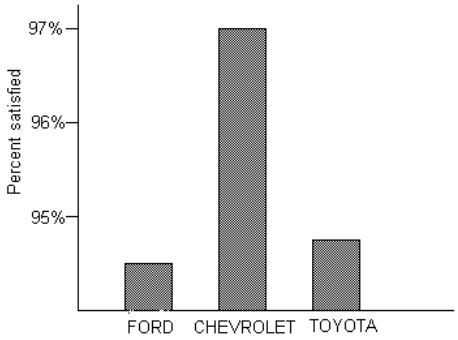
Chapter 1

1. The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380 students took the test. About how many had scores between 35 and 61?
   (a) 26
   (b) 76
   (c) 95
   (d) 190
   (e) None of these

2. The following bar graph gives the percent of owners of three brands of trucks who are satisfied with their truck.

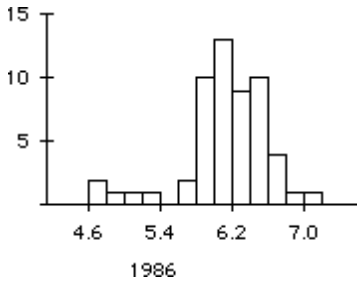


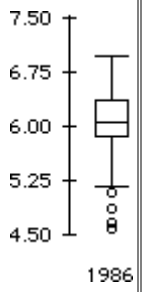
   From this graph we may legitimately conclude that
   (a) owners of other brands of trucks are less satisfied than the owners of these three brands.
   (b) Chevrolet owners are substantially more satisfied than Ford or Toyota owners.
   (c) there is very little difference in the satisfaction of owners for the three brands.
   (d) Chevrolet probably sells more trucks than Ford or Toyota.
   (e) a pie chart would have been a better choice for displaying these data.

3. A reporter wishes to portray baseball players as overpaid. Which measure of center should he report as the average salary of major league players?
   (a) The mean.
   (b) The median.
   (c) Either the mean or median. It doesn't matter since they will be equal.
   (d) Neither the mean nor median. Both will be much lower than the actual average salary.
   (e) The standard deviation should be used to show the great disparity between the astronomical salaries of the few superstars and the salaries of the rest of the players.

4. The mean salary of all female workers is $35,000. The mean salary of all male workers is $41,000. What must be true about the mean salary of all workers?
   (a) It must be $38,000.
   (b) It must be larger than the median salary.
   (c) It could be any number between $35,000 and $41,000.
   (d) It must be larger than $38,000.
   (e) It cannot be larger than $40,000.

5. Consider the following output analyzing pH values of some 1986 data on precipitation events.





Summary stats for 1986
NumNumeric = 55
NumNonNumeric = 0
NumCases = 55
Mean = 6.0673
Median = 6.1000
Std Deviation = 0.47339
Range = 2.4000
Minimum = 4.6000
Maximum = 7
75-th %ile = 6.4000

Which of the following is NOT correct?
(a) The 25th percentile is about 5.9.
(b) Some outliers appear to be present below a pH of 5.2.
(c) About 95% of the observations have pH values in the approximate range 6 ± 1.
(d) About 10% of the values are in the range 5.8 to 6.0.
(e) About 75% of the values are less than 6.4.

6. A sample of 99 distances has a mean of 24 feet and a median of 24.5 feet. Unfortunately, it has just been discovered that an observation which was erroneously recorded as "30" actually had a value of "35." If we make this correction to the data, then
(a) the mean remains the same, but the median is increased.
(b) the mean and median remain the same.
(c) the median remains the same, but the mean is increased.
(d) the mean and median are both increased.
(e) we do not know how the mean and median are affected without further calculations, but the variance is increased.

