1 – Graphs and Tables - Practice Problems

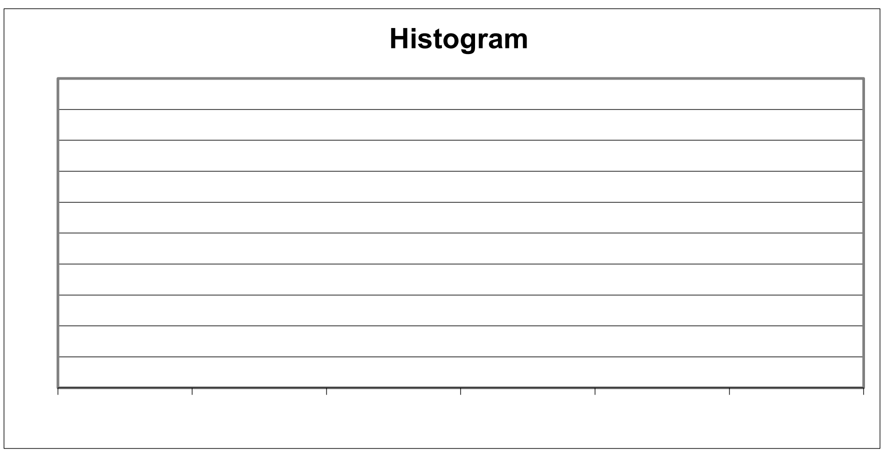
Thickness (")
0.754
0.735
0.754
0.748
0.740
0.752
0.747
0.740
0.751
0.741
0.740
0.742
0.748
0.732
0.750
0.747
0.750
0.752

1. To test new equipment for making ¾" plywood, measurements of thickness were obtained.

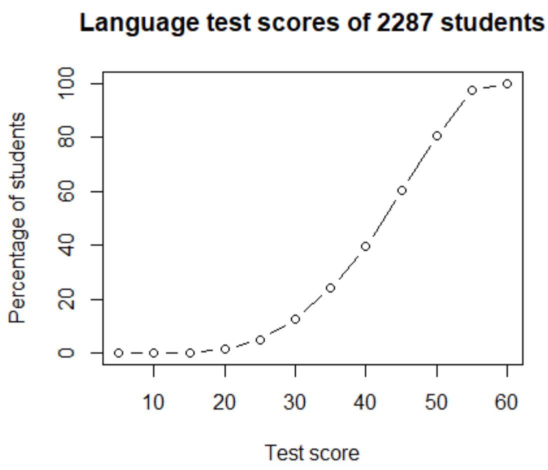
   a. Create a frequency table with five classes:

Class Limits	Frequency	Relative Frequency

   b. Sketch a histogram (or use R):



2. The test scores of 2287 Grade 8 students in the Netherlands are represented by the ogive below.



   a. What proportion of students had scores below 50?
   b. What proportion of students had scores between 30 and 40?
   c. What range of scores represents the top 25% of the class?

3. The times between Old Faithful eruptions, in minutes, are given in the stem plot below.

   The decimal point is 1 digit(s) to the right of the |

   4 | 3
   4 | 55566666777788899999
   5 | 00000111111222223333333444444444
   5 | 555555666677788889999999
   6 | 00000022223334444
   6 | 555667899
   7 | 00001111123333333444444
   7 | 555555556666666666777777777778888888888888999999999999
   8 | 000000001111111111111222222222222333333333333344444444444
   8 | 55555566666667788888999
   9 | 00000012334
   9 | 6

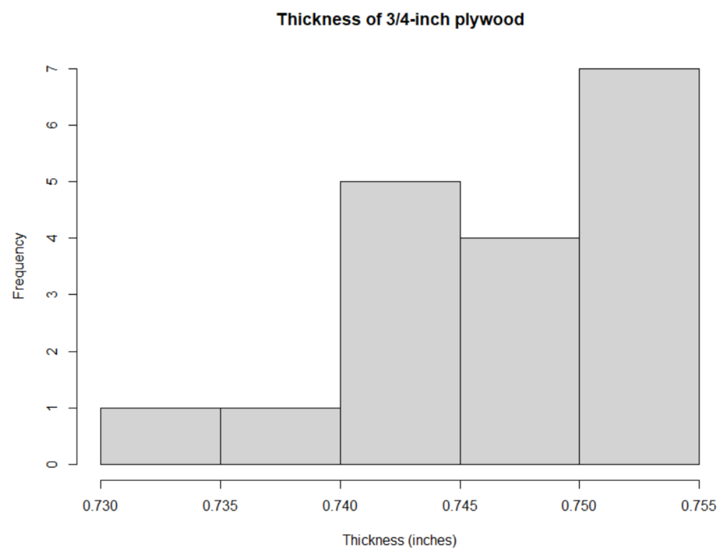
   a. What is the range of the data?
   b. Describe the distribution of the data.
   c. Which 10-minute range contains the largest number of waiting times between eruptions?

Answers

1a

Class Limits	Frequency	Rel. Freq
0.730 - 0.735	1	0.055556
0.735 - 0.740	1	0.055556
0.740 - 0.745	5	0.277778
0.745 - 0.750	4	0.222222
0.750 - 0.755	7	0.388889

1b



2. a. 80%
   b. 25%
   c. 48-60

3. a. 43-96 min
   b. bimodal, with a lot of eruptions between 50 and 55 minutes apart, and a lot between 75 and 80 minutes apart
   c. 75-85 min

2 – Numerical Statistics - Practice Problems

1. To test new equipment for making ¾" plywood, measurements of thickness were obtained.

   Calculate the range and standard deviation of the data.

