The trials are independent.
2. The random variables of interest are *Y*1 , *Y*2 , . . . , *Yk* , where *Yi* equals

the number of trials for which the outcome falls into cell *i*. Notice that *Y*1 +*Y*2 +*Y*3 +···+*Yk* =*n*.