ISE 3293/5013-140

Assignment #2 Guide

Due May 31, 2016

1. [3 marks]

MS 3.36 Page 105-106

Fingerprint expertise. Software engineers are working on developing a fully automated fingerprint identification algorithm. Currently, expert examiners are required to identify the person who left the fingerprint. A study published in *Psychological Science* (August, 2011) tested the accuracy of experts and novices in identifying fingerprints. Participants were presented pairs of fingerprints and asked to judge whether the prints in each pair matched. The pairs were presented under three different conditions: prints from the same individual (*match condition*), non-matching but similar prints (*similar distracter condition*), and nonmatching and very dissimilar prints (*non-similar distracter condition*). The percentages of correct decisions made by the two groups under each of the three conditions are listed in the table.

Condition	Fingerprint experts	Novices
Match	92.12%	74.55%
Similar Distracter	99.32%	44.82%
Non-similar Distracter	100.00%	77.03%

a. Given a pair of matched prints, what is the probability that an expert will fail to identify the match? (The final answer is 0.0788)

b. Given a pair of matched prints, what is the probability that a novice will fail to identify the match? (The final answer is 0.2545)

c. Assume the study included 10 participants, 5 experts and 5 novices. Suppose that a pair of matched prints are presented to a randomly selected study participant and the participant fails to identify the match. Is the participant more likely to be an expert or a novice? (Calculate the probability that the participant is a novice and failed to identify the match and also calculate the probability that the participant is an expert and failed to identify the match. Then compare the probabilities with each other.)

Guide:

Define the following events:

A: {Matched pair is correctly identified}

B: {Similar Distractor pair is correctly identified}

C: {Non-similar Distractor pair is correctly identified}

D: {Participant is an expert} (novices is complement of D or D^c)

According to the table P(A|D)= 0.9212. Find P(B|D), P(C|D).

Also P(A|D^c)= 0.7455. Therefore, you can find P(B|D^c) and P(C|D^c) as well.

We know that P(A^c|B)= 1- P(A|B)

We have the equation for conditional probability:

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

From the above equation we can calculate intersection of A and B:

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

2. [3 marks]

MS 3.52 - Page 111

3.52 Drug testing in athletes. Due to inaccuracies in drug testing procedures (e.g., false positives and false negatives), in the medical field the results of a drug test represent only one factor in a physician's diagnosis. Yet when Olympic athletes are tested for illegal drug use (i.e., doping), the results of a single test are used to ban the athlete from competition. In *Chance* (Spring 2004), University of Texas biostatisticians D. A. Berry and L. Chastain demonstrated **the application of Bayes' Rule** for making inferences about testosterone abuse among Olympic athletes. They used the following example. **In a population of 1,000 athletes**, suppose **100** are illegally using testosterone. Of the users, suppose **50** would test positive for testosterone. Of the nonusers, suppose **9** would test positive.

a. Given the athlete is a user, find the probability that a drug test for testosterone will yield a positive result. This probability represents the *sensitivity* of the drug test.)(The final answer is 0.5)

b. Given the athlete is a nonuser, find the probability that a drug test for testosterone will yield a negative result. (This probability represents the *specificity* of the drug test.) (The final answer is 0.99)

c. If an athlete tests positive for testosterone, use Bayes' Rule to find the probability that the athlete is really doping. (This probability represents the *positive predictive value* of the drug test.) (The final answer is 0.8475)

Guide:

Define the following events to solve the problem:
I: {Athlete is illegally using testosterone}
P: {Drug test yields a positive result}

Read the textbook page 109 and 110 (3.7 Bayes' Rule and the example 3.19)

Bayes' Rule

Given k mutually exclusive and exhaustive states of nature (events), $A_1, A_2, \ldots, A_k$, and an observed event E, then $P(A_i | E)$, for $i = 1, 2, \ldots, k$, is

$$P(A_i | E) = \frac{P(A_i \cap E)}{P(E)} = \frac{P(A_i)P(E | A_i)}{P(A_1)P(E | A_1) + P(A_2)P(E | A_2) + \cdots + P(A_k)P(E | A_k)}$$

3. [1 marks]

MS Theorem 3.1 - Page 113 Prove the theorem in your own words.