Thickness (")
0.754
0.735
0.754
0.748
0.740
0.752
0.747
0.740
0.751
0.741
0.740
0.742
0.748
0.732
0.750
0.747
0.750
0.752

2. Calculate the mean and standard deviation of a sample of 350 steel rods (use R or your calculator's built in functions):

Diameter (mm)	Frequency
10.00	40
10.01	75
10.02	100
10.03	90
10.04	45

3. The following set of data is the *maximum wind velocity* measured at Vancouver International Airport over a one-year period. The data are <u>not</u> ordered by month as given.

    a. Summarize the data set using at least four common sample statistics.
    b. Imagine you are considering the possible wind effects on a new multi-storey hotel to be built in Richmond. Let's assume there is a very low probability that maximum wind velocity will be greater than 3 standard deviations above the mean.
        i. What wind velocity is 3 sample standard deviations above the sample mean?
        ii. What is the probability that maximum wind velocity will exceed this value, assuming that maximum wind velocities follow a normal distribution (Empirical Rule)?

Wind Velocity(km/h)
39.16
115.22
21.11
43.04
37.01
30.1
19.33
70.54
76.11
24.35
109.58
50.27

4. The two sets of sample data (Site A and Site B) are measurements of the drainage rate of the soil in two different building sites (liters per m^2 per s).

   a. Based on relevant sample statistics, which of the sites has the most “relatively-variable” drainage? Conclude which site you would be more likely to recommend constructing a building on.
   b. For Site B data set, assume that the drainage rates are known to be distributed in a bell-shaped curve. Could either (or both) the max and min values could be discounted as being too extreme?
   c. If you were to exclude the minimum data point from the data set for site A, what would be the effect on the mean, median and standard deviation?

Site A	Site B
2.99	1.02
4.75	3.56
8.79	3.5
5.59	3.45
2.32	4.5
1.9	13.6
	4.5
	2.3
	3.5
	2.6
	3.31
	3.1

5. The following soil density measurements were made on a strip mall development site in Richmond before and after compacting the sand.

   Make a graphical comparison and calculate summary statistics to describe the differences before and after. Imagine you are trying to describe how the compacting of the soil has improved the site for building purposes to the client. Use 3 or 4 sample statistics most suitable for making comparisons.

Before Compacting (g/cm^3)	After Compacting (g/cm^3)
0.1	15.46
1.03	17.54
2.69	18.66
3.54	19.03
4.33	19.31
13.41	20.08
14.04	21.2
14.04	21.29
14.83	21.4
53.26	24.68

6. Hank and Pete work in a fabrication facility. Records indicate that Hank completes an average of 62 items per shift, with a standard deviation of 4.2 items. Pete completes an average of 59 items per shift, with a standard deviation of 4.1 items. Which of the two would you consider to be the more consistent worker? Explain.

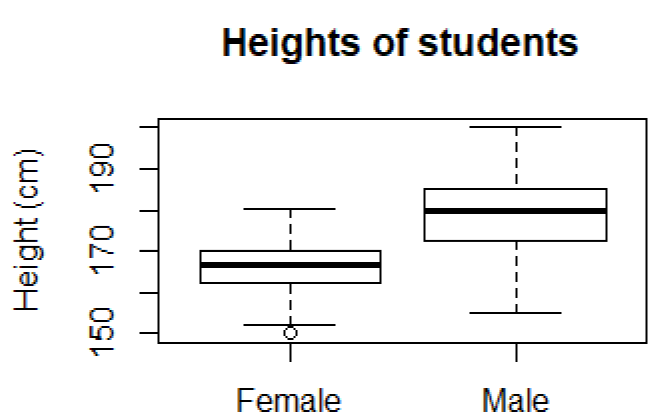
7. Suppose a lumber mill ships 4x4's with a mean moisture content of $\mu = 18\%$ and $\sigma = 0.5\%$. A histogram reveals that the data is bell-shaped.

    a. Use Chebyshev's Rule to determine the minimum percentage of scores that lie between 17% and 19%.
    b. Repeat question a. using the Empirical Rule.
    c. Use the Empirical Rule to estimate the percentage of values greater than 19.5%.
    d. Use the Empirical Rule to estimate the percentage of values between 18% and 18.5%.
    e. Building specifications require that the moisture content be under 19% to be closed in. Use the Empirical Rule to estimate the percentage of 4x4's that are suitable to be closed in.

8. Human body temperatures have a mean of 98.20°F and a standard deviation of 0.62°F. Determine whether each of the following temperatures is usual or unusual.
    a. 101.00°F
    b. 96.90°F
    c. 96.98°F

9. The heights of 237 statistics students are represented in the boxplot below. Complete the sentence below to make one observation about the relative heights of male and female students.

"x% of female students are shorter than…".



Answers

1.

Range	0.022
Standard Dev	0.006560179

2.

Stdev	0.0120078	mm

3a.
The mean maximum monthly wind velocity is 52.99 km/h with a standard deviation of 32.98 km/h. There is considerable variability seen in the standard deviation. Also, the range of values is 95.89 km/h. As a result, it is difficult to predict the maximum wind velocity during a month.

3b.
mean + 3s = 52.98 + 3(32.98) = 151.92 km/h

3c.
It is not reasonable to apply the Empirical Rule since our data does not reflect a normal distribution. If we assume that the population is normal anyway, the probability of being more than 3 standard deviation above the mean is:

$$(100 - 99.7)/2 = 0.15\%$$

4a.
Site B has a larger coefficient of variation, larger standard deviation, and larger range so it is relatively more variable than Site A. I would recommend Site A as its mean drainage is higher, and it is more consistent.

4b.
xbar - 3sigma = 4.07 - 3(3.14) = -5.3
xbar - 2sigma = 4.07 - 2(3.14) = - 2.2
xbar + 2sigma = 4.07 + 2(3.14) = 10.3
xbar + 3sigma = 4.07 + 3(3.14) = 13.5 $L/m^2/s$
The maximum value is more than 3 standard deviations above the mean and thus is a very rare value.

