華南理工大學
South China University of Technology

040100025

Probability and Statistics

Prob. & Stat. course group

Date Last Edited: March 12, 2022

Contents

1 Chapter 2

2.1

1. Four universities-1, 2, 3, and 4-are participating in a holiday basketball tournament. In the first round, 1 will play 2 and 3 will play 4 . Then the two winners will play for the championship, and the two losers will also play. One possible outcome can be denoted by 1324 (1 beats 2 and 3 beats 4 in first-round games, and then 1 beats 3 and 2 beats 4).

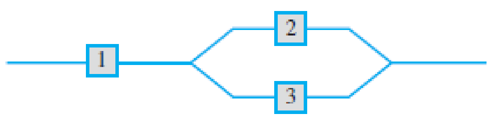
   a. List all outcomes in $\mathcal{S}$.

   b. Let A denote the event that 1 wins the tournament. List outcomes in A.

   c. Let B denote the event that 2 gets into the championship game. List outcomes in B.

   d. What are the outcomes in $A \cup B$ and in $A \cap B$? What are the outcomes in A'?

2. Three components are connected to form a system as shown in the accompanying diagram. Because the components in the 2-3 subsystem are connected in parallel, that subsystem will function if at least one of the two individual components functions. For the entire system to function, component 1 must function and so must the $2-3$ subsystem. The experiment consists of determining the condition of each component [S (success) for a functioning component and F (failure) for a nonfunctioning component].



   a. Which outcomes are contained in the event A that exactly two out of the three components function?

   b. Which outcomes are contained in the event B that at least two of the components function?

   c. Which outcomes are contained in the event C that the system functions?

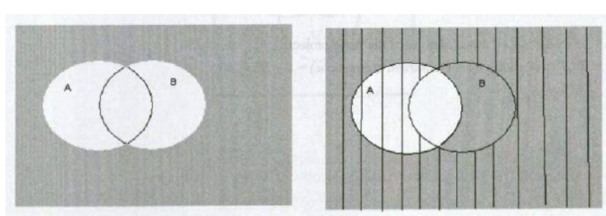
   d. List outcomes in C', $A \cup C$, $A \cap C$, $B \cup C$, and $B \cap C$.

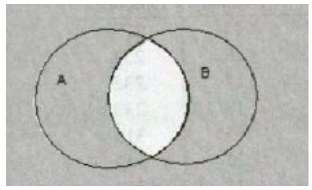
3. Use Venn diagrams to verify the following two relationships for any events A and B (these are called De Morgan's laws):

   a. $(A \cup B)' = A' \cap B'$

   b. $(A \cap B)' = A' \cup B'$

   [Hint: In each part, draw a diagram corresponding to the left side and another corresponding to the right side.]





2.2

4. Let A denote the event that the next request for assistance from a statistical software consultant relates to the SPSS package, and let B be the event that the next request is for help with SAS. Suppose that $P(A)=.30$ and $P(B)=.50$.

   a. Why is it not the case that $P(A)+P(B)=1$?

   b. Calculate $P\left(A^{\prime}\right)$.

   c. Calculate $P(A \cup B)$.

   d. Calculate $P\left(A^{\prime} \cap B^{\prime}\right)$.

5. Human visual inspection of solder joints on printed circuit boards can be very subjective. Part of the problem stems from the numerous types of solder defects (e.g., pad nonwetting, knee visibility, voids) and even the degree to which a joint possesses one or more of these defects. Consequently, even highly trained inspectors can disagree on the disposition of a particular joint. In one batch of 10,000 joints, inspector A found 724 that were judged defective, inspector B found 751 such joints, and 1159 of the joints were judged defective by at least one of the inspectors. Suppose that one of the 10,000 joints is randomly selected.

   a. What is the probability that the selected joint was judged to be defective by neither of the two inspectors?

   b. What is the probability that the selected joint was judged to be defective by inspector B but not by inspector A?

6. The three most popular options on a certain type of new car are a built-in GPS (A), a sunroof (B), and an automatic transmission (C). If 40% of all purchasers request $A, 55\%$ request $B, 70\%$ request $C, 63\%$ request A or B, 77% request A or $C, 80\%$ request B or C, and 85% request A or B or C, determine the probabilities of the following events.

   [Hint: "A or B" is the event that at least one of the two options is requested; try drawing a Venn diagram and labeling all regions.]

   a. The next purchaser will request at least one of the three options.

   b. The next purchaser will select none of the three options.

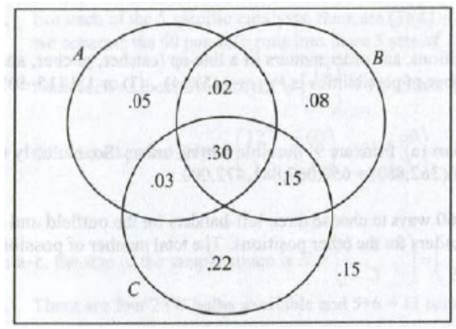
   c. The next purchaser will request only an automatic transmission and not either of the other two options.

   d. The next purchaser will select exactly one of these three options.

2.3

7. The composer Beethoven wrote 9 symphonies, 5 piano concertos (music for piano and orchestra), and 32 piano sonatas (music for solo piano).

   a. How many ways are there to play first a Beethoven symphony and then a Beethoven piano concerto?



b. The manager of a radio station decides that on each successive evening (7 days per week), a Beethoven symphony will be played followed by a Beethoven piano concerto followed by a Beethoven piano sonata. For how many years could this policy be continued before exactly the same program would have to be repeated?

8. A production facility employs 10 workers on the day shift, 8 workers on the swing shift, and 6 workers on the graveyard shift. A quality control consultant is to select 5 of these workers for in-depth interviews. Suppose the selection is made in such a way that any particular group of 5 workers has the same chance of being selected as does any other group (drawing 5 slips without replacement from among 24).

   a. How many selections result in all 5 workers coming from the day shift? What is the probability that all 5 selected workers will be from the day shift?
   b. What is the probability that all 5 selected workers will be from the same shift?
   c. What is the probability that at least two different shifts will be represented among the selected workers?
   d. What is the probability that at least one of the shifts will be unrepresented in the sample of workers?

9. A box in a supply room contains 15 compact fluorescent lightbulbs, of which 5 are rated 13-watt, 6 are rated 18-watt, and 4 are rated 23-watt. Suppose that three of these bulbs are randomly selected.

   a. What is the probability that exactly two of the selected bulbs are rated 23-watt?
   b. What is the probability that all three of the bulbs have the same rating?
   c. What is the probability that one bulb of each type is selected?
   d. If bulbs are selected one by one until a 23-watt bulb is obtained, what is the probability that it is necessary to examine at least 6 bulbs?

10. In five-card poker, a straight consists of five cards with adjacent denominations (e.g., 9 of clubs, 10 of hearts, jack of hearts, queen of spades, and king of clubs). Assuming that aces can be high or low, if you are dealt a five-card hand, what is the probability that it will be a straight with high card 10? What is the probability that it will be a straight? What is the probability that it will be a straight flush (all cards in the same suit)?

2.4

11. The population of a particular country consists of three ethnic groups. Each individual belongs to one of the four major blood groups. The accompanying *joint probability table* gives the proportions of individuals in the various ethnic group-blood group combinations.

		Blood Group			
		O	**A**	**B**	**AB**
Ethnic Group	**1**	.082	.106	.008	.004
	2	.135	.141	.018	.006
	3	.215	.200	.065	.020

Suppose that an individual is randomly selected from the population, and define events by $A = \{\text{type A selected }\}, B = \{\text{type B selected}\}$, and $C = \{\text{ethnic group 3 selected}\}$.

a. Calculate $P(A), P(C)$, and $P(A \cap C)$.

b. Calculate both $P(A|C)$ and $P(C|A)$, and explain in context what each of these probabilities represents.

c. If the selected individual does not have type B blood, what is the probability that he or she is from ethnic group 1?

12. According to a July 31, 2013, posting on cnn.com subsequent to the death of a child who bit into a peanut, a 2010 study in the journal *Pediatrics* found that 8% of children younger than 18 in the United States have at least one food allergy. Among those with food allergies, about 39% had a history of severe reaction.

    a. If a child younger than 18 is randomly selected, what is the probability that he or she has at least one food allergy and a history of severe reaction?

    b. It was also reported that 30% of those with an allergy in fact are allergic to multiple foods. If a child younger than 18 is randomly selected, what is the probability that he or she is allergic to multiple foods?

13. Deer ticks can be carriers of either Lyme disease or human granulocytic ehrlichiosis (HGE). Based on a recent study, suppose that 16% of all ticks in a certain location carry Lyme disease, 10% carry HGE, and 10% of the ticks that carry at least one of these diseases in fact carry both of them. If a randomly selected tick is found to have carried HGE, what is the probability that the selected tick is also a carrier of Lyme disease?

14. If $P(B|A) > P(B)$, show that $P(B'|A) < P(B')$.

15. Components of a certain type are shipped to a supplier in batches of ten. Suppose that 50% of all such batches contain no defective components, 30% contain one defective component, and 20% contain two defective components. Two components from a batch are randomly selected and tested. What are the probabilities associated with 0, 1, and 2 defective components being in the batch under each of the following conditions?

    a. Neither tested component is defective.

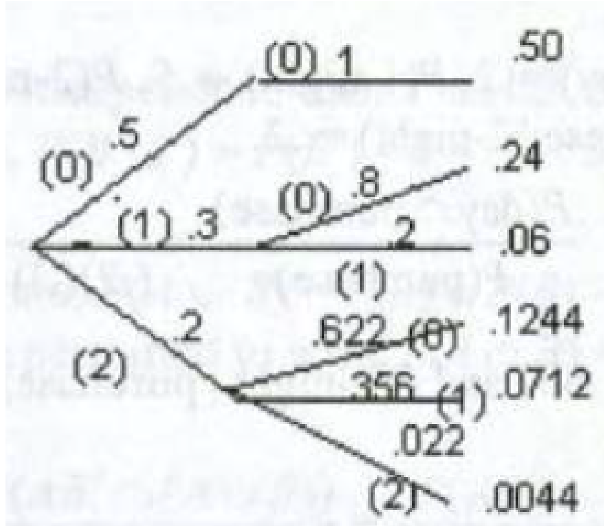
    b. One of the two tested components is defective. [Hint: Draw a tree diagram with three first-generation branches for the three different types of batches.]