THEOREM 3.1

The Multiplicative Rule You have k sets of elements—n_1 in the first set, n_2 in the second set, . . . , and n_k in the kth set. Suppose you want to form a sample of k elements by taking one element from each of the k sets. The number of different samples that can be formed is the product

$$n_1 n_2 n_3 \ldots\ldots n_k$$

Outline of Proof of Theorem 3.1 The proof of Theorem 3.1 can be obtained most easily by examining Table 3.4. Each of the pairs that can be formed from two sets of elements—$a_1, a_2, \ldots, a_{n1}$ and $b_1, b_2, \ldots, b_{n2}$—corresponds to a cell of Table 3.4.

Since the table contains n_1 rows and n_2 columns, there will be $n_1 n_2$ pairs corresponding to each of the $n_1 n_2$ cells of the table. To extend the proof to the case in which $k = 3$, note that the number of triplets that can be formed from three sets of elements—$a_1, a_2, \ldots, a_{n1}$; $b_1, b_2, \ldots, b_{n2}$; and $c_1, c_2, \ldots, c_{n3}$—is equal to the number of pairs that can be formed by associating one of the $a_i b_j$ pairs with one of the c elements. Since there are (n_1, n_2) of the $a_i b_j$ pairs and n_3 of the c elements, we can form $(n_1 n_2)n_3 = n_1 n_2 n_3$ triplets consisting of one a element, one b element, and one c element. The proof of the multiplicative rule for any number, say, k, of sets is obtained by mathematical induction. We leave this proof as an exercise for the student.

TABLE 3.4 **Pairings of $a_1, a_2, \ldots, a_{n1}$, and $b_1, b_2, \ldots, b_{n2}$**

	b_1	b_2	b_3	. . .	b_{n2}
a_1	a_1b_1	a_1b_2	a_1b_3	. . .	a_1b_{n2}
a_2	a_2b_1	. . .	. . .	. . .	. . .
a_3	a_3b_1	. . .	. . .	. . .	. . .
⋮	⋮	⋮	⋮	⋮	⋮
a_{n1}	$a_{n1}b_1$	. . .	. . .	. . .	$a_{n1}b_{n2}$

4. [1 marks]

MS Theorem 3.2 - Page 114 Prove the theorem in your own words.

THEOREM 3.2

Permutations Rule Given a single set of N distinctly different elements, you wish to select n elements from the N and arrange them within n positions in a distinct order. The number of different permutations of the N elements taken n at a time is denoted by P_n^N and is equal to

$$P_n^N = N(N-1)(N-2)\cdot\cdots\cdot(N-n+1) = \frac{N!}{(N-n)!}$$

where $n! = n(n-1)(n-2)\cdot\cdots\cdot(3)(2)(1)$ and is called ***n* factorial**. (Thus, for example, $5! = 5\cdot4\cdot3\cdot2\cdot1 = 120$.) The quantity $0!$ is defined to be equal to 1.

Proof of Theorem 3.2 The proof of Theorem 3.2 is a generalization of the solution to Example 3.23. There are N ways of filling the first position. After it is filled, there are $N-1$ ways of filling the second, $N-2$ ways of filling the third, . . . , and $(N-n+1)$ ways of filling the nth position. We apply the multiplicative rule to obtain

$$P_n^N = (N)(N-1)(N-2)\cdot\cdots\cdot(N-n+1) = \frac{N!}{(N-n)!}$$

5. [1 marks]

MS Theorem 3.3 – Page 116 Prove the theorem in your own words.