4c.
If we remove the minimum from Site A the mean and median should increase and the standard deviation should decrease.

5.

Before Compaction (g/cm3)		After Compaction (g/cm3)	
Mean	12.127	Mean	19.865
Median	8.87	Median	19.695
Mode	14.04	Mode	#N/A
Standard Deviation	15.63321	Standard Deviation	2.50923738
Skewness	2.347438	Skewness	0.17569909
Range	53.16	Range	9.22
Minimum	0.1	Minimum	15.46
Maximum	53.26	Maximum	24.68
Sum	121.27	Sum	198.65
Count	10	Count	10

The mean and median soil density have increased after compaction, so overall compacting the soil increases the density. More importantly, by comparing the Ranges and the Coefficients of Variation, we can see that there is much less variability in the soil density after compaction.

6.

	Mean	St. Dev.	CV	
Hank	62	4.2	6.774194	%
Pete	59	4.1	6.949153	%

Comparing the CV's we see that Hank is a more consistent worker even though his standard deviation is slightly higher than Pete's.

7a.
17% is 2σ below the mean and 19 is 2σ above the mean. Applying Chebyshev with k = 2, we have at least 75% of values between 17% and 19%

7b.
95%

7c.
19.5% is 3σ above the mean. 99.7% are within 3σ of the mean when the data is normally distributed. 0.3% of the data is outside the 3σ range. We are only interested in the upper tail of the distribution so 0.3%/2 = 0.15%

7d.
We know that 68% of the values are between -1σ and 1σ. 34% of the values should then be located between the mean and 1σ.

7e.
19% is 2σ above the mean. We want the entire bell except the part that is more than 2σ.
5%/2 = 2.5%

8a.
$Z = 4.52$; unusual

8b.
$Z = -2.10$; unusual

8c.
$Z = -1.97$; usual

9.
50% of the male students are taller than the tallest female student, who is 180 cm tall.
(Other answers are possible.)

3 – Probability - Practice Problems

1. In an engineering statistics course, the instructor awarded 30 As, 40 Bs, 35 Cs, 15 Ds and 5 Fs.
    a. What is the probability of getting an A?
    b. What is the probability of getting a C or better?

2. The probability that a reckless driver will be fined, get his license revoked or both are, respectively 0.88, 0.60 and 0.55. What is the probability that this driver will be fined or get his license revoked?

3. A carton of 24 light bulbs includes 3 that are defective. If two of the bulbs are chosen at random, what are the probabilities that
    a. neither bulb is defective
    b. exactly one of the bulbs is defective
    c. both bulbs are defective.

4. In the quality control of the production of glass blocks, we find the probability that a glass block is cracked is 0.004, that it has air bubbles is 0.006, and that it is discoloured is 0.009. We also know the appearance of cracks, air bubbles and discolouration are independent. What is the probability that an inspector will find a block that:
    a. is both cracked and has air bubbles?
    b. is either cracked or has air bubbles?
    c. is cracked, has air bubbles, and is discoloured?
    d. is cracked or has air bubbles or is discoloured ("or" being "exclusive or")?

5. A machine produces components for cellular phones. At any given time, the machine is in exactly one of three states:
    - Operational
    - Out of control
    - Down

    Experience shows that the probability that the machine will be out of control at any moment is 0.02 and that probability that the machine is down is 0.015.
    a. The machine can be used to produce a single item unless the machine is down. What is the probability that the machine can produce a single item?
    b. A repair person is called when the machine is <u>not</u> operational. What is the probability that a repair person is called?

6. A smoke detector system uses two devices, A and B. If smoke is present, the probability that it will be detected by device A is 0.95; by device B, 0.98; and by both devices, 0.94. If smoke is present, find the probability that the smoke will be detected by device A or device B.

7. The probability that a certain electronic component fails when first used is 0.10. If it does not fail immediately, the probability that it lasts for 1 year is 0.99. What is the probability that a new component will last 1 year? (i.e., The component must first not fail immediately *and* then last the year.)

8. A fire detection device uses three temperature-sensitive cells acting independently of one another in such a manner that any one or more can activate the alarm. Each cell has a probability of 0.8 of activating the alarm when the temperature reaches 45°C. Find the probability that the alarm will function when the temperature reaches 45°C.

9. Assume that a certain batch of 200 castings contains 5 defectives. Calculate the probability that
   a. a single randomly selected casting will be defective
   b. of three castings selected, all will be defective
   c. of three castings selected, exactly one will be defective
   d. of two castings selected, at least one will be defective

10. A labour management organization wants to study the problem of workers displaced by automation within the industrial engineering field. Case reports for 100 workers whose loss of job is directly attributable to technological advances are selected within the industry. For each worker selected, it is determined whether he or she was given another job within the same company, found a job with another company in industrial engineering, found a job in a new field, or has been unemployed for longer than 6 months. In addition, the union status (union or non-union) of each worker is recorded, with the results shown in the table.

Job status Union status	Same Company	New Company (same field)	New Field	Unemployed
Union	41	12	4	1
Non-Union	16	9	11	6

Suppose a worker is randomly selected from those surveyed.

   a. What is the probability that a selected worker is a union member?
   b. What is the probability that a selected worker is a union member and was given a job with the same company? What is the probability that a selected worker is a non-union member and was given a job in the same company?
   c. *If* the selected worker found a job with a new company in the same field, what is the probability that the worker is a union member?
   d. *If* the worker is not a union member, what is the probability that the worker has been unemployed for longer than 6 months?