2.5

16. An aircraft seam requires 25 rivets. The seam will have to be reworked if any of these rivets is defective. Suppose rivets are defective independently of one another, each with the same probability.

    a. If 15% of all seams need reworking, what is the probability that a rivet is defective?

    b. How small should the probability of a defective rivet be to ensure that only 10% of all seams need reworking?



17. Components arriving at a distributor are checked for defects by two different inspectors (each component is checked by both inspectors). The first inspector detects 90% of all defectives that are present, and the second inspector does likewise. At least one inspector does not detect a defect on 20% of all defective components. What is the probability that the following occur?

    a. A defective component will be detected only by the first inspector? By exactly one of the two inspectors?

    b. All three defective components in a batch escape detection by both inspectors (assuming inspections of different components are independent of one another)?

18. Suppose identical tags are placed on both the left ear and the right ear of a fox. The fox is then let loose for a period of time. Consider the two events $C_1 = \{$ left ear tag is lost $\}$ and $C_2 = \{$right ear tag is lost$\}$. Let $\theta = P(C_1) = P(C_2)$, and assume C_1 and C_2 are independent events. Derive an expression (involving θ) for the probability that exactly one tag is lost, given that at most one is lost ("Ear Tag Loss in Red Foxes", *J. Wildlife Mgmt.*, 1976: 164-167).

Supplementary Exercises

19. One satellite is scheduled to be launched from Cape Canaveral in Florida, and another launching is scheduled for Vandenberg Air Force Base in Califomia. Let A denote the event that the Vandenberg launch goes off on schedule, and let B represent the event that the Cape Canaveral launch goes off on schedule. If A and B are independent events with $P(A) > P(B), P(A \cup B) = .626$, and $P(A \cap B) = .144$, determine the values of $P(A)$ and $P(B)$.

20. Individual A has a circle of five close friends (B, C, D, E, and F). A has heard a certain rumor from outside the circle and has invited the five friends to a party to circulate the rumor. To begin, A selects one of the five at random and tells the rumor to the chosen individual. That individual then selects at random one of the four remaining individuals and repeats the rumor. Continuing, a new individual is selected from those not already having heard the rumor by the individual who has just heard it, until everyone has been told.

    a. What is the probability that the rumor is repeated in the order B, C, D, E, and F?

    b. What is the probability that F is the third person at the party to be told the rumor?

    c. What is the probability that F is the last person to hear the rumor?

d. If at each stage the person who currently ¡°has¡± the rumor does not know who has already heard it and selects the next recipient at random from all five possible individuals, what is the probability that F has still not heard the rumor after it has been told ten times at the party?

21. Disregarding the possibility of a February 29 birthday, suppose a randomly selected individual is equally likely to have been born on any one of the other 365 days.

    a. If ten people are randomly selected, what is the probability that all have different birthdays? That at least two have the same birthday?

    b. With k replacing ten in part (a), what is the smallest k for which there is at least a 50-50 chance that two or more people will have the same birthday?

    c. If ten people are randomly selected, what is the probability that either at least two have the same birthday or at least two have the same last three digits of their Social Security numbers? [Note: The article "Methods for Studying Coincidences" (F. Mosteller and P. Diaconis, *J. Amer. Stat. Assoc.*, 1989: 853-861) discusses problems of this type.]

22. A box contains the following four slips of paper, each having exactly the same dimensions: (1) win prize 1; (2) win prize 2; (3) win prize 3; (4) win prizes 1, 2, and 3. One slip will be randomly selected. Let $A_1 = \{\text{win prize 1}\}$, $A_2 = \{\text{win prize 2}\}$, and $A_3 = \{\text{win prize 3}\}$. Show that A_1 and A_2 are independent, that A_1 and A_3 are independent, and that A_2 and A_3 are also independent (this is pairwise independence). However, show that $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$, so the three events are not mutually independent. [Note: A_1=draw slip 1 or 4.]

2 Chapter 3

3.1

1. **7.** For each random variable defined here, describe the set of possible values for the variable, and state whether the variable is discrete.

   a. X = the number of unbroken eggs in a randomly chosen standard egg carton

   b. Y = the number of students on a class list for a particular course who are absent on the first day of classes

   c. U = the number of times a duffer has to swing at a golf ball before hitting it

   d. X = the length of a randomly selected rattlesnake

   e. Z = the sales tax percentage for a randomly selected amazon.com purchase

   f. Y = the pH of a randomly chosen soil sample

   g. X = the tension (psi) at which a randomly selected tennis racket has been strung

   h. X = the total number of times three tennis players must spin their rackets to obtain something other than UUU or DDD (to determine which two play next)

2. **9.** An individual named Claudius is located at the point 0 in the accompanying diagram (Fig. 1). Using an appropriate randomization device (such as a tetrahedral die, one having four sides), Claudius first moves to one of the four locations B_1, B_2, B_3, B_4. Once at one of these locations, another randomization device is used to decide whether Claudius next returns to 0 or next visits one of the other two adjacent points. This process then continues; after each move, another move to one of the (new) adjacent points is determined by tossing an appropriate die or coin.

   a. Let X = the number of moves that Claudius makes before first returning to 0 . What are possible values of X? Is X discrete or continuous?

   b. If moves are allowed also along the diagonal paths connecting 0 to A_1, A_2, A_3, and A_4, respectively, answer the questions in part (a).

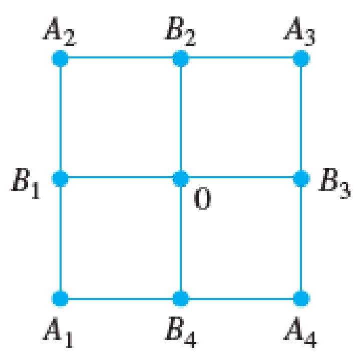


Figure 1: Question 9

3.2

3. **17.** A new battery's voltage may be acceptable (A) or unacceptable (U). A certain flashlight requires two batteries, so batteries will be independently selected and tested until two acceptable ones have been

found. Suppose that 90% of all batteries have acceptable voltages. Let Y denote the number of batteries that must be tested.

a. What is $p(2)$, that is, $P(Y=2)$?

b. What is $p(3)$? [Hint: There are two different outcomes that result in $Y=3$.]

c. To have $Y=5$, what must be true of the fifth battery selected? List the four outcomes for which $Y=5$ and then determine $p(5)$.

d. Use the pattern in your answers for parts (a)-(c) to obtain a general formula for $p(y)$.

4. **23(22).** A branch of a certain bank in New York City has six ATMs. Let X represent the number of machines in use at a particular time of day. The cdf of X is as follows:

$$F(x)=\begin{cases}0 & x<0 \\ .06 & 0 \leq x<1 \\ .19 & 1 \leq x<2 \\ .39 & 2 \leq x<3 \\ .67 & 3 \leq x<4 \\ .92 & 4 \leq x<5 \\ .97 & 5 \leq x<6 \\ 1 & 6 \leq x\end{cases}$$

Calculate the following probabilities directly from the cdf:

a. $p(2)$, that is, $P(X=2)$

b. $P(X>3)$

c. $P(2 \leq X \leq 5)$

d. $P(2<X<5)$

5. **25(24).** In Example 3.12, let $Y=$ the number of girls born before the experiment terminates. With $p=P(B)$ and $1-p=P(G)$, what is the pmf of Y ? [Hint: First list the possible values of Y, starting with the smallest, and proceed until you see a general formula.]

3.3

6. **29(28).** The pmf of the amount of memory X(GB) in a purchased flash drive was given in Example 3.13 as

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

Compute the following:

a. $E(X)$

b. $V(X)$ directly from the definition

c. The standard deviation of X

d. $V(X)$ using the shortcut formula

7. **33(32).** Let X be a Bemoulli rv with pmf as in Example 3.18.

a. Compute $E\left(X^{2}\right)$.

b. Show that $V(X)=p(1-p)$.

c. Compute $E\left(X^{9}\right)$.

8. **37(36).** The n candidates for a job have been ranked $1, 2, 3, \ldots, n$. Let $X =$ the rank of a randomly selected candidate, so that X has pmf

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

(this is called the discrete uniform distribution). Compute $E(X)$ and $V(X)$ using the shortcut formula. [Hint: The sum of the first n positive integers is $n(n+1)/2$, whereas the sum of their squares is $n(n+1)(2n+1)/6$.]

9. **41(40).** Use the definition in Expression (3.13) to prove that $V(aX+b) = a^2 \cdot \sigma_X^2$. [Hint: With $h(X) = aX + b$ $E[h(X)] = a\mu + b$ where $\mu = E(X)$.]

3.4

10. **63(57).**

a.Show that $b(x; n, 1-p) = b(n-x, n, p)$.

b.Show that $B(x; n, 1-p) = 1 - B(n-x-1; n, p)$. [Hint: At most x S's is equivalent to at least $(n-x)$ F's.]

c.What do parts (a) and (b) imply about the necessity of including values of p greater than 5 in Appendix Table A.1?

11. **67(61).**Refer to Chebyshev's inequality given in Exercise 44 . Calculate $P(|X-\mu| \geq k\sigma)$ for $k=2$ and $k=3$ when$X \sim \text{Bin}(20, .5)$, and compare to the corresponding upper bound. Repeat for $X \sim \text{Bin}(20, .75)$.

3.5

12. **73(67).**Twenty pairs of individuals playing in a bridge tournament have been seeded $1, \ldots, 20$. In the first part of the tournament, the 20 are randomly divided into 10 eastwest pairs and 10 north-south pairs.

a. What is the probability that x of the top 10 pairs end up playing east-west?

b. What is the probability that all of the top five pairs end up playing the same direction?

c. If there are $2n$ pairs, what is the pmf of $X =$ the number among the top n pairs who end up playing east-west? What are $E(X)$ and $V(X)$?