THEOREM 3.3

Partitions Rule There exists a single set of N distinctly different elements and you want to partition them into k sets, the first set containing n_1 elements, the second containing n_2 elements, . . . , and the kth set containing n_k elements. The number of different partitions is

$$\frac{N!}{n_1!n_2!\cdot\cdots\cdot n_k!} \quad \text{where } n_1+n_2+n_3+\cdots+n_k = N$$

Proof of Theorem 3.3 Let A equal the number of ways that you can partition N distinctly different elements into k sets. We want to show that

$$A = \frac{N!}{n_1!n_2!\cdot\cdots\cdot n_k!}$$

We will find A by writing an expression for arranging N distinctly different elements in N positions. By Theorem 3.2, the number of ways this can be done is

$$P_N^N = \frac{N!}{(N-N)!} = \frac{N!}{0!} = N!$$

But, by Theorem 3.1, P_N^N is also equal to the product

$$P_N^N = N! = (A)(n_1!)(n_2!)\cdots(n_k!)$$

where A is the number of ways of partitioning N elements into k groups of $n_1, n_2, \ldots, n_k$ elements, respectively; $n_1!$ is the number of ways of arranging the n_1 elements in group 1; $n_2!$ is the number of ways of arranging the n_2 elements in group 2; . . . ; and $n_k!$ is the number of ways of arranging the n_k elements in group k. We obtain the desired result by solving for A:

$$A = \frac{N!}{n_1!n_2!\cdot\cdots\cdot n_k!}$$

6. [1 marks]

MS Theorem 3.4 - Page 117 Prove the theorem in your own words.

THEOREM 3.4

The Combinations Rule A sample of n elements is to be chosen from a set of N elements. Then the number of different samples of n elements that can be selected from N is denoted by $\binom{N}{n}$ and is equal to

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Note that the order in which the n elements are drawn is not important.

Proof of Theorem 3.4 The proof of Theorem 3.4 follows directly from Theorem 3.3. Selecting a sample of n elements from a set of N elements is equivalent to partitioning the N elements into $k = 2$ groups—the n that are selected for the sample and the remaining $(N - n)$ that are not selected. Therefore, by applying Theorem 3.3 we obtain

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

7. [3 marks]

MS 4.2 - Page 138

4.2 Dust mite allergies. A dust mite allergen level that exceeds 2 micrograms per gram (mg/g) of dust has been associated with the development of allergies. Consider a random sample of four homes and let Y be the number of homes with a dust mite level that exceeds 2 mg/g. The probability distribution for Y=y, based on a study by the National Institute of Environmental Health Sciences, is shown in the following table.

y	0	1	2	3	4
$p(y)$	.09	.30	.37	.20	.04

a. Verify that the probabilities for Y in the table sum to 1.
b. Find the probability that three or four of the homes in the sample have a dust mite level that exceeds 2 mg/g. (The final answer is 0.24)
c. Find the probability that fewer than two homes in the sample have a dust mite level that exceeds 2 mg/g. (The final answer is 0.39)

Guide:

Requirements for a Discrete Probability Distribution

1. $0 \leq p(y) \leq 1$
2. $\sum_{\text{all } y} p(y) = 1$

Example 4.2

Probability Distribution for Driver-Side Crash Ratings

CRASH

The National Highway Traffic Safety Administration (NHTSA) has developed a driver-side "star" scoring system for crash-testing new cars. Each crash-tested car is given a rating ranging from one star (*) to five stars (*****); the more stars in the rating, the better the level of crash protection in a head-on collision. Recent data for 98 new cars are saved in the **CRASH** file. A summary of the driver-side star ratings for these cars is reproduced in the MINITAB printout, Figure 4.3. Assume that one of the 98 cars is selected at random and let Y equal the number of stars in the car's driver-side star rating. Use the information in the printout to find the probability distribution for Y. Then find $P(Y \leq 3)$.