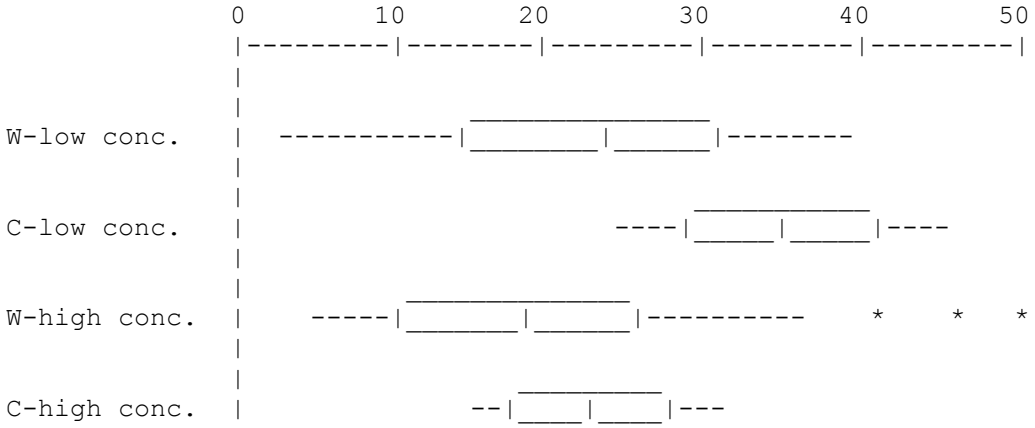
7. Forty students took a statistics examination having a maximum of 50 points. The score distribution is given in the following stem-and-leaf plot:

```
0|28
1|2245
2|01333358889
3|001356679
4|22444466788
5|000
```

The third quartile of the score distribution is equal to
(a) 43
(b) 44
(c) 45
(d) 23
(e) 32

8. Rainwater was collected in water collectors at 30 different sites near an industrial comples and the amount of acidity (pH level) was measured. The mean and standard deviation of the values are 4.60 and 1.10, respectively. When the pH meter was recalibrated back at the laboratory, it was found to be in error. The error can be corrected by adding 0.1 pH units to all of the values and then multiplying the result by 1.2. The mean and standard deviation of the corrected pH measurements are
(a) 5.64, 1.44 (b) 5.64, 1.32 (c) 5.40, 1.44 (d) 5.40, 1.32 (e) 5.64, 1.20

9. An experiment was conducted to investigate the effect of a new weed killer to suppress weed germination in onion crops. Two chemicals were used, the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations. Measurements are made, of the percent weed germination on each of 50 plots for each treatment combination. Here are some boxplots of the results:



Which of the following is NOT a feature of these data?
(a) At either high or low concentrations, the new chemical (W) gives better control of weed germination than the standard weed killer (C).
(b) Fewer weeds germinate at higher concentrations of both chemicals.
(c) The results from the standard chemical are less variable than those from the new chemical.
(d) High or low concentrations of either chemical have approximately the same effects on weed germination.
(e) Some of the results from the low concentration of weed killer W have fewer weeds germinating than some of the results from the high concentration of W.

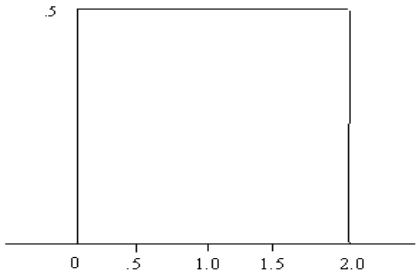
10. A clothing and textiles student is trying to assess the effect of a jacket's design on the time it takes preschool children to put the jacket on. In a pretest, she times 7 children as they put on her prototype jacket. The times (in seconds) are provided below.

n n 65 39 n 43 102

The n's represent children who had not put the jacket on after 120 seconds (in which case the children were allowed to stop). Which of the following would be the best value to use as the "typical" times required to put on the jacket?
(a) The mean time, which was 62.25 seconds.
(b) The mean time, which was 85.6 seconds.
(c) The median time, which was 54 seconds.
(d) The median time, which was 102 seconds.
(e) The missing times (the n's) mean we can't calculate any useful measures of center.