13. **77(71).** Three brothers and their wives decide to have children until each family has two female children. What is the pmf of $X =$ the total number of male children born to the brothers? What is $E(X)$, and how does it compare to the expected number of male children born to each brother?

3.6

14. **91(85).** Suppose that trees are distributed in a forest according to a two-dimensional Poisson process with parameter α, the expected number of trees per acre, equal to 80 .

a. What is the probability that in a certain quarter-acre plot, there will be at most 16 trees?

b. If the forest covers 85,000 acres, what is the expected number of trees in the forest?

c. Suppose you select a point in the forest and construct a circle of radius .1 mile. Let $X =$ the number of trees within that circular region. What is the pmf of X? [Hint: 1 sq mile = 640 acres.]

15. **93(87).**

a. In a Poisson process, what has to happen in both the time interval $(0, t)$ and the interval $(t, t + \Delta t)$ so that no events occur in the entire interval $(0, t + \Delta t)$? Use this and Assumptions 1-3 to write a relationship between $P_0(t + \Delta t)$ and $P_0(t)$.

b. Use the result of part (a) to write an expression for the difference $P_0(t + \Delta t) - P_0(t)$. Then divide by Δt and let $\Delta t \to 0$ to obtain an equation involving (d/dt) $P_0(t)$, the derivative of $P_0(t)$ with respect to t.

c. Verify that $P_0(t) = e^{-\alpha t}$ satisfies the equation of part (b).

d. It can be shown in a manner similar to parts (a) and (b) that the $P_k(t)$ s must satisfy the system of differential equations

$$\frac{d}{dt} P_k(t) = \alpha P_{k-1}(t) - \alpha P_k(t)$$
$$k = 1, 2, 3, \ldots$$

Verify that $P_k(t) = e^{-\alpha t}(\alpha t)^k / k!$ satisfies the system. (This is actually the only solution.)

Supplementary Exercises

16. **99(93).** A k-out-of-n system is one that will function if and only if at least k of the n individual components in the system function. If individual components function independently of one another, each with probability .9, what is the probability that a 3 -out-of-5 system functions?

17. **105(99).** The purchaser of a power-generating unit requires c consecutive successful start-ups before the unit will be accepted. Assume that the outcomes of individual startups are independent of one another. Let p denote the probability that any particular start-up is successful. The random variable of interest is $X =$ the number of startups that must be made prior to acceptance. Give the pmf of X for the case $c = 2$. If $p = .9$, what is $P(X \leq 8)$? [Hint: For $x \geq 5$, express $p(x)$ "recursively" in terms of thepmfevaluatedatthesmallervalues $x - 3, x - 4, \ldots, 2$.] (This problem was suggested by the article "Evaluation of a Start-Up Demonstration Test," J. Quality Technology, 1983: 103-106.)

18. **113(107).** A test for the presence of a certain disease has probability .20 of giving a false-positive reading (indicating that an individual has the disease when this is not the case) and probability . 10 of giving a false-negative result. Suppose that ten individuals are tested, five of whom have the disease and five of whom do not. Let $X =$ the number of positive readings that result.

a. Does X have a binomial distribution? Explain your reasoning.

b. What is the probability that exactly three of the ten test results are positive?

19. **117(111)**. A computer disk storage device has ten concentric tracks, numbered $1, 2, \ldots, 10$ from outermost to innermost, and a single access arm. Let $p_i =$ the probability that any particular request for data will take the arm to track $i (i = 1, \ldots, 10)$. Assume that the tracks accessed in successive seeks are independent. Let $X =$ the number of tracks over which the access arm passes during two successive requests (excluding the track that the arm has just left, so possible X values are $x = 0, 1, \ldots, 9$). Compute the pmf of X. [Hint: P (the arm is now on track i and $X = j$) $= P(X =$ jlarm nowon $i) \cdot p_i$. After the conditional probability is written in terms of $p_1, \ldots, p_{10}$, by the law of total probability, the desired probability is obtained by summing over i.]

20. **119(113).** Use the fact that

$$\sum_{\text{all } x} (x - \mu)^2 p(x) \geq \sum_{x : |x - \mu| = k\sigma} (x - \mu)^2 p(x)$$

to prove Chebyshev's inequality given in Exercise 44.

3 Chapter 4

4.1

1. **5.** A college professor never finishes his lecture before the end of the hour and always finishes his lectures within 2 min after the hour. Let $X =$ the time that elapses between the end of the hour and the end of the lecture and suppose the pdf of X is

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the value of k and draw the corresponding density curve. [Hint: Total area under the graph of $f(x)$ is 1.]

b. What is the probability that the lecture ends within 1 min of the end of the hour?

c. What is the probability that the lecture continues beyond the hour for between 60 and 90 sec?

d. What is the probability that the lecture continues for at least 90sec beyond the end of the hour?

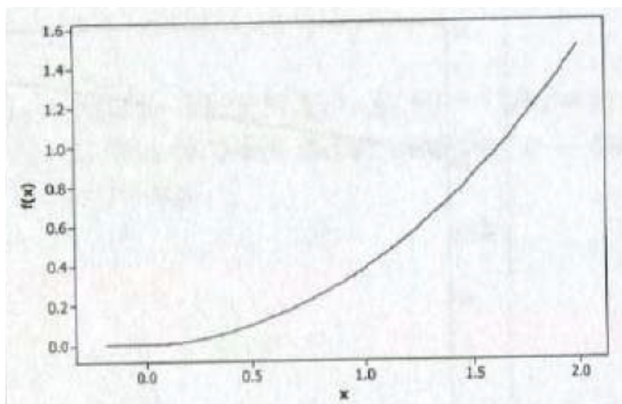


Figure 2: Question 5

4.2

2. **15.** Let X denote the amount of space occupied by an article placed in a $1-\mathrm{ft}^3$ packing container. The pdf of X is

$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Graph the pdf. Then obtain the cdf of X and graph it.

b. What is $P(X \leq .5)$ [i.e., $F(.5)$]?

c. Using the cdf from (a), what is $P(.25 < X \leq .5)$? What is $P(.25 \leq X \leq .5)$?

d. What is the 75 th percentile of the distribution?

e. Compute $E(X)$ and σ_X.

f. What is the probability that X is more than 1 standard deviation from its mean value?

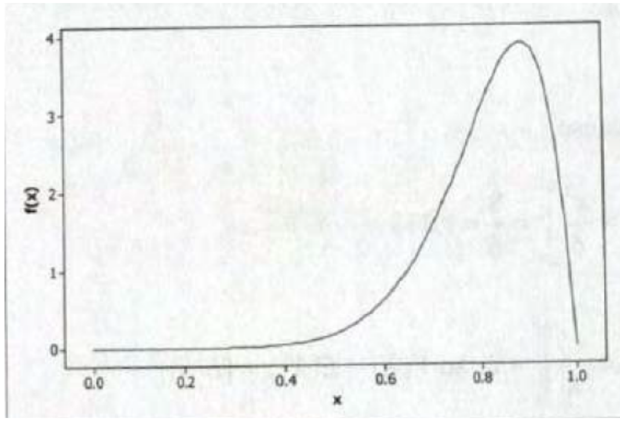


Figure 3: Question 15: pdf

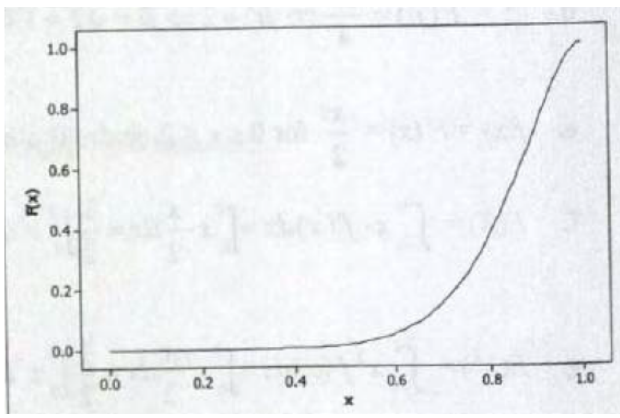


Figure 4: Question 15: cdf

3. **21.** An ecologist wishes to mark off a circular sampling region having radius 10 m. However, the radius of the resulting region is actually a random variable R with pdf

$$f(r)=\begin{cases}\frac{3}{4}\left[1-(10-r)^2\right] & 9 \leq r \leq 11 \\ 0 & \text{otherwise}\end{cases}$$

What is the expected area of the resulting circular region?

4.3

4. **31(29).** Determine z_α for the following values of α :

a. $\alpha=.0055$

b. $\alpha=.09$

c. $\alpha=.663$

5. **51(45).** Chebyshev's inequality, (see Exercise 44, Chapter 3), is valid for continuous as well as discrete distributions. It states that for any number k satisfying $k \geq 1, P(|X-\mu| \geq k\sigma) \leq 1/k^2$ (see Exercise 44 in Chapter 3 for an interpretation). Obtain this probability in the case of a normal distribution for $k=1,2$, and 3 , and compare to the upper bound.

6. **57(51).**

   a. Show that if X has a normal distribution with parameters μ and σ, then $Y = aX + b$ (a linear function of X) also has a normal distribution. What are the parameters of the distribution of Y [i.e., $E(Y)$ and $V(Y)$] ? [Hint: Write the cdf of $Y, P(Y \leq y)$, as an integral involving the pdf of X, and then differentiate with respect to y to get the pdf of Y.]

   b. If, when measured in °C, temperature is normally distributed with mean 115 and standard deviation 2, what can be said about the distribution of temperature measured in °F ?