FIGURE 4.3
MINITAB summary of driver-side star ratings

Tally for Discrete Variables: DRIVSTAR

DRIVSTAR	Count	Percent
2	4	4.08
3	17	17.35
4	59	60.20
5	18	18.37
N=	98	

Solution

Since driver-side star ratings range from 1 to 5, the discrete random variable Y can take on the values 1, 2, 3, 4, or 5. The MINITAB printout gives the percentage of the 98 cars in the **CRASH** file that fall into each star category. These percentages represent the probabilities of a randomly selected car having one of the star ratings. Since none of the 98 cars has a star rating of 1, $P(Y = 1) = p(1) = 0$. The remaining probabilities for Y are as follows: $p(2) = .0408$, $p(3) = .1735$, $p(4) = .6020$, and $p(5) = .1837$. Note that these probabilities sum to 1.

To find $P(Y \leq 3)$, we sum the values $p(1)$, $p(2)$, and $p(3)$.

$$P(Y \leq 3) = p(1) + p(2) + p(3) = 0 + .0408 + .1735 = .2143$$

Thus, about 21% of the driver-side star ratings have three or fewer stars.

8. [4 marks]

MS 4.12 - Page 143

4.12 Downloading "apps" to your cell phone. According to an August, 2011 survey by the Pew Internet & American Life Project, nearly 40% of adult cell phone owners have downloaded an application ("app") to their cell phone. The accompanying table gives the probability distribution for Y, the number of "apps" used at least once a week by cell phone owners who have downloaded an "app" to their phone. (The probabilities in the table are based on information from the Pew Internet & American Life Project survey.)

a. Show that the properties of a probability distribution for a discrete random variable are satisfied.

Requirements for a Discrete Probability Distribution

1. $0 \leq p(y) \leq 1$
2. $\sum_{\text{all } y} p(y) = 1$

b. Find $P(Y \geq 10)$ (The final answer is 0.14)
c. Find the mean and variance of Y. (Mean= 4.655 and variance= 19.856)

Definition 4.4

Let Y be a discrete random variable with probability distribution $p(y)$. Then the **mean** or **expected value** of Y is

$$\mu = E(Y) = \sum_{\text{all } y} yp(y)$$

Definition 4.6

Let Y be a discrete random variable with probability distribution $p(y)$. Then the **variance of Y** is

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

The **standard deviation of Y** is the positive square root of the variance of Y:

$$\sigma = \sqrt{\sigma^2}$$

Read the textbook pages 141-143 (Examples 4.3, 4.4, 4.5 and 4.6)

d. Give an interval that will contain the value of Y with a probability of at least 0.75. (The final answer is 0, 13.56)

Number of "apps" used, y	$p(y)$
0	.17
1	.10
2	.11
3	.11
4	.10
5	.10
6	.07
7	.05
8	.03
9	.02
10	.02
11	.02
12	.02
13	.02
14	.01
15	.01
16	.01
17	.01
18	.01
19	.005
20	.005

9. [4 marks]

MS 4.34 - Page 154

4.34 PhD's in engineering. The National Science Foundation reports that 70% of the U.S. graduate students who earn PhD degrees in engineering are foreign nationals. Consider the number Y of foreign students in a random sample of 25 engineering students who recently earned their PhD.

a. Find P(Y=10). (The final answer is 0.0013)

Guide:

The experiment consists of $n = 25$trials.

Each trial results in an S (graduate student earning PhD degree in engineering is a foreign national) or an F (graduate student earning PhD degree in engineering is not a foreign national).

The probability of success, ***p***, is **0.70** and ***q*** *=1- p= 1- 0.7=* ***0.3***

We assume the trials are independent. Therefore,

Y = the number of graduate students earning a PhD in engineering who are foreign nationals in 25 trials and Y has a **binomial** distribution with $n = 25$and $p = 0.70$.

