Statistics

Your actual assignment that's due, a part of your midterm project, is a formula sheet containing all of the formulas in the text up to our current topics.

You can tell it's something you write an entry for, if there is a gray box, surrounding a formula.

If the grey box, does not have a formula, you do not need it on the stats sheet.

Definition 1.1

The mean of a sample

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

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 $n - 1$. Systematically, the sample variance is

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

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 *populations* standard deviation is denoted by $\sigma = \sqrt{\sigma^2}$.

Definition 2.6:

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

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

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

$Axiom\ 3: If\ A_1, A_2, A_3, \ldots from\ a\ sequence\ of\ pairwise\ mutually\ exclusive\ events\ in\ S\ \Big(that\ is, A_i \cap A_j$

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

Theorem 2.2

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

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 the *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 $(n-r+1)$. Hence, the total number of distinct arrangements is

$$P_r^n = n(n-1)(n-2)\ldots(n-r+1).$$

Expressed in terms of factorials,

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

Where $n! = n(n-1)\ldots(2)(1)\text{and } 0! = 1$.

Theorem 2.3

The number of ways of partitioning *n* distinct objects into *k* distinct groups containing $n_1, n_2, \ldots, n_k$ objects respectively, where each object appears in exactly one group and $\sum_{i=1}^{k} n_i = n$, is

$$N = \binom{n}{n_1\ n_2\ \ldots\ n_k} = \frac{n!}{n_1!\ n_2!\ \ldots\ n_k!}.$$

N is the number of distinct arrangement of n objects in a row for a case in which rearrangements of the objects within a group does not count. For example, the letters *a* to *l* are arranges in three groups, where $n_1 = 3, n_2 = 4, and\ n_3 = 5$:

$$abc|defg|hijkl$$

Is one such arrangement.

The number of distinct arrangements of the *n* objects, assuming all objects are distinct, is $P_n^n = n!$ (from Theorem 2.2). Then P_n^n 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 the $n_1, n_2, \ldots, n_k$ elements within each group. This application of the extended mn rule gives

$$P_n^n = (N)\cdot(n_1!\ n_2!\, n_3!\ldots n_k!),$$

Where n_1! is the number of distinct arrangements of the n_i objects in group *i*.

Solving for *N, we have*

$$N = \frac{n!}{n_1!\, n_2!\ \ldots n_k!} = \binom{n}{n_1 n_2 \ldots n_k}.$$

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

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

The selecting of *r* objects from a total of *n* is equivalent to partitioning the *n* objects into k = 2 groups, the *r* selected, and the *(n – r)* and, therefore,

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

Definition 2.9

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

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

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.*

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

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

This 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

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

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

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

The proof of the additive law can be followed by inspecting the Venn diagram in Figure 2.10.

Notice that $A \cup B = A \cup (\bar{A} \cap B)$, where *A* and $(\bar{A} \cap B)$ are mutually exclusive events. Further, $B = (\bar{A} \cap B) \cup (A \cap B)$, where $(A \cap B)$ and $(A \cap B)$ are mutually exclusive events. Then by Axiom 3,

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) \text{ and } P(B) = P(\bar{A} \cap B) + P(A \cap B).$$

The equality given on the right implies that $P(\bar{A} \cap B) = P(A) - P(A \cap B)$. Substituting this expression for $P(\bar{A} \cap B)$ into the expression for $P(A \cup B)$ given in the left-hand equation of the preceding pair, we obtain the desired result:

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

Theorem 2.7

If *A* is an event, then

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

Observe that $S = A \cup \bar{A}$. Because *A* and $\bar{A}$ are mutually exclusive events, it follows that $P(S) = P(A) + P(\bar{A}) = 1$ and the result follows.

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. $B_i \cap B_j = 0$, for *I ≠ j*.

