Exercises

1. Consider the following population: $1, 2, 4, 8, 9$.

   (a) How many possible samples of $n = 3$ measurements can be chosen from this population?

   (b) Write out all the possible samples of $n = 3$ measurements.

2. For quality control purposes, 40 machines need to be split randomly into two groups of size 20. Explain how you could use a simple random sample to accomplish this.

3. A beverage company produces one brand of pop, in regular and diet versions. Today the company produced 10,650 cans of regular pop and 4,350 cans of diet pop. A random sample of 600 cans is required from today's production. How many of each type of pop should be included in the sample?

4. Name each sampling method described below:

   (a) Starting with a random ball bearing, every 50th ball bearing coming off the manufacturing line is selected for further inspection.

   (b) Twenty soil samples are in test tubes labelled $1, 2, \ldots, 20$. A random number generator is used to select 5 soil samples for analysis.

   (c) A manufacturing company's drill bits are divided into two sizes: 40% are long and 60% are short. A random sample of 12 long drill bits and a random sample of 18 short drill bits are selected for further inspection.

   (d) A mining company's operations are divided over 23 sites, each containing several mines. A random sample of 3 sites is selected and every mine at the selected sites is investigated.

5. [3, p. 619] A car company tested a new engine and found the following results in 20 tests of the number of litres of fuel used by a certain model for each 100 km travelled:

   $$5.3, 5.8, 5.6, 5.4, 5.9, 5.4, 6.0, 5.8, 5.8, 5.4, 6.3, 5.6, 5.7, 5.6, 5.7, 5.9, 5.5, 6.1, 5.9, 5.8$$

   (a) Make a frequency distribution table with 5 classes.

   (b) Find the relative frequencies for each of the classes in part (a).

   (c) Draw a histogram for the data in part (a).

Exercises

1. Calculate the mean and median for the following population of test scores:

$$49, 61, 67, 68, 74, 77, 79, 82, 91$$

2. Calculate the mean and median for the following sample of masses (in grams):

$$22, 25, 28, 23, 22, 27, 27, 29$$

3. A math class has four tests. A student has earned the following marks on the first three tests: 52, 69, 73. What mark does the student need on the fourth test in order to have an average of 70 on the four tests?

4. Leah's class of 29 students has a mean test score of 78. Pat's class of 36 students has a mean test score of 72. Find the mean test score if the two sets of test scores are combined into one population.

5. Calculate the mean and median for the sample of temperature readings:

Temperature (°C)	Frequency
36.8	4
37.1	6
37.2	2
37.4	5
37.7	8

6. Find the mean and median for the following sample of house prices:

Price ($)	Relative Frequency
600,000	0.1
800,000	0.45
1,000,000	0.3
1,200,000	0.1
1,400,000	0.05

7. Calculate the variance and standard deviation for the following population of test scores: $71, 76, 76, 79, 83$.

8. Calculate the variance and standard deviation for the following sample of masses (in grams): $22, 25, 27, 28$.

9. Which data set is more spread out, or are they equally spread out?

   (a) Set 1: $5, 7, 11, 13, 19$ Set 2: $15, 17, 21, 23, 29$.

   (b) Set 1: $5, 7, 11, 13, 19$ Set 2: $10, 14, 22, 26, 38$.

10. A small engineering firm has three employees with the following salaries: $35,000, $60,000 and $100,000. State what happens to the mean, median and standard deviation of the salaries in each situation below (i.e. do they increase, decrease or stay the same?)

   (a) Each employee gets a $5,000 raise.

   (b) Each employee gets a 10% raise.

   (c) The lowest salary is bumped up to $50,000.

Exercises

1. An experiment consists of flipping a coin three times. Find the probability that you will get:

   (a) at most one head.

   (b) exactly two tails.

2. Five employees have the following years of experience: $1, 3, 7, 11$ and 13. If two of these employees are randomly selected for a project, what is the probability that they have at least 15 years of experience in total?

3. An experiment consists of rolling a pair of 4-sided dice. Find the probability that the two rolls:

   (a) sum to 4.

   (b) sum to 3 or 4.

   (c) don't sum to 6.

4. You will be assigned two of four different products to analyze. Call the products A,B,C and D. Find the probability that:

   (a) you are assigned products A and C.

   (b) product B is assigned to you.

   (c) product D is not assigned to you.

5. An experiment consists of randomly selecting a number from 1 to 30 (inclusive). What is the probability that the number is divisible by 3 or 5?

6. Below is the make-up of employees at an engineering firm.

	Male	Female
Contract	37	41
Permanent	98	55

   Find the probability that an employee is:

   (a) female.