The Binomial Probability Distribution

The probability distribution for a binomial random variable Y is given by

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

where

p = Probability of a success on a single trial

$q = 1 - p$

n = Number of trials

y = Number of successes in n trials

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

The mean and variance of the binomial random variable are, respectively,

$$\mu = np \quad \text{and} \quad \sigma^2 = npq$$

Read the textbook pages 150-152 (Examples 4.9, 4.10 and 4.11)

b. Find $P(Y \leq 5)$. (The final answer is 0.)
c. Find the mean μ and standard deviation s for Y. (The mean is 17.5 and the standard deviation is 2.2913)
d. Interpret the results, part **c**.

10. [2 marks]

MS 4.46 - Page 158

4.46 Railway track allocation. Refer to the *Journal of Transportation Engineering* (May, 2013) investigation of the assignment of tracks to trains at a busy railroad station, Exercise 2.8 (p. 28). Ideally, engineers will assign trains to tracks in order to minimize waiting time and bottlenecks. Assume there are 10 tracks at the railroad station and the trains will be randomly assigned to a track. Suppose that in a single day there are 50 trains that require track assignment.

a. What is the probability that exactly 5 trains are assigned to each of the 10 tracks? (The final answer is 0.000000491 which is almost 0.)
b. A track is considered underutilized if fewer than 2 trains are assigned to the track during the day. Find the probability that Track #1 is underutilized. (The final answer is 0.0338)

Guide:

This experiment consists of 50 identical trials. There are ten possible outcomes on each trial and each outcome has an equal chance of being selected. Assuming the trials are independent, this is a **multinomial experiment** with *n=50, k=10, p=0.1, i* = 1,2,...,10.

Read the textbook pages 153-15 (4.7 The Multinomial Probability Distribution example 4.12)

Properties of the Multinomial Experiment

1. The experiment consists of n identical trials.
2. There are k possible outcomes to each trial.
3. The probabilities of the k outcomes, denoted by $p_1, p_2, \ldots, p_k$, remain the same from trial to trial, where $p_1 + p_2 + \cdots + p_k = 1$.
4. The trials are independent.
5. The random variables of interest are the counts $Y_1, Y_2, \ldots, Y_k$ in each of the k classification categories.

The Multinomial Probability Distribution

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \cdots y_k!}(p_1)^{y_1}(p_2)^{y_2} \cdots (p_k)^{y_k}$$

where

p_i = Probability of outcome i on a single trial
$p_1 + p_2 + \cdots + p_k = 1$
$n = y_1 + y_2 + \cdots + y_k$ = Number of trials

y_i = Number of occurrences of outcome i in n trials

The mean and variance of the multinomial random variable y_i are, respectively,

$$\mu_i = np_i \quad \text{and} \quad \sigma_i^2 = np_i(1 - p_i)$$

11. [2 marks]

MS 4.54 - Page 162

4.54 Is a product "green"? Refer to the *ImagePower Green Brands Survey* of international consumers, Exercise 3.4 (p. 84). Recall that a "green" product is one built from recycled materials that has minimal impact on the environment. The reasons why a consumer identifies a product as green are summarized in

the next table. Consider interviewing consumers, at random. Let *Y* represent the number of consumers who must be interviewed until one indicates something other than information given directly on the product's label or packaging as the reason a product is green.
a. Give a formula for the probability distribution of *Y*.
b. What is E(*Y*)? Interpret the result. (The final answer is 2.5.)
c. Find P(Y=1) (The final answer is 0.4)
d. Find P(Y>2) (The final answer is 0.36)

Reason for saying a product is green	Percentage of consumers
Certification mark on label	45
Packaging	15
Reading information about the product	12
Advertisement	6
Brand website	4
Other	18
TOTAL	100

Guide:

Let S=Reason given is something other than information given on label or package. (Then calculate p)

F= Reason given is found on label and packaging. (Then calculate q)

Y= The number of consumers until one indicates something other than information given on label or packaging, then the probability distribution of Y is a **geometric**.