Chapter 2

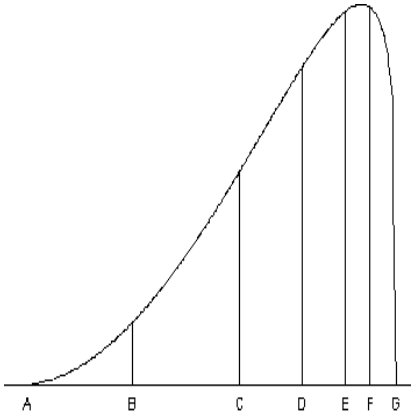
11. The density curve shown to the right takes the value 0.5 on the interval $0 \le x \le 2$ and takes the value 0 everywhere else. What percent of the observations lie between 0.5 and 1.2?
    (a) 25%
    (b) 35%
    (c) 50%
    (d) 68%
    (e) 70%



12. The proportion of observations from a standard Normal distribution that take values greater than 1.15 is about
    (a) 0.1251 (b) 0.8531 (c) 0.8729 (d) 0.8749 (e) 0.8770

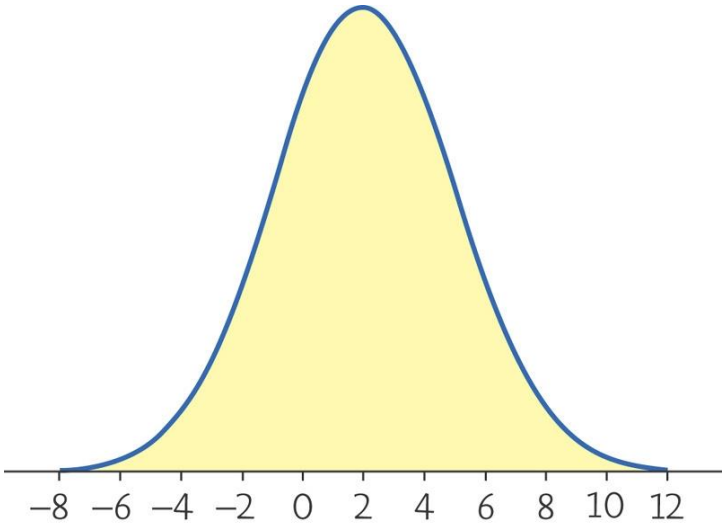
13. If the median of a set of data is equal to the mean, then
    (a) The data are Normally distributed.
    (b) The data are approximately Normally distributed.
    (c) The distribution is skewed.
    (d) The distribution is symmetric.
    (e) One can't say anything about the shape of the distribution with any certainty.

14. The figure at the right is the density curve of a distribution: Five of the seven points marked on the density curve make up the five-number summary for this distribution. Which two points are *not* part of the five-number summary?
    (a) B and E.
    (b) C and F.
    (c) C and E.
    (d) B and F.
    (e) A and G.



15. If the heights of American men follow a Normal distribution, and 99.7% have heights between 5′0″ and 7′0″, what is your estimate of the standard deviation of the heights of American men?
    (a) 1″
    (b) 3″
    (c) 4″
    (d) 6″
    (e) 12″

The figure below shows a Normal curve. Questions 16 and 17 refer to this figure.



16. The mean of this distribution is
(a) 0 (b) 1 (c) 2 (d) 3 (e) 5

17. The standard deviation of this Normal distribution is
(a) 0 (b) 1 (c) 2 (d) 3 (e) 5

18. The average yearly snowfall in Chillyville is Normally distributed with a mean of 55 inches. If the snowfall in Chillyville exceeds 60 inches in 15% of the years, what is the standard deviation?
(a) 4.83 inches (b) 5.18 inches (c) 6.04 inches (d) 8.93 inches
(e) The standard deviation cannot be computed from the given information.

Chapter 3

19. A study found correlation $r = 0.61$ between the sex of a worker and his or her income. You conclude that
(a) women earn more than men on the average.
(b) women earn less than men on average.
(c) an arithmetic mistake was made; this is not a possible value of r.
(d) this is nonsense because r makes no sense here.
(e) the correlation should have been $r = -0.61$.

20. A copy machine dealer has data on the number x of copy machines at each of 89 customer locations and the number y of service calls in a month at each location. Summary calculations give $\bar{x} = 8.4$, $s_x = 2.1$, $\bar{y} = 14.2$, $s_y = 3.8$, and $r = 0.86$. What is the slope of the least-squares regression line of number of service calls on number of copiers?
(a) 0.86
(b) 1.56
(c) 0.48
(d) None of these
(e) Can't tell from the information given

21. In the setting of the previous problem, about what percent of the variation in the number of service calls is explained by the linear relation between number of service calls and number of machines?
    (a) 86%
    (b) 93%
    (c) 74%
    (d) None of these
    (e) Can't tell from the information given

22. If data set A of (x, y) data has correlation coefficient $r = 0.65$, and a second data set B has correlation $r = -0.65$, then
    (a) the points in A exhibit a stronger linear association than B.
    (b) the points in B exhibit a stronger linear association than A.
    (c) neither A nor B has a stronger linear association.
    (d) you can't tell which data set has a stronger linear association without seeing the data or seeing the scatterplots.
    (e) a mistake has been made—r cannot be negative.

23. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least-squares fit of some data collected by a biologist gives the model $\hat{y} = 25.2 + 3.3x$, $9 < x < 25$, where x is the number of chirps per minute and $\hat{y}$ is the estimated temperature in degrees Fahrenheit. What is the estimated increase in temperature that corresponds to an increase of 5 chirps per minute?
    (a) 3.3°F (b) 16.5°F (c) 25.2°F (d) 28.5°F (e) 41.7°F

24. Which of the following relationships is most likely to result in a strong negative correlation?
    (a) The number of people showering in a college dorm and the water pressure in each shower.
    (b) The outdoor temperature and the number of fans running in non-air-conditioned dorm rooms.
    (c) The comfort rating of a mattress and the number of hours of uninterrupted sleep obtained.
    (d) The price of a home and its square footage.
    (e) The fuel efficiency of a car (miles per gallon) and its speed.

25. A set of data relates the amount of annual salary raise and the performance rating. The least squares regression equation is $\hat{y} = 1400 + 2000x$ where y is the raise amount and x is the performance rating. Which of statements (a) to (d) is *not* correct?
    (a) For each increase of one point in performance rating, the raise will increase on average by $2000.
    (b) This equation produces predicted raises with an average error of 0.
    (c) A rating of 0 will yield a predicted raise of $1400.
    (d) The correlation between salary raise and performance rating is positive.
    (e) All of the above are true.

26. Leonardo da Vinci, the renowned painter, speculated that an ideal human would have an armspan (distance from outstretched fingertip of left hand to outstretched fingertip of right hand) that was equal to his height. The following computer regression printout shows the results of a least-squares regression on height and armspan, in inches, for a sample of 18 high school students.

Dependent variable is: **Height**
No Selector
R squared = 87.1% R squared (adjusted) = 86.3%
s = 1.613 with 18 - 2 = 16 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	280.631	1	280.631	108
Residual	41.6185	16	2.60116	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	11.5474	5.6	2.06	0.0558
Armspan	0.840424	0.08091	10.4	≤ 0.0001

Which of the following statements is *false*?
(a) This least-squares regression model would make a prediction that is 1.63 inches higher than da Vinci projected for a 62-inch tall student.
(b) One of the students in the sample had a height of 70.5 inches and an armspan of 68 inches. The residual for this student is 1.83 inches.
(c) Da Vinci's projection is lower than the prediction that this least-squares line will make for any height.
(d) For every one-inch increase in armspan, the regression model predicts about a 0.84-inch increase in height.
(e) For a student 66 inches tall, our model would predict an armspan of about 67 inches.

Chapter 4

27. A variable grows exponentially over time if
(a) it increases by the addition of a fixed amount to the variable as time increases by a fixed amount.
(b) it increases by squaring its value whenever time is increased by a certain fixed amount.
(c) it increases by multiplication by a fixed amount as time increases by a fixed amount.
(d) it increases by the logarithm of its value whenever time is increased by a certain fixed amount.
(e) none of these.

A business has two types of employees, managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The number of male and female managers at each salary level and the number of male and female workers at each salary level are given in the two tables below.

	Managers	
	Male	Female
$100,000	80	20
$200,000	20	30

	Workers	
	Male	Female
$10,000	30	20
$20,000	20	80

28. The proportion of male managers who make $200,000 per year is
(a) 0.067. (b) 0.133. (c) 0.200. (d) 0.400. (e) 0.667.

29. From these data we may conclude that
(a) the mean salary of female managers is greater than that of male managers.
(b) the mean salary of males in this business is greater than the mean salary of females.
(c) the mean salary of female workers is greater than that of male workers.
(d) this is an example of Simpson's paradox.
(e) all of the above.