4.4

7. **59(58).** Let $X =$ the time between two successive arrivals at the drive-up window of a local bank. If X has an exponential distribution with $\lambda = 1$ (which is identical to a standard gamma distribution with $\alpha = 1$), compute the following:

   a. The expected time between two successive arrivals

   b. The standard deviation of the time between successive arrivals

   c. $P(X \leq 4)$

   d. $P(2 \leq X \leq 5)$

8. **69(62).** A system consists of five identical components connected in series as shown (Fig. 5): As

Figure 5: Question 69

   soon as one component fails, the entire system will fail. Suppose each component has a lifetime that is exponentially distributed with $\lambda = .01$ and that components fail independently of one another. Define events $A_i = \{\text{ith component lasts at least } t \text{ hours }\}, i = 1, \ldots, 5$, so that the $A_i s$ are independent events. Let $X =$ the time at which the system fails - that is, the shortest (minimum) lifetime among the five components.

   a. The event $\{X \geq t\}$ is equivalent to what event involving $A_1, \ldots, A_5$?

   b. Using the independence of the A_i 's, compute $P(X \geq t)$. Then obtain $F(t) = P(X \leq t)$ and the pdf of X. What type of distribution does X have?

   c. Suppose there are n components, each having exponential lifetime with parameter λ. What type of distribution does X have?

9. **71(65).**

   a. The event $\left\{X^2 \leq y\right\}$ is equivalent to what event involving X itself?

   b. If X has a standard normal distribution, use part (a) to write the integral that equals $P\left(X^2 \leq y\right)$. Then differentiate this with respect to y to obtain the pdf of X^2 [the square of a $N(0,1)$ variable]. Finally, show that X^2 has a chi-squared distribution with $\nu = 1$ df [see (4.10)]. [Hint: Use the following identity.]

$$\frac{d}{dy}\left\{\int_{a(y)}^{b(y)} f(x) dx\right\} = f[b(y)] \cdot b'(y) - f[a(y)] \cdot a'(y)$$

4.5

10. **75(69).** Let X have a Weibull distribution with the pdf from Expression (4.11). Verify that $\mu = \beta\Gamma(1+1/\alpha)$. [Hint: In the integral for $E(X)$, make the change of variable $y = (x/\beta)^\alpha$, so that $x = \beta y^{1/\alpha}$.]

11. **85(77).** Let X have a standard beta density with parameters α and β.

    a. Verify the formula for $E(X)$ given in the section.

    b. Compute $E\left[(1-X)^m\right]$. If X represents the proportion of a substance consisting of a particular ingredient, what is the expected proportion that does not consist of this ingredient?

Supplementary Exercises

12. **99(89).** A 12-in. bar that is clamped at both ends is to be subjected to an increasing amount of stress until it snaps. Let $Y =$ the distance from the left end at which the break occurs. Suppose Y has pdf

$$f(y) = \begin{cases} \left(\frac{1}{24}\right) y \left(1 - \frac{y}{12}\right) & 0 \leq y \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

    Compute the following:

    a. The cdf of Y, and graph it.

    b. $P(Y \leq 4), P(Y > 6)$, and $P(4 \leq Y \leq 6)$

    c. $E(Y), E\left(Y^2\right)$, and $V(Y)$

    d. The probability that the break point occurs more than 2 in. from the expected break point.

    e. The expected length of the shorter segment when the break occurs.

13. **101(91).** The completion time X for a certain task has cdf $F(x)$ given by

$$\begin{cases} 0 & x < 0 \\ \frac{x^3}{3} & 0 \leq x < 1 \\ 1 - \frac{1}{2}\left(\frac{7}{3} - x\right)\left(\frac{7}{4} - \frac{3}{4}x\right) & 1 \leq x \leq \frac{7}{3} \\ 1 & x > \frac{7}{3} \end{cases}$$

    a. Obtain the pdf $f(x)$ and sketch its graph.

    b. Compute $P(.5 \leq X \leq 2)$.

    c. Compute $E(X)$.

14. **107(97).** Let X denote the temperature at which a certain chemical reaction takes place. Suppose that X has pdf

$$f(x) = \begin{cases} \frac{1}{9}\left(4 - x^2\right) & -1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

    a. Sketch the graph of $f(x)$.

    b. Determine the cdf and sketch it.

    c. Is 0 the median temperature at which the reaction takes place? If not, is the median temperature smaller or larger than 0 ?

    d. Suppose this reaction is independently carried out once in each of ten different labs and that the pdf of reaction time in each lab is as given. Let $Y =$ the number among the ten labs at which the temperature exceeds 1. What kind of distribution does Y have? (Give the names and values of any parameters.)

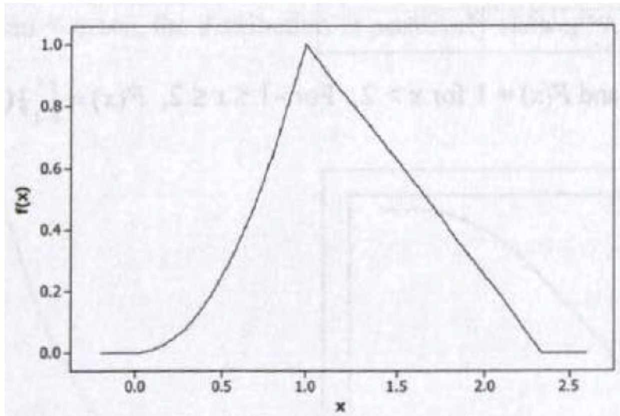


Figure 6: Question 101

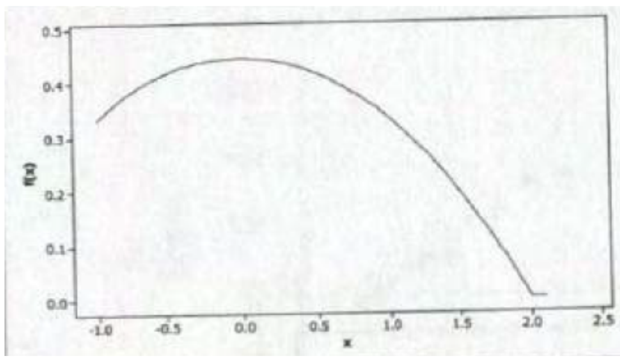


Figure 7: Question 107

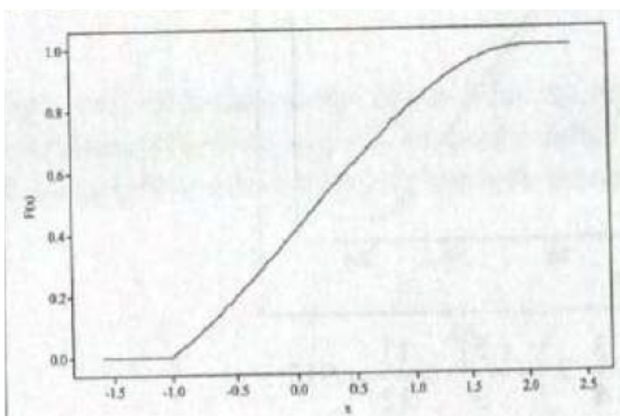


Figure 8: Question 107

15. **109(99).** The article "The Prediction of Corrosion by Statistical Analysis of Corrosion Profiles" (Corrosion Science, 1985: 305-315) suggests the following cdf for the depth X of the deepest pit in an experiment involving the exposure of carbon manganese steel to acidified seawater.

$$F(x,\alpha,\beta) = e^{-e^{(x-\alpha/\beta)}} - \infty < x < \infty$$

The authors propose the values $\alpha = 150$ and $\beta = 90$. Assume this to be the correct model.

a. What is the probability that the depth of the deepest pit is at most 150? At most 300? Between 150 and 300 ?

b. Below what value will the depth of the maximum pit be observed in 90% of all such experiments?

c. What is the density function of X ?

d. The density function can be shown to be unimodal (a single peak). Above what value on the measurement axis does this peak occur? (This value is the mode.) e. It can be shown that $E(X) \approx .5772\beta + \alpha$. What is the mean for the given values of α and β, and how does it compare to the median and mode? Sketch the graph of the density function. [Note: This is called the largest extreme value distribution.]

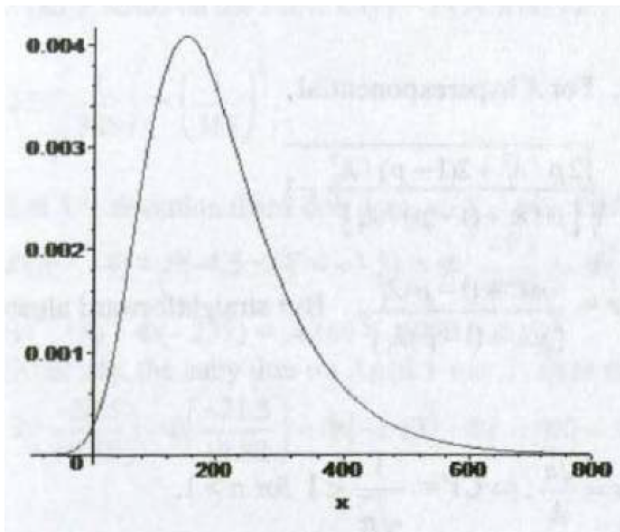


Figure 9: Question 109

16. **111(101).** The mode of a continuous distribution is the value x^*that maximizes $f(x)$.

    a. What is the mode of a normal distribution with parameters μ and σ ?

    b. Does the uniform distribution with parameters A and B have a single mode? Why or why not?

    c. What is the mode of an exponential distribution with parameter λ ? (Draw a picture.)

    d. If X has a gamma distribution with parameters α and β, and $\alpha > 1$, find the mode. [Hint: $\ln[f(x)]$ will be maximized iff $f(x)$ is, and it may be simpler to take the derivative of $\ln[f(x)]$.]

    e. What is the mode of a chi-squared distribution having ν degrees of freedom?

17. **115(105).** Let I_i be the input current to a transistor and I_0 be the output current. Then the current gain is proportional to $\ln(I_0/I_i)$. Suppose the constant of proportionality is 1 (which amounts to choosing a particular unit of measurement), so that current gain $= X = \ln(I_0/I_i)$. Assume X is normally distributed with $\mu = 1$ and $\sigma = .05$.

    a. What type of distribution does the ratio I_0/I_i have?

    b. What is the probability that the output current is more than twice the input current?

    c. What are the expected value and variance of the ratio of output to input current?

18. **117(107).** Let Z have a standard normal distribution and define a new rv Y by $Y = \sigma Z + \mu$. Show that Y has a normal distribution with parameters μ and σ. [Hint: $Y \leq y$ iff Z $\leq$? Use this to find the cdf of Y and then differentiate it with respect to y.]