The Geometric Probability Distribution

$$p(y) = pq^{y-1} \quad (y = 1, 2, \dots)$$

where

Y = Number of trials until the first success is observed

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{q}{p^2}$$

12. [2 marks]

MS 4.66 - Page 168

4.66 On-site disposal of hazardous waste. The Resource Conservation and Recovery Act mandates the tracking and disposal of hazardous waste produced at U.S. facilities. *Professional Geographer* (Feb. 2000) reported the hazardous waste generation and disposal characteristics of 209 facilities. Only eight of these facilities treated hazardous waste on-site.

a. In a random sample of 10 of the 209 facilities, what is the expected number in the sample that treats hazardous waste on-site? Interpret this result. (The final answer is 0.3828)
b. Find the probability that 4 of the 10 selected facilities treat hazardous waste on-site. (The final answer is 0.000169)

Guide:

Let Y = the number facilities that treat hazardous waste on-site. Then Y has a **hypergeometric** distribution with $N = 209$, $r = 8$, and $n = 10$.

Characteristics That Define a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, r of which are S's (for Success) and $(N - r)$ of which are F's (for Failure).
2. The sample size n is large relative to the number N of elements in the population, i.e., $n/N > .05$.
3. The hypergeometric random variable Y is the number of S's in the draw of n elements.

The Hypergeometric Probability Distribution

The hypergeometric probability distribution is given by

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, \quad y = \text{Maximum } [0, n - (N - r)], \ldots, \text{Minimum } (r, n)$$

where

N = Total number of elements
r = Number of S's in the N elements
n = Number of elements drawn
y = Number of S's drawn in the n elements

The mean and variance of a hypergeometric random variable are, respectively,

$$\mu = \frac{nr}{N} \qquad \sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)}$$

Read the textbook pages 164-167 examples 4.15, 4.16 and 4.17.

13. [3 marks]

MS 4.78 - Page 173
4.78 Deep-draft vessel casualties. Engineers at the University of New Mexico modeled the number of casualties (deaths or missing persons) experienced by a deep-draft U.S. flag vessel over a 3-year period as a Poisson random variable, Y. The researchers estimated $E(Y)$ to be .03. (*Management Science*, Jan. 1999.)
a. Find the variance of Y. (The final answer is 0.03.)

b. Discuss the conditions that would make the researchers' Poisson assumption plausible.
c. What is the probability that a deep-draft U.S. flag vessel will have no casualties in a 3-year time period? (The final answer is 0.9704)

Guide:

Read the textbook pages 168- 172 (4.10 The Poisson Probability Distribution and example 4.18)

Characteristics of a Poisson Random Variable

1. The experiment consists of counting the number of times Y a particular (rare) event occurs during a given unit of time or in a given area or volume (or weight, distance, or any other unit of measurement).
2. The probability that an event occurs in a given unit of time, area, or volume is the same for all the units. Also, units are mutually exclusive.
3. The number of events that occur in one unit of time, area, or volume is independent of the number that occur in other units.

The Poisson Probability Distribution

The probability distribution* for a Poisson random variable Y is given by

$$p(y) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad (y = 0, 1, 2, \ldots)$$

where

λ = Mean number of events during a given unit of time, area, or volume
$e = 2.71828\ldots$

The mean and variance of a Poisson random variable are, respectively,

$\mu = \lambda$ and $\sigma^2 = \lambda$

14. [4 marks]

MS 5.2 - Page 191

Let c be a constant and consider the density function for the random variable Y:

$$f(y) = \begin{cases} c(2 - y) & \text{if } 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

a. Find the value of c. (The final answer is 2/3.)
b. Find the cumulative distribution function $F(y)$.
c. Compute $F(0.4)$. (The final answer is 0.48.)
d. Compute $P(0.1 \leq Y \leq 0.6)$ (The final answer is 0.55.)

Guide:

Properties of a Density Function for a Continuous Random Variable Y

1. $f(y) \geq 0$
2. $\int_{-\infty}^{\infty} f(y)\, dy = F(\infty) = 1$
3. $P(a < Y < b) = \int_a^b f(y)\, dy = F(b) - F(a)$, where a and b are constants

Read example 5.1 and 5.2 in the textbook.