An article in the student newspaper of a large university had the headline "A's swapped for evaluations?" The article included the following.

> According to a new study, teachers may be more inclined to give higher grades to students, hoping to gain favor with the university administrators who grant tenure. The study examined the average grade and teaching evaluation in a large number of courses in order to investigate the effects of grade inflation on evaluations. "I am concerned with student evaluations because instruction has become a popularity contest for some teachers," said Professor Smith, who recently completed the study.

Results showed that higher grades directly corresponded to a more positive evaluation.

30. Which of the following would be a valid conclusion to draw from the study?
    (a) A teacher can improve his or her teaching evaluations by giving good grades.
    (b) A good teacher, as measured by teaching evaluations, helps students learn better, resulting in higher grades.
    (c) Teachers of courses in which the mean grade is above average apparently tend to have above-average teaching evaluations.
    (d) Teaching evaluations should be conducted before grades are awarded.
    (e) All of the above.

Chapter 5

31. A nutritionist wants to study the effect of storage time (6, 12, and 18 months) on the amount of vitamin C present in freeze dried fruit when stored for these lengths of time. Vitamin C is measured in milligrams per 100 milligrams of fruit. Six fruit packs were randomly assigned to each of the three storage times. The treatment, experimental unit, and response are respectively:
    (a) A specific storage time, amount of vitamin C, a fruit pack
    (b) A fruit pack, amount of vitamin C, a specific storage time
    (c) Random assignment, a fruit pack, amount of vitamin C
    (d) A specific storage time, a fruit pack, amount of vitamin C
    (e) A specific storage time, the nutritionist, amount of vitamin C

32. We wish to investigate if a new medicine is effective in reducing the length and severity of the flu. We take the next 20 patients who come to the walk-in clinic complaining of flu and, after a medical exam to verify that the patients do have the flu, give them the new medicine and tell them about the new drug we are giving them. One week later, the patients are contacted and 15 patients state the new remedy was helpful in reducing the severity and length of the illness. Which of the following is *not* correct?
    (a) This is a poor experiment because there is no control group. We do not know how many would feel better in a week without treatment.
    (b) This is a poor experiment because it is not double-blind. The patients may feel relief because they thought the drug should work.
    (c) This is a poor experiment because a convenience sample was selected. Patients who come to the walk-in clinic may have more severe flu than people who do not.
    (d) This is a poor experiment because we didn't give the remedy to people without the flu to assess its effect in a control group.
    (e) This is a poor experiment because the sample size is likely to be too small to detect anything but a large improvement when measuring the proportion of people reporting an improvement.

33. An experiment to measure the effect of giving growth hormones to girls affected by Turner's Syndrome was carried out recently in Vancouver. All 34 girls in the study were given the growth hormone and their heights were measured at the time the hormone was given and again one year later. No measurements were made on their final adult heights. Which of the following is *not* a problem with this experiment:
    (a) There was no blinding.
    (b) There was no control group.
    (c) Nonresponse bias
    (d) There was insufficient attention to the placebo effect.
    (e) Because final heights were not measured, it is impossible to tell if the hormone affected final height or only accelerated growth and made no difference to final height.

34. A survey is to be administered to recent nursing graduates in order to compare the starting salaries of women and men. For each graduate, three variables are to be recorded: sex, starting salary, and area of specialization.
    (a) Sex and starting salary are explanatory variables; area of specialization is a response variable.
    (b) Sex is an explanatory variable; starting salary and area of specialization are response variables.
    (c) Sex is an explanatory variable; starting salary is a response variable; area of specialization is a possible confounding variable.
    (d) Sex is a response variable; starting salary is an explanatory variable; area of specialization is a possible confounding variable.
    (e) Sex and area of specialization are response variables; starting salary is an explanatory variable.