19. **119(109).** In Exercises 117 and 118 , as well as many other situations, one has the pdf $f(x)$ of X and wishes to know the pdf of $y = h(X)$. Assume that $h(\cdot)$ is an invertible function, so that $y = h(x)$ can be solved for x to yield $x = k(y)$. Then it can be shown that the pdf of Y is

$$g(y) = f[k(y)] \cdot |k'(y)|$$

a. If X has a uniform distribution with $A = 0$ and $B = 1$, derive the pdf of $Y = -\ln(X)$.

b. Work Exercise 117, using this result.

c. Work Exercise 118(b), using this result.

20. **123(111).** Let U have a uniform distribution on the interval $[0, 1]$. Then observed values having this distribution can be obtained from a computer's random number generator. Let $X = -(1/\lambda)\ln(1 - U)$.

a. Show that X has an exponential distribution with parameter λ. [Hint: The odf of X is $F(x) = P(X \simeq x)$; $X \leq x$ is equivalent to $U \leq$?]

b. How would you use part (a) and a random number generator to obtain observed values from an exponential distribution with parameter $\lambda = 10$?

4 Chapter 7

7.1

1. Suppose that a random sample of 50 bottles of a particular brand of cough syrup is selected and the alcohol content of each bottle is determined. Let μ denote the average alcohol content for the population of all bottles of the brand under study. Suppose that the resulting 95% confidence interval is (7.8, 9.4).

   a. Would a 90% confidence interval calculated from this same sample have been narrower or wider than the given interval? Explain your reasoning.

   b. Consider the following statement: There is a 95% chance that μ is between 7.8 and 9.4. Is this statement correct? Why or why not?

   c. Consider the following statement: We can be highly confident that 95% of all bottles of this type of cough syrup have an alcohol content that is between 7.8 and 9.4. Is this statement correct? Why or why not?

   d. Consider the following statement: If the process of selecting a sample of size 50 and then computing the corresponding 95% interval is repeated 100 times, 95 of the resulting intervals will include μ. Is this statement correct? Why or why not?

2. Assume that the helium porosity (in percentage) of coal samples taken from any particular seam is normally distributed with true standard deviation .75.

   a. Compute a 95% CI for the true average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85.

   b. Compute a 98% CI for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.

   c. How large a sample size is necessary if the width of the 95% interval is to be .40?

   d. What sample size is necessary to estimate true average porosity to within .2 with 99% confidence?

3. A random sample of $n = 15$ heat pumps of a certain type yielded the following observations on lifetime (in years):

2.0	1.3	6.0	1.9	5.1	.4	1.0	5.3
15.7	.7	4.8	.9	12.2	5.3	.6	

   a. Assume that the lifetime distribution is exponential and use an argument parallel to that of Example 7.5 to obtain a 95% CI for expected (true average) lifetime.

   b. How should the interval of part (a) be altered to achieve a confidence level of 99% ?

   c. What is a 95% CI for the standard deviation of the lifetime distribution? [Hint: What is the standard deviation of an exponential random variable?]

4. Consider the next 1000 95% CIs for μ that a statistical consultant will obtain for various clients. Suppose the data sets on which the intervals are based are selected independently of one another. How many of these 1000 intervals do you expect to capture the corresponding value of μ? What is the probability that between 940 and 960 of these intervals contain the corresponding value of μ? [Hint: Let Y = the number among the 1000 intervals that contain μ. What kind of random variable is Y?]

7.2

5. The article "Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products" (Indoor Air, 2006: 65-73) reported that for a sample of 50 kitchens with gas cooking appliances monitored during a oneweek period, the sample mean CO_2 level (ppm) was 654.16, and the sample standard deviation was 164.43.

a. Calculate and interpret a 95% (two-sided) confidence interval for true average CO_2 level in the population of all homes from which the sample was selected.

b. Suppose the investigators had made a rough guess of 175 for the value of s before collecting data. What sample size would be necessary to obtain an interval width of 50ppm for a confidence level of 95% ?

6. The Pew Forum on Religion and Public Life reported on Dec. 9, 2009, that in a survey of 2003 American adults, 25% said they believed in astrology.

   a. Calculate and interpret a confidence interval at the 99% confidence level for the proportion of all adult Americans who believe in astrology.

   b. What sample size would be required for the width of a 99% CI to be at most .05 irrespective of the value of $\hat{p}$?

7. The superintendent of a large school district, having once had a course in probability and statistics, believes that the number of teachers absent on any given day has a Poisson distribution with parameter μ. Use the accompanying data on absences for 50 days to obtain a largesample CI for μ. [Hint: The mean and variance of a Poisson variable both equal μ, so

$$Z = \frac{\bar{X} - \mu}{\sqrt{\mu/n}}$$

has approximately a standard normal distribution. Now proceed as in the derivation of the interval for p by making a probability statement (with probability $1 - \alpha$) and solving the resulting inequalities for μ (see the argument just after (7.10)).]

Number of absences	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	4	8	10	8	7	5	3	2	1	1

8. Suppose that $X_1, \ldots, X_n$ form a random sample from the exponential distribution with unknown mean μ. Describe a method for constructing a confidence interval for μ with a specified confidence coefficient $\gamma\,(0 < \gamma < 1)$. Hint: Determine constants c_1 and c_2 such that $\mathrm{P}(c_1 < (1/\mu)\sum_{i=1}^{n} X_i < c_2) = \gamma$.

7.3

9. A random sample of $n = 8$ E-glass fiber test specimens of a certain type yielded a sample mean interfacial shear yield stress of 30.2 and a sample standard deviation of 3.1 ("On Interfacial Failure in Notched Unidirectional Glass/Epoxy Composites," J. of Composite Materials, 1985: 276-286). Assuming that interfacial shear yield stress is normally distributed, compute a 95% CI for true average stress (as did the authors of the cited article).

10. A sample of 14 joint specimens of a particular type gave a sample mean proportional limit stress of 8.48 MPa and a sample standard deviation of .79 MPa ("Characterization of Bearing Strength Factors in Pegged Timber Connections," J. of Structural Engr., 1997: 326-332).

    a. Calculate and interpret a 95% lower confidence bound for the true average proportional limit stress of all such joints. What, if any, assumptions did you make about the distribution of proportional limit stress?

    b. Calculate and interpret a 95% lower prediction bound for the proportional limit stress of a single joint of this type.

11. A sample of 26 offshore oil workers took part in a simulated escape exercise, resulting in the accompanying data on time (sec) to complete the escape ("Oxygen Consumption and Ventilation During Escape from an Offshore Platform," Ergonomics, 1997: 281-292):

389	356	359	363	375	424	325	394	402
373	373	370	364	366	364	325	339	393
392	369	374	359	356	403	334	397	

a. Calculate an upper confidence bound for population mean escape time using a confidence level of 95%.

b. Calculate an upper prediction bound for the escape time of a single additional worker using a prediction level of 95%. How does this bound compare with the confidence bound of part (a)?

c. Suppose that two additional workers will be chosen to participate in the simulated escape exercise. Denote their escape times by X_{27} and X_{28}, and let $\bar{X}_{\text{new}}$ denote the average of these two values. Modify the formula for a PI for a single x value to obtain a PI for $\bar{X}_{\text{new}}$, and calculate a 95% two-sided interval based on the given escape data.

7.4

12. The amount of lateral expansion (mils) was determined for a sample of $n = 9$ pulsed-power gas metal arc welds used in LNG ship containment tanks. The resulting sample standard deviation was $s = 2.81$ mils. Assuming normality, derive a 95% CI for σ^2 and for σ.

13. Wire electrical-discharge machining (WEDM) is a process used to manufacture conductive hard metal components. It uses a continuously moving wire that serves as an electrode. Coating on the wire electrode allows for cooling of the wire electrode core and provides an improved cutting performance. The article "HighPerformance Wire Electrodes for Wire ElectricalDischarge Machining - A Review" (J. of Engr. Manuf., 2012: 1757-1773) gave the following sample observations on total coating layer thickness (in μm) of eight wire electrodes used for WEDM:

21	16	29	35	42	24	24	25

Calculate a 99% CI for the standard deviation of the coating layer thickness distribution. Is this interval valid whatever the nature of the distribution? Explain.

Supplementary Exercises

14. Suppose a certain type of component has a lifetime distribution that is exponential with parameter λ so that expected lifetime is $\mu = 1/\lambda$. A sample of n such components is selected, and each is put into operation. Suppose component lifetimes are independent. Let Y_1 denote the time at which the first failure occurs, Y_2 the time at which the second failure occurs, and so on, so that $T_r = Y_1 + \cdots + Y_r + (n-r)Y_r$ is the total accumulated lifetime at termination. Then it can be shown that $2\lambda T_r$ has a chi-squared distribution with $2r$ df. Use this fact to develop a $100(1-\alpha)\%$ CI formula for true average lifetime $1/\lambda$. As an example, suppose 20 components are tested and $r = 10$. Then if the first ten failure times are 11, 15, 29, 33, 35, 40, 47, 55, 58, and 72. Compute a 95% CI from the data.

15. Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous probability distribution having median $\tilde{\mu}$ (so that $P(X_i \leq \tilde{\mu}) = P(X_i \geq \tilde{\mu}) = .5)$.

a. Show that

$$P(\min(X_i) < \tilde{\mu} < \max(X_i)) = 1 - \left(\frac{1}{2}\right)^{n-1}$$

so that $(\min(x_i), \max(x_i))$ is a $100(1-\alpha)\%$ confidence interval for $\tilde{\mu}$ with $\alpha = \left(\frac{1}{2}\right)^{n-1}$. [Hint: The complement of the event $\{\min(X_i) < \tilde{\mu} < \max(X_i)\}$ is $\{\max(X_i) \leq \tilde{\mu}\} \cup \{\min(X_i) \geq \tilde{\mu}\}$. But $\max(X_i) \leq \tilde{\mu}$ iff $X_i \leq \tilde{\mu}$ for all i.]

b. For each of six normal male infants, the amount of the amino acid alanine (mg/100 mL) was determined while the infants were on an isoleucine-free diet, resulting in the following data:

$$2.84 \quad 3.54 \quad 2.80 \quad 1.44 \quad 2.94 \quad 2.70$$

Compute a 97% CI for the true median amount of alanine for infants on such a diet ("The Essential Amino Acid Requirements of Infants," Amer. J. of Nutrition, 1964: 322-330).

c. Let $x_{(2)}$ denote the second smallest of the x_i's and $x_{(n-1)}$ denote the second largest of the x_i's. What is the confidence level of the interval $\left(x_{(2)}, x_{(n-1)}\right)$ for $\tilde{\mu}$?

16. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[0, \theta]$, so that

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

Then if $Y = \max(X_i)$, it can be shown that the rv $U = Y/\theta$ has density function

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Use $f_U(u)$ to verify that

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} \leq (1-\alpha/2)^{1/n}\right) = 1-\alpha$$

and use this to derive a $100(1-\alpha)\%$ CI for θ.

b. Verify that $P\left(\alpha^{1/n} \leq Y/\theta \leq 1\right) = 1-\alpha$, and derive a $100(1-\alpha)\%$ CI for θ based on this probability statement.

c. Which of the two intervals derived previously is shorter? If my waiting time for a morning bus is uniformly distributed and observed waiting times are $x_1 = 4.2, x_2 = 3.5, x_3 = 1.7, x_4 = 1.2$, and $x_5 = 2.4$, derive a 95% CI for θ by using the shorter of the two intervals.

17. Let $0 \leq \gamma \leq \alpha$. Then a $100(1-\alpha)\%$ CI for μ when n is large is

$$\left(\bar{x} - z_\gamma \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha-\gamma} \cdot \frac{s}{\sqrt{n}}\right)$$

The choice $\gamma = \alpha/2$ yields the usual interval derived in Section 7.2; if $\gamma \neq \alpha/2$, this interval is not symmetric about $\bar{x}$. The width of this interval is $w = s\left(z_\gamma + z_{\alpha-\gamma}\right)/\sqrt{n}$. Show that w is minimized for the choice $\gamma = \alpha/2$, so that the symmetric interval is the shortest. [Hints: (a) By definition of $z_\alpha, \Phi(z_\alpha) = 1-\alpha$, so that $z_\alpha = \Phi^{-1}(1-\alpha)$; (b) the relationship between the derivative of a function $y = f(x)$ and the inverse function $x = f^{-1}(y)$ is $(d/dy)f^{-1}(y) = 1/f'(x)$.]

18. a. Use the results of Example 7.5 to obtain a 95% lower confidence bound for the parameter λ of an exponential distribution, and calculate the bound based on the data given in the example.

b. If lifetime X has an exponential distribution, the probability that lifetime exceeds t is $P(X > t) = e^{-\lambda t}$. Use the result of part (a) to obtain a 95% lower confidence bound for the probability that breakdown time exceeds 100 min.

19. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with $X_1 \sim \text{Exponential}(\lambda)$. Show that $2\lambda\sum_{i=1}^n X_i$ has a chi-squared distribution with $2n$ degrees of freedom, i.e., $2\lambda\sum_{i=1}^n X_i \sim \chi^2(2n)$.

5 Chapter 8

8.1

1. Let μ denote the true average radioactivity level (pico-curies per liter). The value 5pCi/L is considered the dividing line between safe and unsafe water. Which of the following two options would you recommend testing:
(a) $H_0 : \mu = 5$ versus $H_a : \mu > 5$; or
(b) $H_0 : \mu = 5$ versus $H_a : \mu < 5$?

2. Before agreeing to purchase a large order of polyethylene sheaths for a particular type of high-pressure oilfilled submarine power cable, a company wants to see conclusive evidence that the true standard deviation of sheath thickness is less than .05 mm. What hypotheses should be tested, and why? In this context, what are the type I and type II errors?

3. Two different companies have applied to provide cable television service in a certain region. Let p denote the proportion of all potential subscribers who favor the first company over the second. Consider testing $H_0 : p = .5$ versus $H_\mathrm{a} : p \neq .5$ based on a random sample of 25 individuals. Let the test statistic X be the number in the sample who favor the first company and x represent the observed value of X.

    a. Describe type I and II errors in the context of this problem situation.
    b. Suppose that $x = 6$. Which values of X are at least as contradictory to H_0 as this one?
    c. What is the probability distribution of the test statistic X when H_0 is true? Use it to compute the P-value when $x = 6$.
    d. If H_0 is to be rejected when P-value $\leq .044$, compute the probability of a type II error when $p = .4$, again when $p = .3$, and also when $p = .6$ and $p = .7$. [Hint: P-value $> .044$ is equivalent to what inequalities involving x (see Example 8.4)?]
    e. Using the test procedure of (d), what would you conclude if 6 of the 25 queried favored company 1?

4. The calibration of a scale is to be checked by weighing a 10-kg test specimen 25 times. Suppose that the results of different weighings are independent of one another and that the weight on each trial is normally distributed with $\sigma = .200$ kg. Let μ denote the true average weight reading on the scale.

    a. What hypotheses should be tested?
    b. With the sample mean itself as the test statistic, what is the P-value when $\bar{x} = 9.85$, and what would you conclude at significance level .01?
    c. For a test with $\alpha = .01$, what is the probability that recalibration is judged unnecessary when in fact $\mu = 10.1$? When $\mu = 9.8$?

8.2

5. Let μ denote the true average reaction time to a certain stimulus. For a z test of $H_0 : \mu = 5$ versus $H_\mathrm{a} : \mu > 5$, determine the P-value for each of the following values of the z test statistic.
a. 1.42 b. 90 c. 1.96 d. 2.48 e. $-.11$

6. Answer the following questions for the tire problem in Example 8.7.

    a. If $\bar{x} = 30,960$ and a level $\alpha = .01$ test is used, what is the decision?
    b. If a level .01 test is used, what is $\beta(30,500)$?
    c. If a level .01 test is used and it is also required that $\beta(30,500) = .05$, what sample size n is necessary?

d. If $\bar{x} = 30,960$, what is the smallest α at which H_0 can be rejected (based on $n = 16$)?

7. The melting point of each of 16 samples of a certain brand of hydrogenated vegetable oil was determined, resulting in $\bar{x} = 94.32$. Assume that the distribution of the melting point is normal with $\sigma = 1.20$.

   a. Test $H_0 : \mu = 95$ versus $H_{\text{a}} : \mu \neq 95$ using a two tailed level .01 test.

   b. If a level .01 test is used, what is $\beta(94)$, the probability of a type II error when $\mu = 94$?

   c. What value of n is necessary to ensure that $\beta(94) = .1$ when $\alpha = .01$?

8. The desired percentage of SiO_2 in a certain type of aluminous cement is 5.5. To test whether the true average percentage is 5.5 for a particular production facility, 16 independently obtained samples are analyzed. Suppose that the percentage of SiO_2 in a sample is normally distributed with $\sigma = .3$ and that $\bar{x} = 5.25$.

   a. Does this indicate conclusively that the true average percentage differs from 5.5?

   b. If the true average percentage is $\mu = 5.6$ and a level $\alpha = .01$ test based on $n = 16$ is used, what is the probability of detecting this departure from H_0?

   c. What value of n is required to satisfy $\alpha = .01$ and $\beta(5.6) = .01$?

9. Show that for any $\Delta > 0$, when the population distribution is normal and σ is known, the two-tailed test satisfies $\beta\left(\mu_0 - \Delta\right) = \beta\left(\mu_0 + \Delta\right)$, so that $\beta\left(\mu'\right)$ is symmetric about μ_0.

10. For a fixed alternative value μ', show that $\beta\left(\mu'\right) \to 0$ as $n \to \infty$ for either a one-tailed or a two-tailed z test in the case of a normal population distribution with known σ.

8.3

11. The true average diameter of ball bearings of a certain type is supposed to be .5 in. A one-sample t test will be carried out to see whether this is the case. What conclusion is appropriate in each of the following situations?
    a. $n = 13, t = 1.6, \alpha = .05$
    b. $n = 13, t = -1.6, \alpha = .05$
    c. $n = 25, t = -2.6, \alpha = .01$
    d. $n = 25, t = -3.9$

12. The paint used to make lines on roads must reflect enough light to be clearly visible at night. Let μ denote the true average reflectometer reading for a new type of paint under consideration. A test of $H_0 : \mu = 20$ versus $H_{\text{a}} : \mu > 20$ will be based on a random sample of size n from a normal population distribution. What conclusion is appropriate in each of the following situations?
    a. $n = 15, t = 3.2, \alpha = .05$
    b. $n = 9, t = 1.8, \alpha = .01$
    c. $n = 24, t = -.2$

13. The article "Uncertainty Estimation in Railway Track Life-Cycle Cost" (J. of Rail and Rapid Transit, 2009) presented the following data on time to repair (min) a rail break in the high rail on a curved track of a certain railway line.

    159 120 480 149 270 547 340 43 228 202 240 218

    A normal probability plot of the data shows a reasonably linear pattern, so it is plausible that the population distribution of repair time is at least approximately normal. The sample mean and standard deviation are 249.7 and 145.1, respectively.

a. Is there compelling evidence for concluding that true average repair time exceeds 200 min? Carry out a test of hypotheses using a significance level of .05.

b. Using $\sigma = 150$, what is the type II error probability of the test used in (a) when true average repair time is actually 300 min ? That is, what is $\beta(300)$?

14. Reconsider the accompanying sample data on expense ratio (%) for large-cap growth mutual funds first introduced in Exercise 1.53.

0.52	1.06	1.26	2.17	1.55	0.99	1.10	1.07	1.81	2.05
0.91	0.79	1.39	0.62	1.52	1.02	1.10	1.78	1.01	1.15

A normal probability plot shows a reasonably linear pattern.

a. Is there compelling evidence for concluding that the population mean expense ratio exceeds 1% ? Carry out a test of the relevant hypotheses using a significance level of .01.

b. Referring back to (a), describe in context type I and II errors and say which error you might have made in reaching your conclusion. The source from which the data was obtained reported that $\mu = 1.33$ for the population of all 762 such funds. So did you actually commit an error in reaching your conclusion?

c. Supposing that $\sigma = .5$, determine and interpret the power of the test in (a) for the actual value of μ stated in (b).