11. **(Past Test Question)** A storage bin contains 200 injection molded parts from 4 cavities. One of the 4 cavities has a problem which causes all the parts it produces to be defective. If two of the 200 parts are randomly selected from the bin, what is the probability that:
    a. both are from the faulty cavity (i.e. both defective)
    b. only one is from the faulty cavity

12. **(Past Test Question)** Four machines, a drill, a lathe, a grinder and a miller operate independently of each other. Their utilizations currently are:

Drill 52 %	Miller 68 %
Lathe 60 %	Grinder 85 %

    a. What is the chance of all of the machines being idle? (i.e. not in use)
    b. What is the probability of either the Drill or the Lathe (or both) being utilized?
    c. What is the probability that at least two of the machines will be utilized?

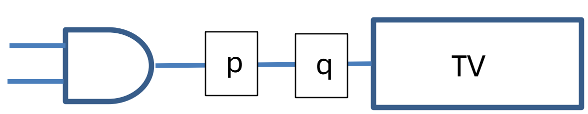
13. It is required that commercial aircraft have two independent radios. Assume that for a typical flight, the probability of radio failure is 0.002. What is the probability that both radios will fail on a particular flight?

14. When testing for current in a cable with five colour-coded wires, an electrician used a meter to test two wires at a time. How many different tests are required for every possible pairing of two wires?

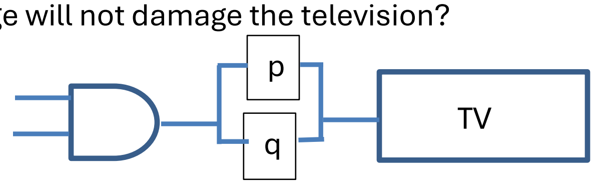
15. A lot consists of 10 articles. 6 of the articles have defects and the rest are good. Two articles are selected at random. Determine the probability that both have defects.

16. Two surge protectors *p* and *q* are used to protect an expensive television. If there is a surge in the voltage, the surge protector reduces it to a safe level. Assume that each surge protector has a 99% chance of working properly when a voltage surge occurs.

    a. If two surge protectors are arranged in series, what is the probability that a voltage surge will not damage the television?



    b. If two surge protectors are arranged in parallel, what is the probability that a voltage surge will not damage the television?



17. A manufacturer of mining safety equipment plans to ship 15 gas detectors to a customer, and as a precaution, orders that a sample of 3 of the detectors be inspected and checked. All 3 were found to be satisfactory and so the entire 15 detectors were shipped to the customer. The customer immediately put all 15 detectors to use and discovered that 2 of them were actually defective. What is the probability that the 3 detectors checked will include no defective units, when in fact 2 of the 15 detectors are defective?

18. An online service randomly generates 8-letter passwords. What is the probability that such a password consists of 8 *different* letters (assume 26 letters in alphabet)?

19. A quality control engineer inspects a random sample of 3 batteries from each lot of 24 car batteries ready to be shipped. If such a lot contains 6 batteries with slight defects, what are the probabilities that the inspector's sample will contain none of the batteries with defects?

20. An expert shot hits a target 95% of the time. What is the probability that the expert will miss the target for the first time on the 15^{th} shot?

21. Ten horses are in a race. If you randomly guess the first, second, and third place finishers, what is the probability you were correct?

22. Suppose you have not attended classes or done any homework for a course in which you are to write a ten-question multiple choice test where each question has four choices. Therefore, you have to guess on every question and have a 1/4 chance of getting each question correct.
    a. What is the probability you get none of the answers correct?
    b. What is the probability you get all of the questions correct?

23. When testing blood samples for HIV, the procedure can be made more efficient and less expensive by combining samples. If samples from three people are combined and test negative, no further testing is needed. If the combined sample tests positive, three individual tests are done. The probability of an at-risk person being HIV-positive is 0.1. What is the probability that a combined sample from three at-risk people tests positive?

24. In Mathtown, there are two brands of bicycles: Binomial Bikes and Calculus Cycles. 80% the town's cyclists ride Binomial Bikes and 20% ride Calculus Cycles. Furthermore, 5% of people who ride Binomial Bikes and 8% of people who ride Calculus Cycles bring their bikes into the local shop for service. What is the probability that a randomly-chosen bike that has been brought in for service is a Binomial Bike?

25. If a person has recently been sick with COVID-19, their blood contains antibodies that make the person immune to contracting the virus again. Serology testing can be used to determine the presence of antibodies. This has the potential to be very useful, since many people who contract COVID-19 have no symptoms.

    The testing method, of course, is not perfect. A particular serology test has a sensitivity of 90%; that is, 90% of people who have recently been sick with COVID-19 “test positive” (i.e., the test detects the presence of antibodies). The specificity of the test is 92%; that is, 92% of people who have not recently been sick with COVID-19 “test negative” (i.e., the test does not detect the presence of antibodies). It is known that 5% of people have been sick with COVID-19.

    a. Suppose a person tests positive for COVID-19 antibodies.
       What is the probability the person has recently been sick with COVID-19?
    b. Suppose a person tests negative for COVID-19 antibodies. What is the probability that the person has not recently been sick with COVID-19?

    Hint: Define events
    $A =$ person tests positive for COVID-19 antibodies
    $B =$ person’s blood actually contains COVID-19 antibodies

Answers

1a. 0.2400
1b. 0.8400
2. 0.9300
3a. 0.7609
3b. 0.2283
3c. 0.0109
4a. 0.000024
4b. 0.009976
4c. 2.16×10^{-7}
4d. 0.0188
5a. 0.985
5b. 0.035
6. 0.9900
7. 0.8910
8. 0.9920
9a. 0.025
9b. 7.61×10^{-6}
9c. 7.20×10^{-2}
9d. 0.0495
10a. 0.58
10b. 0.41 and 0.16
10c. 0.5714
10d. 0.1429
11a. 0.06156
11b. 0.3769
12a. 0.009216
12b. 0.808
12c. 0.895168
13. 0.000004
14. 10
15. 0.3333
16a. 0.9999
16b. 0.9801
17. 0.6286
18. 0.3016
19. 0.4032
20. 0.02438
21. 0.001389
22a. 0.05631
22b. 9.536E-7
23. 0.271
24. 0.7143
25a. 0.3719
25b. 0.9943

4 - Discrete Probability Distributions – Practice Problems

1. Engineers are trying to determine where to locate a mine. At one site, 70% of the ore samples are below the minimum size needed for studying. If we randomly select 4 ore samples from all the available samples at this site, what is the probability that:
    a. all four samples will be undersized?
    b. none of the samples will be undersized?
    c. Three or fewer of the samples will be undersized?