15. [3 marks]

MS 5.10 - Page 196

5.10 Time a train is late. Refer to Exercise 5.5 (p. 191) the amount of time Y (in minutes) that a commuter train is late is a continuous random variable with probability density

$$f(y) = \begin{cases} \frac{3}{500}(25 - y^2) & \text{if } -5 < y < 5 \\ 0 & \text{elsewhere} \end{cases}$$

a. Find the mean and variance of the amount of time in minutes the train is late.
(Mean = 0 and variance = 5.)
b. Find the mean and variance of the amount of time in hours the train is late. (Mean=0 and variance = 0.0014.)
c. Find the mean and variance of the amount of time in seconds the train is late (Mean = 0 and variance = 18000.)

Guide:

Definition 5.4

Let Y be a continuous random variable with density function $f(y)$, and let $g(Y)$ be any function of Y. Then the **expected values** of Y and $g(Y)$ are

$$E(Y) = \int_{-\infty}^{\infty} yf(y)\, dy$$

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

THEOREM 5.2

Let Y be a continuous random variable with $E(Y) = \mu$. Then

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

Read the textbook pages 192-196 (Examples 5.3, 5.4, 5.5 and 5.6)

16. [3 marks]

MS 5.36 - Page 205

5.36 Alkalinity of river water. The alkalinity level of water specimens collected from the Han River in Seoul, Korea, has a mean of 50 milligrams per liter and a standard deviation of 3.2 milligrams per liter. (*Environmental Science & Engineering*, Sept. 1, 2000.) Assume the distribution of alkalinity levels is approximately **normal** and find the probability that a water specimen collected from the river has an alkalinity level

a. exceeding 45 milligrams per liter. (The final answer is 0.9406.)
b. below 55 milligrams per liter.(The final answer is 0.9406.)
c. between 51 and 52 milligrams per liter. (The final answer is 0.1140.)

Guide:

THEOREM 5.4

If Y is a normal random variable with mean μ and variance σ^2, then $Z = (Y - \mu)/\sigma$ is a normal random variable with mean 0 and variance 1.* The random variable Z is called a **standard normal variable**.

The areas for the standard normal variable,

$$Z = \frac{Y - \mu}{\sigma}$$

are given in Table 5 of Appendix B. Recall from Section 2.6 that Z is the distance between the value of the normal random variable Y and its mean μ, measured in units of its standard deviation σ.

Let Y = alkalinity level of water specimens collected from the Han River. Then Y is normally distributed with $\mu = 50$ and $\sigma = 3.2$.

(Read the textbook pages 201-204 and examples 5.8-5.11)

Use Table 5 Appendix B.

17. [5 marks]

MS 5.38- Page 205

5.38 NHTSA crash safety tests. Refer to the National Highway Traffic Safety Administration (NHTSA) crash test data for new cars, introduced in Exercise 2.74 (p. 70) and saved in the **CRASH** file. One of the variables measured is the severity of a driver's head injury when the car is in a headon collision with a fixed barrier while traveling at 35 miles per hour. The more points assigned to the head injury rating, the more severe the injury. The head injury ratings can be shown to be approximately normally distributed with a mean of 605 points and a standard deviation of 185 points. One of the crash-tested cars is randomly selected from the data and the driver's head injury rating is observed.

a. Find the probability that the rating will fall between 500 and 700 points (The final answer is 0.4107.)
b. Find the probability that the rating will fall between 400 and 500 points (The final answer is 0.1508.)
c. Find the probability that the rating will be less than 850 points (The final answer is 0.9066.)
d. Find the probability that the rating will exceed 1,000 points (The final answer is 0.0162.)
e. What rating will only 10% of the crash-tested cars exceed? (The final answer is 841.8)

Guide:

Let Y = driver's head injury rating. Then Y is normally distributed with $\mu = 605$ and $\sigma = 185$.