   (b) male or on contract.

   (c) female and permanent.

7. The probability that Device A fails is 2.3%. The probability that Device B fails is 3.1%. The probability that both devices fail is 0.3%. Find the probability that neither device fails.

8. Out of 62 job applicants, 35 have their P.Eng. qualification and 23 are fluent in French. Of those who are fluent in French, 17 have their P.Eng. qualification. What is the probability that an applicant has their P.Eng. but does not speak French?

9. Canadian postal codes have the following format:
letter-number-letter   number-letter-number, where numbers 0-9 are used. The letters D,F,I,O,Q,U are never used; the letters W and Z cannot be used in the first position.

   (a) How many postal codes are possible?

   (b) How many postal codes begin with A?

   (c) How many postal codes don't end with 0?

   (d) How many postal codes begin with B or end with 9?

10. Your employer's computer network requires a case-sensitive alphanumeric password that is four symbols long. Find the probability that a password:

   (a) contains at least one number.

   (b) starts or ends with d.

Exercises

1. Consider the probability distribution of a random variable X:

x	$P(x)$
9.12	0.41
10.89	0.28
12.31	0.13
14.22	0.11
15.06	0.07

   Find:

   (a) $P(10 \leq X \leq 15)$.

   (b) the mean (or expected value) of X.

   (c) the variance of X.

   (d) the standard deviation of X.

   (e) the probability that a value of X lies within two standard deviations of the mean.

2. Consider the probability distribution of a random variable Y:

y	$P(y)$
-4	0.4
-3	0.2
2	0.1
7	0.3

   Find:

   (a) $P(Y < 2)$.

   (b) the mean (or expected value) of Y.

   (c) the variance of Y.

   (d) the standard deviation of Y.

   (e) the probability that a value of Y lies within one standard deviation of the mean.

3. Project A has a 60% probability of earning \$10,000, a 30% probability of earning \$5,000 and a 10% probability of earning nothing. Project B has an 80% probability of earning \$25,000 and a 20% probability of earning nothing.

   (a) Find the expected earnings of Project A.

   (b) Find the expected earnings of Project B.

   (c) Find the variance of earnings for Project A.

   (d) Find the variance of earnings for Project B.

   (e) In terms of earnings, which project is more risky? (Which project's earnings has a larger variance?)

4. Your engineering firm is considering competing for Project Alpha. The cost of competing for Project Alpha is $10,000. The firm has a 35% probability of success, which will mean revenue of $80,000.

   (a) Find the probability distribution of the earnings from Project Alpha, where earnings=revenue−cost.

   (b) Find the expected earnings.

   (c) Find the standard deviation of the earnings.

   (d) Project Beta's earnings has a standard deviation of $25,000. In terms of earnings, which project is more risky: Project Alpha, or Project Beta?

5. You want to insure a used car worth $4,000 against theft (not damage) for one year by paying a premium m. The probability of theft during the year is 1.3%.

   (a) Find the probability distribution of the insurance company's gain.

   (b) Find the premium if the insurance company expects to gain $60.

Exercises

1. A basketball player makes 72% of his free throws. He does not improve with practice. Find the probability that in his next six free throw attempts:

   (a) he makes exactly five of them.

   (b) he makes at least four of them.

   (c) he makes at least two of them.

2. A multiple choice test has 20 questions, each of which has 3 possible answers. A student guesses randomly on each question. What is the probability that the student gets:

   (a) exactly 6 questions right?

   (b) between 5 and 7 (inclusive) questions right ?

   (c) at most 3 questions right?

3. For a particular cement mix, the average number of cracks per cubic metre of concrete is 1.7. Find the probability that a randomly-chosen cubic metre of concrete has:

   (a) at least one crack.

   (b) at most three cracks.

4. Suppose 400 typos are distributed randomly throughout a textbook that is 1000 pages long. Find the probability that a given page contains:

   (a) exactly two typos.

   (b) more than one typo.

5. A web server receives an average of 3 requests per 15-minute interval. What is the probability that the server receives at most 4 requests in the next hour?

Exercises

1. The probability density function for X is

$$f(x) = \begin{cases} x & 0 < x \leq 1 \\ \frac{1}{4} & 1 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

   Find:

   (a) $P(X = 1.5)$.

   (b) $P(0.5 < X < 1.5)$.

   (c) $P(0.5 \leq X \leq 1.5)$.

   (d) $P(X > 1.5)$.

   (e) $P(X < 0.5)$.

   (f) the mean and standard deviation of X .

2. Consider a continuous random variable, X, with probability density function

$$f(x) = \begin{cases} kx^4 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find the value of k that makes $f(x)$ a valid probability density function.