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

Theorem 2.8
Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of *S* (see Definition 2.11) such that $P(B_i) > 0$, for *I* = 1, 2, …, *k.* Then for any event *A*

$$P(A) = \sum_{i=1}^{k} P(A|B_I)P(B_{i)}.$$

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

$$A = A \cap S = A \cap (B_1 \cup B_1 \cup \ldots \cup B_k)$$
$$= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k).$$

Notice that, because $\{B_1, B_2, \ldots, B_k\}$ is a partition of *S,* if *i ≠ j,*

$$(A \cap B_i) \cap \left(A \cap B_j\right) = A \cap \left(B_i \cap B_j\right) = A \cap 0 = 0$$

And that $(A \cap B_i)$ and $(A \cap B_j)$ are mutually exclusive events. Thus,

$$P(A) = P(A \cap B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)$$
$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)$$
$$= \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

Theorem 2.9
Baye's Rule Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of *S* (see Definition 2.11) such that $P(B_i) > 0, \text{for } i = 1, 2, \ldots, k$. Then

$$P\left(B_j\middle|A\right) = \frac{P\left(A\middle|B_j\right)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

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

$$P\left(B_j\middle|A\right) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Theorem 3.1
For any discrete probability distribution, the following must be true:
$1.\ 0 \le p(y) \le 1$ 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]

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

Theorem 3.2
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

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

Proof: We prove the result in the case where the random variable *Y* takes on the finite number of values $y_1, y_2, \ldots, y_n$. Because the function *g(y)* may not be one to-one, suppose that *g(Y)* takes on values $g_1, g_2, \ldots, g_m$ (where *m ≤ n).* It follows that *g(Y)* is a random variable such that for *i = 1,2, … , m,*

$$P[g(Y) = \ g_i] = \sum_{all\ y_1 such\ that\ g(y_j)=g_i} p(y_i) = p^*(g_i),$$

Thus, by Definition 3.4,

$$\begin{aligned} E[g(Y))] &= \sum_{i=1}^{,} g_i p^*(g_i) \\ &= \sum_{i=1}^{m} g_i \left\{ \sum_{all\ y_j\ such\ that\ g(y_{j\_}=g_i} p(y_i) \right\} \\ &= \sum_{i=1}^{m} \sum_{all\ y_j such\ that\ g(y_j)=g_i} g_i p(y_j) \\ &= \sum_{j=1}^{n} g(y_j)p(y_j). \end{aligned}$$

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 - μ²).* 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.*
Proof: Consider the function $g(Y) \equiv c$. By Theorem 3.2,

$$E(c) = \sum_{y} cp(y) = c\sum_{y} p(y).$$

But $\sum_y p(y) = 1$ (Theorem 3.1) and, hence, *E(c) = c(1) = 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

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

Proof: By Theorem 3.2,

$$E[cg(Y)] = \sum_y cg(y)p(y) = c\sum_y g(y)p(y) = cE[g(Y)].$$

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* function of *Y*. Then

$$E[g_1(Y) + g_2(Y) + \ldots + g_k(Y)] = E[g_2(Y)] + \ldots + E[g_k(Y)].$$

Proof: We will demonstrate the proof only for the case *k = 2*, but analogous steps will hold for any finite *k.* By Theorem 3.2,

$$E[g_1(Y) + g_2(Y)] = \sum_y [g_1(y) + g_2(y)]p(y)$$

$$= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y)$$

$$= E[g_1(Y)] + E[g_2(Y)].$$

Theorem 3.6

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

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

Proof: $\sigma^2 = E[(Y-\mu)^2 = E(Y^2 - 2\mu Y + \mu^2)$

$= E(Y^2) - E(2\mu Y) + E(\mu^2)$ (by Theorem 3.5).

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

$$\sigma^2 = E(Y^2) - 2\mu E(Y) + \mu^2.$$

But μ = E*(Y)* and, therefore,

$$\sigma^2 = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2.$$

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 *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

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

Theorem 3.7
Let *Y* be a binomial random variable based on *n* trials and success probability *p*. Then

$$\mu = E(Y) = np \quad \text{and} \quad \sigma^2 = V(Y) = npq.$$

Proof: By Definition 3.4 and 3.7,

$$E(Y) = \sum_y yp(y) = \sum_{y=0}^{n} \binom{n}{y} p^y q^{n-y}.$$

Notice that the first term in the sum is 0 and hence that

$$E(Y) = \sum_{y=1}^{n} y \frac{n!}{(n-y)!\,y!} p^y q^{n-y}$$
$$= \sum_{y=1}^{n} \frac{n!}{(n-y)!\,(y-1)!} p^y q^{n-y}.$$