2. Based on health data, there is a 0.9989 chance that a randomly-selected 30-year-old male will live through the year and a 0.9971 that a randomly-selected 50-year-old female will live through the year.
    a. A life insurance company charges a 30-year-old male $180 for insuring his life for the year. If he dies, the policy pays out $100,000 as a benefit. What is the expected cost of the insurance policy?
    b. How much should the company charge a 50-year-old female on a $75,000 life insurance policy in order for it to have the same expected cost as the 30-year-old male's $100,000 policy?

3. Suppose you have not attended classes or done any homework for a course in which you are to write a ten-question multiple choice test where each question has four choices. You have to guess on every question and have a 1/4 chance of getting each question correct.
    a. What is the probability you get none of the answers correct?
    b. Suppose you need a mark of 8/10 or better in order to pass. What is the probability you will pass the test?

4. An airline has a policy of overbooking flights. The random variable X gives the probability that a flight cannot be boarded because there are more passengers than seats. Its probability distribution is given in the table on the right. What is the expected number of passengers who will not be boarded?

X	P(X)
0	0.067
1	0.141
2	0.274
3	0.331
4	0.187

5. A student sends job applications one at a time until he gets an interview. Each employer has a 15% chance of granting the student an interview.
    a. How many applications should the student expect to send out?
    b. What is the probability the student's fifth application is the one that results in an interview?
    c. What is the probability the student will have to send out three or more applications?

6. In one year, there were 116 homicide deaths in Richmond, Virginia. For a randomly-selected day, find the probability that the number of homicide deaths was:
    a. 0
    b. 1
    c. 2
    d. 3
    e. 4

(The actual data was as follows: 268 days (73.4% of days) with no homicides; 79 days (21.6%) with 1 homicide; 1 day with 3 homicides (0.294%); no days with more than 3 homicides.)

7. A Company produces a wood product that is shipped in lots of 20. To ensure quality 5 of every 20 products produced is examined for a minimum strength. Suppose that this lot has 4 pieces that do not meet the minimum strength requirement.
    a) What is the probability that the inspector will find exactly one part that does not meet the minimum strength requirement?

    b) What is the probability that the inspector will find at most one part that does not meet the minimum strength requirement?

    c) What is the probability that the inspector will find at least one part that does not meet the minimum strength requirement?

8. A quality control engineer inspects a random sample of 3 batteries from lots of 24 car batteries that are ready to be shipped. If such a lot contains 6 batteries with defects, what are the probabilities that the inspector's sample will contain:

    a) None of the batteries with defects.
    b) Only one of the batteries with defects
    c) At least two of the batteries with defects.

9. What is the probability that a tax auditor will catch at least 2 income tax returns with illegitimate deductions, if she randomly selects 6 returns from among the 18 returns on her desk, of which 8 actually contain illegitimate deductions?

10. A newborn baby is considered to have a low birth weight if it weighs less than 2.5kg. Dutchess County, New York has been experiencing a mean of 210 low birth weight babies born every year.
    a. Find the probability that on a given day, there is more than one baby born with low birth weight.
    b. Would it be unusual to have two babies with low birth weight born in a day?

11. If you buy a ticket on the Lotto 6/49, what are your chances of winning \$10? To win \$10 you must match 3 of the 6 (distinct) numbers drawn.

12. Radioactive atoms are unstable because they have too much energy. When they release their extra energy, they are said to decay. When studying cesium-137, a nuclear engineer found that over 365 days, 1,000,000 radioactive atoms decayed to 977,287 radioactive atoms. Find the probability that on a given day, 50 radioactive atoms decayed.

13. Consider customers arriving at a cafeteria at an average rate of 0.3 per minute.
    a. Find the probability that exactly 2 customers arrive in a 10 minute span.
    b. Find the probability that 2 or more customers arrive in a 10 minute span.
    c. Find the probability that exactly one customer arrives in a 5 minute span and one customer arrives in the next 5 minute span.

14. A lot contains 50 items, 6 of which are defective.
    a. What is the probability that a random sample of 5 items from the lot will contain no defective items?
    b. What is the probability that a random sample of 5 items will contain not more than one defective?
    c. What is the probability that a random sample of 5 items will contain more than two defectives?

15. A manufacturer of mining safety equipment plans to ship 15 gas detectors to a customer, and as a precaution, orders that a sample of 3 of the detectors be inspected and checked. All 3 were found to be satisfactory and so the entire 15 detectors were shipped to the customer. The customer immediately put all 15 detectors to use, and discovered that 2 of them were actually defective. What is the probability that the 3 detectors checked will include no defective units, when in fact 2 of the 15 detectors are defective?