   (b) Find the mean of X.

   (c) Find the standard deviation of X.

3. Let X represent how often a student studies alone, as a proportion of their total study time. (For example, $X = 0.35$ indicates that a student spends 35% of their total study time alone.) The probability density function of X is:

$$f(x) = \begin{cases} \frac{1}{(\ln 2)(x+1)} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   Find the probability that a student studies alone:

   (a) exactly 20% of the time.

   (b) less than 20% of the time.

   (c) at least 30% of the time.

   (d) between 50% and 75% of the time.

4. The time it takes students to complete a certain project is a uniform continuous random variable with values between 2 and 11 hours. Find:

   (a) the probability density function for the completion time.

   (b) the probability that a student takes between 3 and 8 hours to complete their project.

   (c) the probability that a student takes more than 7 hours to complete their project.

   (d) the probability that a student takes less than 4 hours to complete their project.

5. The lifetime of a certain machine part (in years) has probability density function

$$f(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0 & x \le 0 \end{cases}$$

   (a) Find the probability that a part lasts less than 0.1 years.

   (b) Using part (a), find the probability that a part lasts at least 0.1 years.

6. The shelf life of a brand of tomato soup (in months) has probability density function

$$f(x) = \begin{cases} 0.1e^{-0.1x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

   (a) Find the probability that a can of soup has a shelf life between two and five months.

   (b) Find the average shelf life given that $\int_0^\infty x e^{-kx} \, dx = \frac{1}{k^2}$ for $k > 0$.

Exercises

1. Let z be the standard normal random variable. Find:

   (a) $P(0 \leq z \leq 1.20)$

   (b) $P(-2.81 \leq z \leq 0)$

   (c) $P(-1.98 \leq z \leq 3.41)$

   (d) $P(1.35 \leq z \leq 1.85)$

   (e) $P(-2.93 \leq z \leq -1.90)$

   (f) $P(z \geq 2.46)$

   (g) $P(z \leq -1.34)$

2. Find the proportion of x-values that are within three standard deviations of the mean for a normal distribution.

3. [1, p. 45] Spray drift is a constant concern for pesticide applicators and agricultural producers. The inverse relationship between droplet size and drift potential is well known. The normal distribution with $\mu = 1050$ μm and $\sigma = 150$ μm can be used as a model for droplet size for water sprayed through a 760 ml/min nozzle.

   (a) What proportion of all droplets have a size that is less than 1500 μm?

   (b) What proportion of all droplets have a size that is between 1000 and 1500 μm?

   (c) How would you characterize the smallest 2% of all droplets?

4. [1, p. 60] The bursting strength of wine bottles of a certain type is normally distributed with parameters $\mu = 250$ psi and $\sigma = 30$ psi. What proportion of these bottles have a bursting strength greater than 300 psi?

5. [1, p. 46] Based on extensive data from an urban freeway near Toronto, it is assumed that vehicle speeds can best be represented by a normal distribution. The values of $\mu = 119$ km/h and $\sigma = 13.1$ km/h were reported.

   (a) What percentage of vehicles have speeds that are between 100 and 120 km/h?

   (b) What speed characterizes the fastest 10% of all speeds?

   (c) The posted speed limit was 100 km/h. What percentage of vehicles were traveling at speeds exceeding this posted limit?

   (d) Find the value a such that 90% of all vehicle speeds are between $\mu - a$ and $\mu + a$ km/h.

6. [2, p. 158] The actual amount of instant coffee that a filling maching puts into "4-ounce" jars can be modeled as a random variable having a normal distribution with $\sigma = 0.04$ ounces. If only 2% of the jars are to contain less than 4 ounces, what should be the mean fill of these jars?

Exercises

1. At a walk-in clinic the average amount of time patients spend with a doctor is 7.1 minutes. The standard deviation is 5.2 minutes. Find the probability that for a random sample of 60 patients, the mean visit time is between 6 and 8 minutes.

2. The test scores in a large Calculus class have a mean of 71 and a standard deviation of 9. Find the probability that a random sample of n tests have an average score of less than 70 if:

    (a) $n = 40$.

    (b) $n = 250$.

3. [2, p. 213] If a 1-gallon can of paint covers an average area of 513.3 square feet with a standard deviation of 31.5 square feet, what is the probability that the sample mean area covered by a sample of 40 of these cans of paint will be

    (a) less than 510 sq ft?

    (b) more than 520 sq ft?

    (c) between 510 and 520 sq ft?

4. At a large engineering firm, employees worked a mean of 45.5 hours last week, with a standard deviation of 6 hours. One hundred and sixty employees are selected at random. What is the probability that their work hours totalled more than 7344 hours?