8.4

15. Consider using a z test to test $H_0 : p = .6$. Determine the P-value in each of the following situations.
a. $H_{\mathrm{a}} : p > .6, z = 1.47$
b. $H_{\mathrm{a}} : p < .6, z = -2.70$
c. $H_{\mathrm{a}} : p \neq .6, z = -2.70$
d. $H_{\mathrm{a}} : p < .6, z = .25$

16. A common characterization of obese individuals is that their body mass index is at least 30[BMI = weight/(height)^2, where height is in meters and weight is in kilograms]. The article "The Impact of Obesity on Illness Absence and Productivity in an IndustrialPopulation of Petrochemical Workers" (Annals of Epidemiology, 2008: 8-14) reported that in a sample of female workers, 262 had BMIs of less than 25,159 had BMIs that were at least 25 but less than 30 , and 120 had BMIs exceeding 30 . Is there compelling evidence for concluding that more than 20% of the individuals in the sampled population are obese?

a. State and test appropriate hypotheses with a significance level of .05.

b. Explain in the context of this scenario what constitutes type I and II errors.

c. What is the probability of not concluding that more than 20% of the population is obese when the actual percentage of obese individuals is 25% ?

17. A plan for an executive travelers' club has been developed by an airline on the premise that 5% of its current customers would qualify for membership. A random sample of 500 customers yielded 40 who would qualify.

a. Using this data, test at level .01 the null hypothesis that the company's premise is correct against the alternative that it is not correct.

b. What is the probability that when the test of part (a) is used, the company's premise will be judged correct when in fact 10% of all current customers qualify?

Supplementary Exercises

18. When $X_1, X_2, \ldots, X_n$ are independent Poisson variables, each with parameter μ, and n is large, the sample mean $\bar{X}$ has approximately a normal distribution with $\mu = E(\bar{X})$ and $V(\bar{X}) = \mu/n$. This implies that

$$Z = \frac{\bar{X} - \mu}{\sqrt{\mu/n}}$$

has approximately a standard normal distribution. For testing $H_0 : \mu = \mu_0$, we can replace μ by μ_0 in the equation for Z to obtain a test statistic. This statistic is actually preferred to the large-sample statistic with denominator $S/\sqrt{n}$ (when the X_i 's are Poisson) because it is tailored explicitly to the Poisson assumption. If the number of requests for consulting received by a certain statistician during a 5-day work week has a Poisson distribution and the total number of consulting requests during a 36-week period is 160 , does this suggest that the true average number of weekly requests exceeds 4.0 ? Test using $\alpha = .02$.

19. When the population distribution is normal and n is large, the sample standard deviation S has approximately a normal distribution with $E(S) \approx \sigma$ and $V(S) \approx \sigma^2/(2n)$. We already know that in this case, for any $n, \bar{X}$ is normal with $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$.

    a. Assuming that the underlying distribution is normal, what is an approximately unbiased estimator of the 99th percentile $\theta = \mu + 2.33\sigma$?

    b. When the X_i's are normal, it can be shown that $\bar{X}$ and S are independent rv's (one measures location whereas the other measures spread). Use this to compute $V(\hat{\theta})$ and $\sigma_{\hat{\theta}}$ for the estimator $\hat{\theta}$ of part (a). What is the estimated standard error $\hat{\sigma}_{\hat{\theta}}$?

    c. Write a test statistic for testing $H_0 : \theta = \theta_0$ that has approximately a standard normal distribution when H_0 is true. If soil pH is normally distributed in a certain region and 64 soil samples yield $\bar{x} = 6.33, s = .16$, does this provide strong evidence for concluding that at least 99% of all possible samples would have a pH of less than 6.75? Test using $\alpha = .01$.

20. Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with parameter λ. Then it can be shown that $2\lambda\Sigma X_i$ has a chi-squared distribution with $\nu = 2n$ (by first showing that $2\lambda X_i$ has a chi-squared distribution with $v = 2$).

    a. Use this fact to obtain a test statistic for testing $H_0 : \mu = \mu_0$. Then explain how you would determine the P-value when the alternative hypothesis is $H_\mathrm{a} : \mu < \mu_0$. [Hint: $E(X_i) = \mu = 1/\lambda$, so $\mu = \mu_0$ is equivalent to $\lambda = 1/\mu_0$.]

    b. Suppose that ten identical components, each having exponentially distributed time until failure, are tested. The resulting failure times are

    95 16 11 3 42 71 225 64 87 123

    Use the test procedure of part (a) to decide whether the data strongly suggests that the true average lifetime is less than the previously claimed value of 75 . [Hint: Consult Table A.7.]

6 Chapter 9

9.1

1. **5.** Persons having Reynaud's syndrome are apt to suffer a sudden impairment of blood circulation in fingers and toes. In an experiment to study the extent of this impairment, each subject immersed a forefinger in water and the resulting heat output $\left(\mathrm{cal/cm^2/min}\right)$ was measured. For $m = 10$ subjects with the syndrome, the average heat output was $\bar{x} = .64$, and for $n = 10$ nonsufferers, the average output was 2.05. Let μ_1 and μ_2 denote the true average heat outputs for the two types of subjects. Assume that the two distributions of heat output are normal with $\sigma_1 = .2$ and $\sigma_2 = .4$.

   a. Consider testing $H_0 : \mu_1 - \mu_2 = -1.0$ versus $H_2 : \mu_1 - \mu_2 < -1.0$ at level .01. Describe in words what H_2 says, and then carry out the test.

   b. What is the probability of a type II error when the actual difference between μ_1 and μ_2 is $\mu_1 - \mu_2 = -1.2$?

   c. Assuming that $m = n$, what sample sizes are required to ensure that $\beta = .1$ when $\mu_1 - \mu_2 = -1.2$?

2. **11.** The level of lead in the blood was determined for a sample of 152 male hazardous-waste workers ages $20 - 30$ and also for a sample of 86 female workers, resulting in a mean $\pm$ standard error of 5.5 ± 0.3 for the men and 3.8 ± 0.2 for the women (”Temporal Changes in Blood Lead Levels of Hazardous Waste Workers in New Jersey, 1984-1987,” Environ. Monitoring and Assessment, 1993: 99-107). Calculate an estimate of the difference between true average blood lead levels for male and female workers in a way that provides information about reliability and precision.

9.2

3. **19.** Suppose μ_1 and μ_2 are true mean stopping distances at 50mph for cars of a certain type equipped with two different types of braking systems. Use the two-sample t test at significance level .01 to test $H_0 : \mu_1 - \mu_2 = -10$ versus $H_\mathrm{a} : \mu_1 - \mu_2 < -10$ for the following data: $m = 6, \quad \bar{x} = 115.7, s_1 = 5.03, n = 6, \bar{y} = 129.3, \quad \text{and } s_2 = 5.38$.

4. **23.** Fusible interlinings are being used with increasing frequency to support outer fabrics and improve the shape and drape of various pieces of clothing. The article ”Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments” (Textile Res. J., 1997: 137-142) gave the accompanying data on extensibility (%) at 100gm/cm for both high-quality (H) fabric and poor-quality (P) fabric specimens.

H	1.2	.9	.7	1.0	1.7	1.7	1.1	.9	1.7
	1.9	1.3	2.1	1.6	1.8	1.4	1.3	1.9	1.6
	.8	2.0	1.7	1.6	2.3	2.0			
P	1.6	1.5	1.1	2.1	1.5	1.3	1.0	2.6	

   a. Construct normal probability plots to verify the plausibility of both samples having been selected from normal population distributions.

   b. Construct a comparative boxplot. Does it suggest that there is a difference between true average extensibility for high-quality fabric specimens and that for poor-quality specimens?

   c. The sample mean and standard deviation for the highquality sample are 1.508 and .444, respectively, and those for the poor-quality sample are 1.588 and .530. Use the two-sample t test to decide whether true average extensibility differs for the two types of fabric.

Normal Probability Plot for High Quality Fabric (Fig. 10)

Normal Probability Plot for Poor Quality Fabric (Fig. 11)

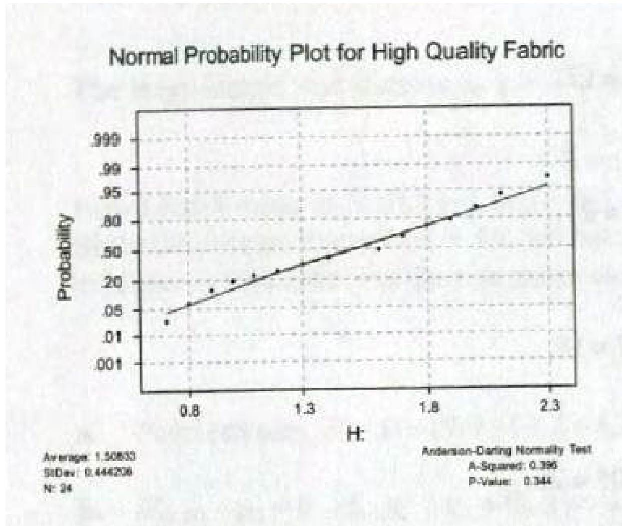


Figure 10: Question 23

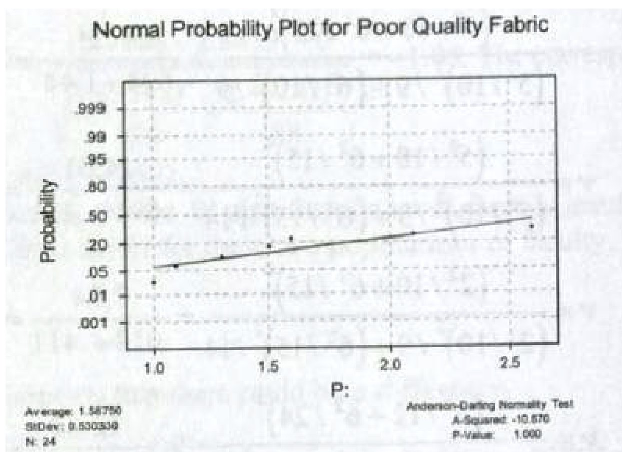


Figure 11: Question 23

Comparative Box Plot for High Quality and Poor Quality Fabric (Fig. 12)

5. **29.** The article Strength of Beverage Cans and Plastic Bottles" (J. of Testing and Evaluation, 1993: 129-131)"Effect of Internal Gas Pressure on the Compression includes the accompanying data on compression strength (lb) for a sample of 12 -oz aluminum cans filled with strawberry drink and another sample filled with cola Does the data suggest that the extra carbonation of cola results in a higher average compression strength? Base your answer on a P-value. What assumptions are necessary for your analysis?