Answers

1a. 0.2401
1b. 0.0081
1c. 0.7599
2a. -$70
2b. $287.50
3a. 0.05631
3b. 4.158E-4
4. 2.43
5a. 6.667
5b. 0.0783
5c. 0.7225
6a. 0.728
6b. 0.231
6c. 0.0368
6d. 0.00389
6e. 0.000309
7a. 0.4696
7b. 0.7513
7c. 0.7183
8a. 0.4032
8b. 0.4536
8c. 0.1432
9. 0.8801
10a. 0.113
10b. no ($|z|<2$)
11. 0.01765 or 1 in 56.6 chance
12. 0.0155
13a. 0.2240
13b. 0.8009
13c. 0.1120
14a. 0.5126
14b. 0.8970
14c. 0.0093
15. 0.6286

5/6 – Continuous Probability Distributions and Central Limit Theorem

1. Suppose X is a normally distributed random variable with $\mu = 16.2$ and $\sigma^2 = 1.3225$. Find the probability that
    a. X is greater than 16.8
    b. X is less than 14.9
    c. X is between 13.6 and 18.8
    d. X is between 16.5 and 16.7

2. Human body temperatures are normally distributed with a mean of 98.20°F and a standard deviation of 0.62°F. A temperature greater than 100.6°F is a fever. What proportion of people have fevers?

3. Men's heights are normally distributed with a mean of 69.0 inches and a standard deviation of 2.8 inches. Women's heights are normally distributed with a mean of 63.6 inches and a standard deviation of 2.5 inches. The Boeing 757 airplane's doors are 72 inches from top to bottom.
    a. What proportion of men can fit through the door without bending?
    b. What proportion of women can fit through the door without bending?

4. Birth weights in Norway are normally distributed with a mean of 3570g and a standard deviation of 500g.
    a. If a hospital requires special treatment for babies that weigh less than 2700g, what percentage of newborn babies require special treatment?
    b. If the hospital requires special treatment for the smallest 3% of babies, what birth weight separates babies that require special treatment from those that don't?

5. In a certain antipsychotic medication (tablet form) the mass of the active ingredient is normally distributed with mean 51 mg and standard deviation 2.5 mg.
    a. If the rated content of the active ingredient in the tablets is 50 mg, then what percentage of these tablets will have less than the rated amount of active ingredient?
    b. Suppose a patient receives 10 such tablets. What is the probability that at least one of the tablets will have less than the rated amount of the active ingredient? [*Hint: normal + binomial*]
    c. If the acceptable limits of the amount of active ingredient are 47 mg to 55 mg, then what percentage of tablets will lie outside the acceptable limits?
    d. It turns out that a simple adjustment on the machine used to manufacture the tablets allows one to change the mean content without changing the standard deviation. To what level (in mg) should the mean be raised in order that only 1% of the tablets lie below the lower acceptable limit? At this setting for the mean, what percentage of tablets lie above the upper acceptable limit?

e. With some effort it is possible to reduce the standard deviation. With the mean set at 51 mg, to what value must the standard deviation be reduced in order that only 5% of all tablets will have a mass of active ingredient which is outside the acceptable limits quoted above?

6. Specifications for a mechanical product require metal washers with an inside diameter of 0.300 ± 0.002 cm. If the inside diameters of the washers supplied by a given manufacture are normal distributed with $\mu = 0.301$ cm and $\sigma = 0.001$ cm, then what percentage of these washers will meet specifications?

7. The length of a structural component of a device is an approximately normally distributed random variable with a standard deviation of 0.90 mm. The fabricating machine can be adjusted to achieve any desired mean value. What must the mean value be so that 90% of the components have a length of 12.10 mm or greater?

8. The amount of time T that a surveillance camera will run without having to be reset is a continuous random variable that is exponentially distributed with a mean of 50 days. Find the probability that such a camera will
    a. have to be reset in less than 25 days (i.e., $T < 25$)
    b. last at least 65 days
    c. last between 60 and 80 days

9. The waiting time T between arrivals at a passport office is modelled by an exponential random variable with a rate of 0.2 per minute. Find the probability that T is:
    a. less than 5 minutes
    b. between 7 and 9 minutes
    c. more than 10 minutes

10. An air-actuated electric switch has an exponential life distribution with mean of 1000 hours.
    a. What proportion of the devices last at least 1150 hours?
    b. Find the median lifetime of the devices.

11. An integrated circuit chip has an exponential failure rate of 0.048 per thousand hours. What is the probability that it will operate satisfactorily for at least 15000 hours?

12. The elevator in a women's gym is limited to 10 passengers. Women's weights are approximately normally distributed with mean 154 lbs and standard deviation of 33 lbs.
    a. If 10 women are randomly selected, find the probability that their total weight will not exceed the maximum capacity of 1750 lb.
    b. If we want a 99.99% probability that the elevator will not be overloaded whenever 10 women are randomly selected as passengers, what should be the maximum allowable weight?

13. The new Lucky Lady Casino wants to increase revenue by providing buses that can transport gamblers from other cities. Research shows that these gamblers tend to be older, they tend to play slot machines only, and they have losses with a mean of $82 and a standard deviation of $60. The buses carry 40 gamblers per trip. The bus costs $200 to operate, and the casino gives each bus passenger $50 worth of vouchers. The casino needs to recover its costs in order to make a profit. Find the probability that if a bus is filled with 40 passengers, the casino makes a profit.

14. A gym has 210 members. The weights of members have a distribution that is approximately normal with a mean of 163 lbs and a standard deviation of 32 lbs. The design for a new building includes an elevator with a capacity limited to 12 passengers.
    a. If the elevator is designed to safely carry a load up to 2100 lbs, what is the probability that 12 randomly-selected gym members will exceed the limit?
    b. What is the maximum number of passengers that should be allowed if we want a 99.9% chance that the elevator is not overloaded when it is filled with gym members?