The summand in this expression bear a striking resemblance to binomial probabilities. In fact, ig we factor *np* out of each term in the sum and let *z = y – 1,*

$$E(Y) = np \sum_{y=1}^{n} \frac{(n-1)1}{(n-y)!\,(y-1)!} p^{y-1} q^{n-y}$$
$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)!\,z!} p^z q^{n-1-z}$$
$$= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{n-1-z}.$$

Notice that $p(z) = \binom{n-1}{z} p^z q^{n-1-z}$ is the binomial probability function based on *(n-1)* trials. Thus, $\sum_z p(z) = 1$, and it follows that

$$\mu = E(Y) = np.$$

From Theorem 3.6, we that $\sigma^2 = V(Y) = E(Y^2) - \mu^2$. Thus, μ^2 can be calculated if we find $E(Y^2)$. Finding $E(Y^2)$ directly is difficult because

$$E(Y^2) = \sum_{y=0}^{n} y^2 \binom{n}{y} p^y q^{n-y} = \sum_{y=0}^{n} y^2 \frac{n!}{y!\,(n-y)!} p^y q^{n-y}$$

and the quantity y^2 does not appear as a factor of *y!*. Where do we from here? Notice that

$$E[Y(Y-1)] = E(Y^2 - Y) = E(Y^2) - E(Y)$$

and, therefore,

$$E(Y^2) = E[Y(Y-1)] + E(Y) = E[Y(Y-1] + \mu.$$

In this case,

$$E[Y(Y-1)] = \sum_{y=0}^{n} (y(y-1) \frac{n!}{y!\,(n-y)!} p^y q^{n-y}.$$

The first and second terms of this sum equal zero (when *y = 0* and *y = 1).*

$$E[Y(Y-1)] = \sum_{y=2}^{n} \frac{n!}{(y-2)!\,(n-y)!} p^y q^{n-y}.$$

(Notice the cancellation that led to this result. The anticipation of this cancellation is what actually motivated the consideration of *E[Y(Y-1)].)* Again, the summand in the last expression look very much like binomial probabilities. Factor $n(n-1)p^2$ out of each term in the sum and let $z = y - 2$ to obtain

$$\begin{aligned} E[Y(Y-1)] &= n(n-1)p^2 \sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!\,(n-y)!} p^{y-2} q^{n-y} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{(n-2-z)!} p^z q^{n-2-z} \\ &= np^2 \sum_{z=0}^{n-2} \binom{n-2}{z} p^z q^{n-2-z}. \end{aligned}$$

Again note that $p(z) = \binom{n-2}{z} p^z q^{n-2-z}$ is the binomial probability function based on *(n – 2)* trials. Then $\sum_{z=0}^{n-2} p(z) = 1$ (again using the device illustrated in the derivation of the mean) and

$$E[Y(Y-1)] = n(n-1)p^2.$$

Thus,

$$E(Y^2) = E[Y(Y-1)] + \mu = n(n-1)p^2 + np$$

and

$$\begin{aligned} \sigma^2 &= E(Y^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2 \\ &= np[(n-1)p + 1 - np] = np(1-p) = npq \end{aligned}$$

Definition 3.8

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

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

Theorem 3.8

If *Y* is a random variable with a geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \text{ and } \quad \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Proof:

$$E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p\sum_{y=1}^{\infty} yq^{y-1}$$

This series might seem to be difficult to sum directly. Actually, it can be summed easily if we take into account that, for *y ≥ 1,*

$$\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right) = \sum_{y=1}^{\infty} yq^{y-1}.$$

(The interchanging of derivative and sum here can be justified.) Substituting, we obtain

$$E(Y) = p\sum_{y=1}^{\infty} yq^{y-1} = p\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right).$$

The latter sum is the geometric series, $q + q^2 + q^3 + \ldots,$ which is equal to $q/(1-q)$ (see Appendix A1.11). Therefore,

$$E(Y) = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left[\frac{1}{(1-q)^2}\right] = \frac{p}{p^2} = \frac{1}{p}.$$

To summarize, our approach is to express a series that cannot be summed directly as the derivative of a series for which the sum can be readily obtained. Once we evaluate the more easily handled series, we differentiate to complete the process.