Beverage	Sample Size	Sample Mean	Sample SD
Strawberry drink	15	540	21
Cola	15	554	15

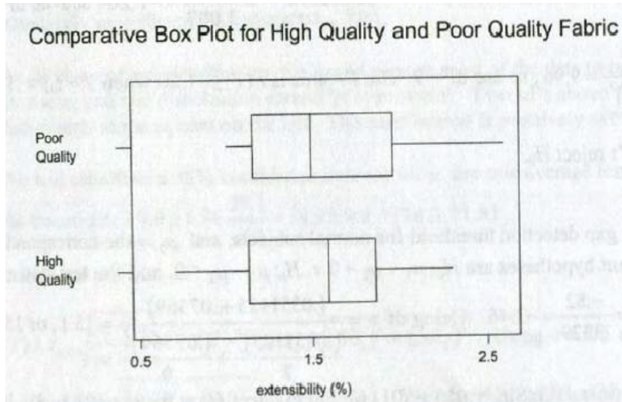


Figure 12: Question 23

9.3

6. **39.** Scientists and engineers frequently wish to compare two different techniques for measuring or determining the value of a variable. In such situations, interest centers on testing whether the mean difference in measurements is zero. The article "Evaluation of the Deuterium Dilution Technique Against the Test Weighing Procedure for the Determination of Breast Milk Intake" (Amer. J. of Clinical Nutr., 1983: 996-1003) reports the accompanying data on amount of milk ingested by each of 14 randomly selected infants.

	Infant				
	1	2	3	4	**5**
DD method	1509	1418	1561	1556	2169
TW method	1498	1254	1336	1565	2000
Difference	11	164	225	−9	169

	Infant				
	6	7	8	9	10
DD method	1760	1098	1198	1479	1281
TW method	1318	1410	1129	1342	1124
Difference	442	−312	69	137	157

	Infant			
	11	12	13	14
DD method	1414	1954	2174	2058
TW method	1468	1604	1722	1518
Difference	−54	350	452	540

a. Is it plausible that the population distribution of differences is normal?

b. Does it appear that the true average difference between intake values measured by the two methods is something other than zero? Determine the P-value of the test, and use it to reach a conclusion at significance level .05.

7. **43.** Cushing's disease is characterized by muscular weakness due to adrenal or pituitary dysfunction. To provide effective treatment, it is important to detect childhood Cushing's disease as early as possible. Age at onset of symptoms and age at diagnosis (months) for 15 children suffering from the disease were given in the article "Treatment of Cushing's Disease in Childhood and Adolescence by Transphenoidal Microadenomectomy" (New Emgl. J. of Med., 1984: 889). Here are the values of the differences between age at onset of symptoms and age at diagnosis:

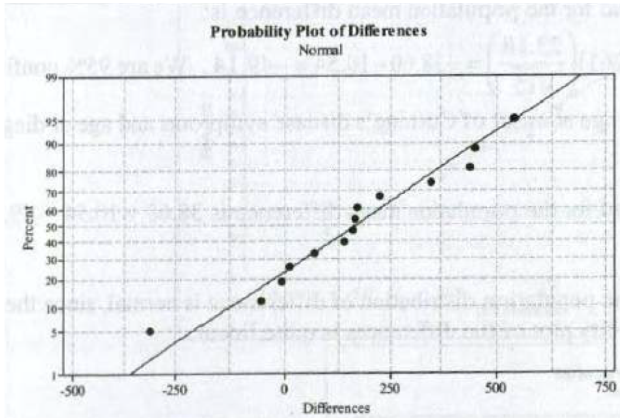


Figure 13: Question 39

−24	−12	−55	−15	−30	−60	−14	−21
−48	−12	−25	−53	−61	−69	−80	

a. Does the accompanying normal probability plot (Fig. 14) cast strong doubt on the approximate normality of the population distribution of differences?

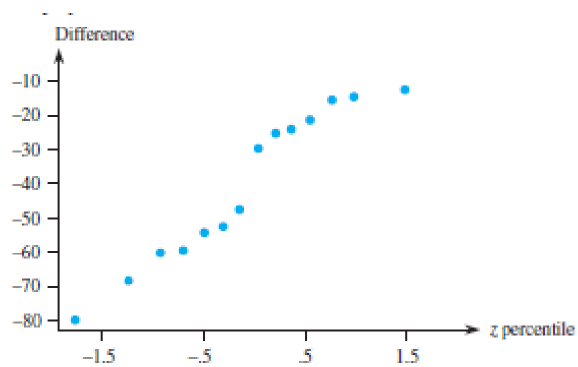


Figure 14: Question 43

b. Calculate a lower 95% confidence bound for the population mean difference, and interpret the resulting bound.

c. Suppose the (age at diagnosis) - (age at onset) differences had been calculated. What would be a 95% upper confidence bound for the corresponding population mean difference?

9.4

8. **55(53).** In medical investigations, the ratio $\theta = p_1/p_2$ is often of more interest than the difference $p_1 - p_2$ (e.g., individuals given treatment 1 are how many times as likely to recover as those given treatment 2?). Let $\hat{\theta} = \hat{p}_1/\hat{p}_2$. When m and n are both large, the statistic $\ln(\hat{\theta})$ has approximately a normal distribution with approximate mean value $\ln(\theta)$ and approximate standard deviation $[(m-x)/(mx)+(n-y)/(ny)]^{1/2}$.

   a. Use these facts to obtain a large-sample 95% CI formula for estimating $\ln(\theta)$, and then a CI for θ itself.

b. Return to the heart-attack data of Example 1.3, and calculate an interval of plausible values for θ at the 95% confidence level. What does this interval suggest about the efficacy of the aspirin treatment?

9. **57(55).** Two different types of alloy, A and B, have been used to manufacture experimental specimens of a small tension link to be used in a certain engineering application. The ultimate strength (ksi) of each specimen was determined, and the results are summarized in the accompanying frequency distribution.

	A	B
$26- < 30$	6	4
$30- < 34$	12	9
$34- < 38$	15	19
$38- < 42$	7	10
	$m = 40$	$m = 42$

Compute a 95%CI for the difference between the true proportions of all specimens of alloys A and B that have an ultimate strength of at least 34ksi.

9.5

10. **59(57).** Obtain or compute the following quantities:

a. $F_{.05,5,8}$

b. $F_{.05,8,5}$

c. $F_{95,5,8}$

d. $F_{.95,8,5}$

e. The 99th percentile of the F distribution with $v_1 = 10, v_2 = 12$

f. The 1st percentile of the F distribution with $v_1 = 10, v_2 = 12$

g. $P(F \leq 6.16)$ for $v_1 = 6, v_2 = 4$

h. $P(.177 \leq F \leq 4.74)$ for $v_1 = 10, v_2 = 5$

11. **65(63).** The article "Enhancement of Compressive Properties of Failed Concrete Cylinders with Polymer Impregnation" (J. of Testing and Evaluation, 1977: 333-337) reports the following data on impregnated compressive modulus ($\text{psi} \times 10^6$) when two different polymers were used to repair cracks in failed concrete.

Epoxy	1.75	2.12	2.05	1.97
MMA prepolymer	1.77	1.59	1.70	1.69

Obtain a 90% CI for the ratio of variances by first using the method suggested in the text to obtain a general confidence interval formula.

Supplementary Exercises

12. **67(65).** The accompanying summary data on compression strength (lb) for $12 \times 10 \times 8$ in. boxes appeared in the article "Compression of Single-Wall Corrugated Shipping Containers Using Fixed and Floating Test Platens" (J. Testing and Evaluation, 1992: 318-320). The authors stated that "the difference between the compression strength using fixed and floating platen method was found to be small compared to normal

variation in compression strength between identical boxes." Do you agree? Is your analysis predicated on any assumptions?

Method	Sample Size	Sample Mean	Sample SD
Fixed	10	807	27
Floating	10	757	41

13. **69(67).** Is the response rate for questionnaires affected by including some sort of incentive to respond along with the questionnaire? In one experiment, 110 questionnaires with no incentive resulted in 75 being returned, whereas 98 questionnaires that included a chance to win a lottery yielded 66 responses ("Charities, No; Lotteries, No; Cash, Yes," Public Opinion Quarterly, 1996: 542-562). Does this data suggest that including an incentive increases the likelihood of a response? State and test the relevant hypotheses at significance level .10.

14. **71(69).** The article "Quantitative MRI and Electrophysiology of Preoperative Carpal Tunnel Syndrome in a Female Population" (Ergonomics, 1997: 642-649) reported that $(-473.13, 1691.9)$ was a large-sample 95% confidence interval for the difference between true average thenar muscle volume $\left(\mathrm{mm}^3\right)$ for sufferers of carpal tunnel syndrome and true average volume for nonsufferers. Calculate and interpret a 90% confidence interval for this difference.

15. **89(85).** Suppose a level .05 test of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_\alpha : \mu_1 - \mu_2 > 0$ is to be perfonned, assuming $\sigma_1 = \sigma_2 = 10$ and normality of both distributions, using equal sample sizes $(m = n)$. Evaluate the probability of a type II error when $\mu_1 - \mu_2 = 1$ and $n = 25, 100, 2500$, and 10,000 . Can you think of real problems in which the difference $\mu_1 - \mu_2 = 1$ has little practical significance? Would sample sizes of $n = 10,000$ be desirable in such problems?

16. **95(91).** Referring to Exercise 94 , develop a large-sample confidence interval formula for $\mu_1 - \mu_2$. Calculate the interval for the data given there using a confidence level of 95%.