35. A researcher wishes to compare the effects of two fertilizers on the yield of a soybean crop. She has 20 plots of land available and she decides to use a paired experiment — using 10 pairs of plots. Thus, she will
    (a) use a table of random digits to divide the 20 plots into 10 pairs and then, for each pair, flip a coin to assign the fertilizers to the 2 plots.
    (b) subjectively divide the 20 plots into 10 pairs (making the plots within a block as similar as possible) and then, for each pair, flip a coin to assign the fertilizers to the 2 plots.
    (c) use a table of random digits to divide the 20 plots into 10 pairs and then use the table of random digits a second time to decide upon the fertilizer to be applied to each pair.
    (d) flip a coin to divide the 20 plots into 10 pairs and then, for each pair, use a table of random digits to assign the fertilizers to the 2 plots.
    (e) use a table of random digits to assign the 2 fertilizers to the 20 plots and then use the table of random digits a second time to place the plots into 10 pairs.

36. A study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same sex, age, and race who did not have brain cancer, then asked about use of cell phones. This is
    (a) an observational study.
    (b) an uncontrolled experiment.
    (c) a randomized comparative experiment.
    (d) a matched pairs experiment.
    (e) a survey.

37. A class in marketing designs two videos advertising an expensive Mercedes sports car. They test the videos by asking fellow students to view both (in random order) and say which makes them more likely to buy the car. Mercedes should be reluctant to agree that the video favored in this study will sell more cars because
    (a) there is no control group.
    (b) there is no placebo.
    (c) the study used a matched pairs design instead of a completely randomized design.
    (d) this is an observational study, not an experiment.
    (e) results from students may not generalize to the older and richer customers who might buy a Mercedes.

38. Consider an experiment to investigate the effectiveness of different insecticides in controlling pests and their effects on subsequent yield. What is the best reason for randomly assigning treatment levels (spraying or not spraying) to the experimental units (farms)?
    (a) Randomization makes the experiment easier to conduct since we can apply the insecticide in any pattern rather than in a systematic fashion.
    (b) Randomization will tend to average out all other uncontrolled factors such as soil fertility so that they are not confounded with the treatment effects.
    (c) Randomization makes the analysis easier since the data can be collected and entered into the computer in any order.
    (d) Randomization is required by statistical consultants before they will help you analyze the experiment.
    (e) Randomization implies that it is not necessary to be careful during the experiment, during data collection, and during data analysis.

Chapter 6

39. An assignment of probabilities must obey which of the following?
    (a) The probability of any event must be a number between 0 and 1, inclusive.
    (b) The sum of the probabilities of all outcomes in the sample space must be exactly 1.
    (c) The probability of an event is the sum of the outcomes in the sample space that make up the event.
    (d) All of the above.
    (e) Only (a) and (b) are true.

40. Event A occurs with probability 0.2. Event B occurs with probability 0.8. If A and B are disjoint (mutually exclusive), then
    (a) P(A and B) = 0.16.
    (b) P(A or B) = 1.0.
    (c) P(A and B) = 1.0.
    (d) P(A or B) = 0.16.
    (e) both (a) and (b) are true.

41. A fair coin is tossed four times, and each time the coin lands heads up. If the coin is then tossed 1996 more times, how many heads are most likely to appear for these 1996 additional tosses?
    (a) 996
    (b) 998
    (c) 1000
    (d) 1996
    (e) None of the above. The answer is _____.

42. A die is loaded so that the number 6 comes up three times as often as any other number. What is the probability of rolling a 1 or a 6?
    (a) 1/3
    (b) 1/4
    (c) 1/2
    (d) 2/3
    (e) None of the above. The answer is _____.

Questions 43 and 44 relate to the following: In a particular game, a fair die is tossed. If the number of spots showing is either four or five, you win $1. If the number of spots showing is six, you win $4. And if the number of spots showing is one, two, or three, you win nothing. You are going to play the game twice.

43. The probability that you win $4 both times is
    (a) 1/6.
    (b) 1/3.
    (c) 1/36.
    (d) 1/4.
    (e) 1/12.

44. The probability that you win at least $1 both times is
    (a) 1/2.
    (b) 4/36.
    (c) 1/36.
    (d) 1/4.
    (e) 3/4.

Questions 45 and 46 relate to the following: An event A will occur with probability 0.5. An event B will occur with probability 0.6. The probability that both A and B will occur is 0.1.

45. The conditional probability of A, given B
    (a) is 0.5.
    (b) is 0.3.
    (c) is 0.2.
    (d) is 1/6.
    (e) cannot be determined from the information given.