Answers:

1a. 0.301
1b. 0.129
1c. 0.976
1d. 0.0652
2. 0.005%
3a. 85.77%
3b. 99.96%
4a. 4.09%
4b. 2630g
5a. 0.3446
5b. 0.9854
5c. 10.96%
5d. 52.82; 19.12%
5e. 2.041 mg
6. 0.8400
7. 13.25mm
8a. 0.3935
8b. 0.2725
8c. 0.09930
9a. 0.6321
9b. 0.08130
9c. 0.1353
10a. 0.3166
10b. 693.1h
11. 0.4868
12a. 97.78%
12b. 1927 lbs
1. 99.8%
14a. 0.0918
14b. 10 passengers

7 – Confidence Intervals

For each confidence interval question, you should include the steps listed below:

- Verify relevant conditions
- Record the confidence level $1-\alpha$
- Calculate the critical value $t_{\alpha/2}$ or $Z_{\alpha/2}$
- Calculate the margin of error E
- Calculate lower and upper limits
- Write a one-sentence conclusion.

1. A sample of 75 toner cartridges produced a mean of 19300 copies and a standard deviation of 2900 copies. Construct a 95% confidence interval for the population mean number of copies produced by a toner cartridge.

2. In measuring breaking points of struts for ducts, 100 struts are measured and the average breaking force is found to be 2000 kN. The population standard deviation of the struts is believed to be 223 kN. Assume you wish 95% confidence.
   a. What is the 95% confidence interval for the mean breaking force of all beams of this type?
   b. What would your answer be if the sample size was 1000 instead of 100?

3. Suppose we wish to determine the mean starting salary of an ELEX graduate within a \$5000 margin of error with 95% confidence. How large a sample do we need? In a previous study of ELEX grads, the standard deviation of salaries was \$23000.

4. The water flow of the sprinkler system for a building is measured at 18 locations. The mean is found to be $0.12\ \mathrm{m}^3/s$ with a standard deviation of $0.045\ \mathrm{m}^3/s$. Assuming that the measurements indicated a normally distributed population, what is the interval estimate (at 99% level) for the mean flow of the sprinkler system of this site?

5. Twelve randomly-selected players were observed playing video games and the duration times of game play (in seconds) are listed below. The data suggests a normally-distributed population.

4049	3844	3859	4027	4318	4813	4657	4033	5004	4823	4334	4317

   Find a 95% confidence interval for the mean duration of game play.

6. A random sample of $n = 100$ BCIT students found that 65 were iPhone users. Find a 90% confidence interval for the population proportion p of BCIT students who use an iPhone.

Answers

1. Condition: $n \geq 30$
   Conf level: $1 - \alpha = 0.95$
   Critical t: 1.9925
   Margin of err: 667.2
   Interval: (18633, 19967)
   Sentence: We are 95% confident that the mean number of copies produced by a toner cartridge is between 18633 and 19967.

2. Part a:
   Condition: $n \geq 30$
   Conf level: $1 - \alpha = 0.95$
   Critical Z: 1.960
   Margin of err: 43.7 kN
   Interval: (1956.3 kN, 2043.7 kN)
   Sentence: We are 95% confident that the mean breaking force is between 1956.3 kN and 2043.7 kN.
   Part b: (1986.2 kN, 2013.8 kN)

3. 83

4. Condition: underlying X is normal
   Conf level: $1 - \alpha = 0.99$
   Critical t: 2.898
   Margin of err: 0.0307
   Interval: (0.089 m^3/s, $0.151\ \mathrm{m}^3/s$)
   Sentence: We are 99% confident that the mean water flow in the sprinkler system is between $0.089\ \mathrm{m}^3/s$ and $0.151\ \mathrm{m}^3/s$.

5. Condition: underlying X is normal
   Conf level: $1 - \alpha = 0.95$
   Critical t: 2.201
   Margin of err: 253.7 sec
   Interval: (4086.1 s, 4593.5 s)
   Sentence: We are 95% confident that the mean duration of game play is between 4086.1 seconds and 4593.5 seconds.

6. Condition: $n \cdot \hat{p} = 65 \geq 5$ and $n \cdot \hat{q} = 35 \geq 5$
   Conf level: $1 - \alpha = 0.9$
   Critical Z: 1.645
   Margin of err: 0.078
   Interval: $(0.572, 0.728)$
   Sentence: We are 90% confident that the proportion of BCIT students who are iPhone users is between 57.2% and 72.8%.

8 – Hypothesis Testing

Solve each problem using both the rejection region method and the p-value method.

1. Tests of older baseballs showed that when dropped 24 ft onto a concrete surface, they bounced an average of 235.8 cm. In a test of 40 new baseballs, the bounce had a mean of 233.4 cm and standard deviation of 4.5 cm. At $\alpha = 0.05$, do new baseballs bounce to a lower mean height than they used to?

2. The totals of individual weights of garbage discarded by 62 households in one week have a mean of 27.443 lbs. Assume the standard deviation is 12.458 lbs. Test the claim that the population mean weight of garbage is less than 30 lbs at a 5% significance level.

3. A random sample of 26 cans of coke has a mean volume of 12.19 oz. The sample is normally distributed and has a standard deviation of 0.11 oz. At $\alpha = 0.01$ significance level, test the claim that the cans of regular Coke have a mean volume of 12 oz.

4. Researchers collected a simple random sample of the times that 81 students took to earn their bachelor's degrees. The sample has a mean of 4.9 years and a standard deviation of 2.3 years. At $\alpha = 0.1$, do we have sufficient evidence to conclude that the average student takes more than 4.5 years to complete a bachelor's degree?

5. A developer claims that houses in a new subdivision have a mean acreage of at least 5 acres. A quick measurement of 15 lots randomly chosen yields a mean lot size of 4.95 acres and a standard deviation of 0.3 acres. At $\alpha = 0.05$, is the developer's claimed mean value supported by your test?

6. You have a manufacturer of plastic piping claiming that the bursting pressure of their product has a mean of 200 psi. You take a sample of 21 pipes and find the mean bursting pressure of the sample to be 203 psi with a standard deviation of 8 psi.
   a. If you use 95% confidence, does your sample support the manufacturer's claim of a mean value of 200 psi or do you reject their claim as being incorrect?
   b. If you made a mistake in a), what kind of mistake was it (Type I or Type II)? How could you have reduced the probability of making such a mistake?

7. The mean replacement time for a filter is listed as 40 days. A test of 6 samples from a new manufacturer's first batch yields a mean value of 38 days with a standard deviation of 2.2 days.

   a. Assuming a confidence of 90% is required, does the sample of filters indicate that the mean filter replacement time is 40 days?
   b. If you made a mistake in a), what kind of mistake was it (Type I or Type II)? How could you have reduced the probability of making such a mistake?

8. In a random sample of $n = 200$ BCIT students, you find that 85 students use an iPhone.

   a. Find a 95% confidence interval for the proportion p of BCIT students who use an iPhone.
   b. Find a 90% confidence interval for the proportion p of BCIT students who use an iPhone.
   c. Use the traditional method to test the claim (at the 5% significance level) that 50% of BCIT students use an iPhone. State your conclusion as a complete sentence with reference to the significance level.
   d. Use the p-value method to test the claim (at the 5% significance level) that less than half of BCIT students use an iPhone. State your conclusion as a complete sentence including the p-value.

9. Suppose you roll a die $n = 1000$ times and find that 210 times you get a six.

   a. Is there statistically significant evidence (at the level $\alpha = 0.05$) to conclude that the die is loaded (i.e., not fair)?
   b. Determine the 95% confidence interval for the probability p that the die turns up a six.

10. Suppose you flip a coin 100 times and obtain X heads. What is the largest possible value of X that would *not* count as statistically significant evidence that the coin:

    a. is biased towards heads?
    b. is not fair (i.e., is biased in either direction)?

Answers

1. $t_{\text{test}} = -3.373$;
   rejection region: $t < -1.685$;
   sufficient evidence that baseball bounce to a lower height than they used to

2. $t_{\text{test}} = -1.616$;
   rejection region: $t < -1.670$;
   insufficient evidence that the mean is less than 30 lbs

3. $t_{\text{test}} = 8.8$;
   rejection region: $t < -2.787$ or $t > 2.787$;
   sufficient evidence mean is not 12

4. $t_{\text{test}} = 1.565$;
   rejection region: $t > 1.292$
   sufficient evidence average student takes more than 4.5 years

5. $t_{\text{test}} = -0.645$;
   rejection region: $t < -1.761$;
   insufficient evidence that the mean lot is at least 5 acres

6. part a:
   $t_{\text{test}} = 1.718$;
   rejection region: $t < -2.086$ or $t > 2.086$
   we do no reject the claim that the mean is 200 psi

   part b:
   type II error;
   we could increase α and/or increase n

7. part a;
   $t_{\text{test}} = -2.227$;
   rejection region: $t < -2.015$ or $t > 2.015$
   the mean filter replacement time is not equal to the listed 40 days

   part b:
   type I error;
   decrease α to reduce the risk

8. part a: 95% C.I. $(0.356, 0.494)$
   part b: 90% C.I $(0.368, 0.482)$
   part c: There is sufficient evidence at the 5% significance level to reject the claim that 50% of BCIT students use an iPhone.
   part d: Less than half of students at BCIT use an iPhone (p-val = $.0169$).
   [p-val $= 0.0202$ using continuity correction in prop.test]

9. part a: Yes, the evidence is statistically significant at a 5% significance level.
   [$Z = 3.68$ (without continuity correction)]
   part b: 95% C.I. $(0.185, 0.235)$

10. part a: $X = 58$ is the largest value with p-value > 0.05 [p-val = 0.067 using prop.test]
    part b: $X = 60$ is the largest value with p-value > 0.05 [p-val = 0.057 using prop.test]

9 - Correlation and Regression

1. In assessing the efficiency (**Eff**) of a ground-loop heat pump system for heating in the winter with changing ground temperature (**Temp**) the data from the table below was gathered.

Temp (°C)	**Eff** (%)
5.6	85
4.5	87
3.4	93
1.5	95
-0.3	99
-2.3	103
-4.5	111
-8.7	210
-10.2	350
-13.5	650

   a. Calculate the linear correlation coefficient r.

   b. At $\alpha = 0.05$ significance level, is a linear model appropriate for this data? (You must conduct a formal hypothesis test, as shown in class.)

   c. Find the best fit line to the data.

   d. What efficiency would you predict for a ground temperature of 2°C?

   e. Find the 95% prediction interval for this efficiency. (Use R's predict function.)

2. Listed below are the overhead widths (in cm) of seals measured from photographs, and the weights of the seals (in kg).

Width	7.2	7.4	9.8	9.4	8.8	8.4
Weight	116	154	245	202	200	191

a. At $\alpha = 0.05$ significance level, is a linear model appropriate for this data? (You must conduct a formal hypothesis test, as shown in class.)

b. Find the best fit line to the data.

c. Find the predicted weight of a seal given that the width from an overhead photograph is 9.0 cm.

d. Find the standard error S_e.

e. Find a 95% prediction interval for the weight of a seal given that the width from an overhead photograph is 9.0 cm. (Use R's predict function.)

Answers

1. part a: $r = -0.8291$
   part b: yes; $t_{\alpha/2} = 2.306$; $t_{\text{test}} = 4.19$
   part c: $\widehat{\text{Eff}} = 132.2 - 22.90 \cdot \text{Temp}$
   part d: $S_e = 86.4\ \%$
   part e: $(-181.3\%, 354.1\%)$

2. part a : yes: $t_{\alpha/2} = 2.7765$; $t_{\text{test}} = 5.99$
   part b : $\hat{y} = -156.88 + 40.18X$
   part c : 205 kg
   part d : 15.74 kg
   part e : $157\ \text{kg} < y < 253\ \text{kg}$