46. We may conclude that
    (a) events A and B are independent.
    (b) events A and B are disjoint.
    (c) either A or B always occurs.
    (d) events A and B are complementary.
    (e) none of the above is correct.

47. Experience has shown that a certain lie detector will show a positive reading (indicates a lie) 10% of the time when a person is telling the truth and 95% of the time when a person is lying. Suppose that a random sample of 5 suspects is subjected to a lie detector test regarding a recent one-person crime. Then the probability of observing no positive reading if all suspects plead innocent and are telling the truth is
(a) 0.409.
(b) 0.735.
(c) 0.00001.
(d) 0.591.
(e) 0.99999.

48. If you buy one ticket in the Provincial Lottery, then the probability that you will win a prize is 0.11. If you buy one ticket each month for five months, what is the probability that you will win at least one prize?
(a) 0.55
(b) 0.50
(c) 0.44
(d) 0.45
(e) 0.56

Chapter 7

50. A random variable Y has the following distribution:

Y	-1	0	1	2
P(Y)	3C	2C	0.4	0.1

The value of the constant C is:
(a) 0.10.
(b) 0.15.
(c) 0.20.
(d) 0.25.
(e) 0.75.

51. A random variable X has a probability distribution as follows:

X	0	1	2	3
P(X)	2k	3k	13k	2k

Then the probability that $P(X < 2.0)$ is equal to
(a) 0.90.
(b) 0.25.
(c) 0.65.
(d) 0.15.
(e) 1.00.

52. Cans of soft drinks cost $ 0.30 in a certain vending machine. What is the expected value and variance of daily revenue (Y) from the machine, if X, the number of cans sold per day has E(X) = 125, and Var(X) = 50?
(a) E(Y) = 37.5, Var(Y) = 50
(b) E(Y) = 37.5, Var(Y) = 4.5
(c) E(Y) = 37.5, Var(Y) = 15
(d) E(Y) = 37.5, Var(Y) = 30
(e) E(Y) = 125, Var(Y) = 4.5

53. A rock concert producer has scheduled an outdoor concert. If it is warm that day, she expects to make a $20,000 profit. If it is cool that day, she expects to make a $5000 profit. If it is very cold that day, she expects to suffer a $12,000 loss. Based upon historical records, the weather office has estimated the chances of a warm day to be 0.60; the chances of a cool day to be 0.25. What is the producer's expected profit?
(a) $5000
(b) $13,000
(c) $15,050
(d) $13,250
(e) $11,450

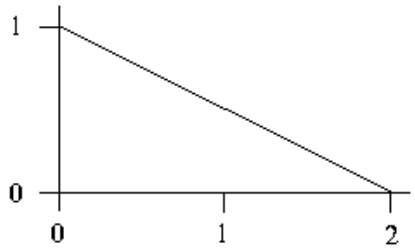
54. In a particular game, a fair die is tossed. If the number of spots showing is either 4 or 5, you win $1, if the number of spots showing is 6, you win $4, and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win. The expected value of X is
(a) $0.00.
(b) $1.00.
(c) $2.50.
(d) $4.00.
(e) $6.00.

Questions 55 and 56 use the following: Suppose X is a random variable with mean μ_X and standard deviation σ_X. Suppose Y is a random variable with mean μ_Y and standard deviation σ_Y.

55. The mean of X + Y is
(a) $\mu_X + \mu_Y$.
(b) $(\mu_X / \sigma_X) + (\mu_Y/\sigma_Y)$.
(c) $\mu_X + \mu_Y$, but only if X and Y are independent.
(d) $(\mu_X/\sigma_X) + (\mu_Y/\sigma_Y)$, but only if X and Y are independent.
(e) None of these.

56. The variance of X + Y is
(a) $\sigma_X + \sigma_Y$.
(b) $(\sigma_X)^2 + (\sigma_Y)^2$.
(c) $\sigma_X + \sigma_Y$, but only if X and Y are independent.
(d) $(\sigma_X)^2 + (\sigma_Y)^2$, but only if X and Y are independent.
(e) None of these.

57. Suppose X is a continuous random variable taking values between 0 and 2 and having the probability density function below.



$P(1 \le X \le 2)$ has value
(a) 0.50. (c) 0.25. (e) None of these

(b) 0.33. (d) 0.00.