5. A machine is filling cans of pop. The volume per can has a standard deviation of 1.9 mL. What should the volume be set to on the machine (this is μ) in order to ensure that in a random sample of 30 cans, there is a 99% probability that the mean is at least 355.0 mL?

6. [2, p. 216] If the distribution of the weights of all passengers traveling by air on a certain route has a mean of 145 pounds with a standard deviation of 18 pounds, what is the probability that the combined weight of 40 passengers traveling on that route exceeds 6000 pounds?

Exercises

1. A random sample of 40 test marks from a large Calculus class had an average of 78.0 and a standard deviation of 5.0. Find a 95% confidence interval for the average test mark.

2. For a random sample of 50 room temperature readings taken in a lab, the mean temperature is 28.00 °C and the standard deviation is 3.32 °C. Find an upper confidence bound for the true mean temperature in the lab with:

   (a) 95% confidence.

   (b) 98% confidence.

3. At a paper factory, the paper length is known to have a standard deviation of 0.08 inches. In a random sample of 100 sheets, the mean length is found to be 11.00 inches. Find a lower confidence bound for the mean length among all sheets of paper produced at the factory with:

   (a) 90% confidence.

   (b) 99% confidence.

4. [3, p. 640] A toy manufacturer wishes to estimate the average time it takes an adult to assemble a certain "easy to assemble" toy. How many adults must be sampled so that a 99% confidence interval for the true mean assembly time μ will have a maximum margin of error of 2.0 minutes? Use the known standard deviation of 5.9 minutes for a similar model.

5. At ABC Cereal Company, the mass of cereal boxes has a standard deviation of 13 grams. We want a 95% confidence interval for the mean mass among all their cereal boxes with a margin of error less than 2 grams. Find the minimum sample size.

6. Consider a large-sample confidence interval for the population mean. Describe the effect of the following on the margin of error, assuming that the other quantities remain unchanged:

   (a) the sample size increases.

   (b) the standard deviation increases.

   (c) the confidence level increases.

   (d) the sample mean increases.

7. Ten random water samples taken from the inner harbour yield a mean nitrate ion concentration of 25.0 ppm with a standard deviation of 5.1 ppm. Assuming that the ion concentrations are normally distributed throughout the inner harbour, find a 95% confidence interval for the mean ion concentration in the inner harbour.

8. Fifteen randomly selected ropes had a mean breaking strength of 69.1 pounds, with a standard deviation of 3.5 pounds. The breaking strengths of this brand of rope are known to be normally distributed. Find a 99% confidence interval for the mean breaking strength of this brand of rope.

9. Fuel efficiencies for city driving are measured for a random sample of twelve 2012 Prius cars. The mean fuel efficiency was 51 miles per gallon, with a standard deviation of 2 miles per gallon. The fuel efficiencies of 2012 Prius cars are normally distributed. Find an upper confidence bound for the mean fuel efficiency among all 2012 Prius cars with:

   (a) 90% confidence.

   (b) 97.5% confidence.

10. At a bottling plant, volumes are measured for a random sample of 20 cans of pop. The mean volume was 356.1 mL, with a standard deviation of 1.9 mL. The volumes among all cans at the bottling plant are normally distributed. Find a lower confidence bound for the mean volume among all cans at the bottling plant with:

    (a) 95% confidence.

    (b) 99% confidence.

11. [2, p. 231] Measurement errors in scientific experiments are normally distributed. In fact, "the normal curve of errors" was first studied in the eighteenth century when scientists observed an astonishing degree of regularity in errors of measurement.

    In six determinations of the melting point of tin, a chemist obtained a mean of 232.26 °C with a standard deviation of 0.14 °C. If he uses this mean to estimate the actual melting point of tin, what is the 98% margin of error?

Exercises

1. For each data set below, use a scatterplot to decide whether the linear association is best described as positive, negative, zero, or nonlinear.

   (a)
x	10	15	25	30	40	50
y	78	71	45	41	20	3

   (b)
x	1	5	7	11	13	14
y	33	18	2	5	19	24

   (c)
x	1.1	1.4	1.6	1.6	1.8	1.9
y	1.3	1.7	2.1	2.3	2.6	2.9

   (d)
x	5	10	15	20	25	30	35
y	7	9	6	8	7	9	6

2. Consider a bivariate data set. What percentage of the variation in y is accounted for by the best-fit line if the correlation coefficient is:

   (a) 0.8?

   (b) −0.9?

3. Bivariate Data Set A has a correlation coefficient of 0.84. Bivariate Data Set B has a correlation coefficient of −0.96. Which data set has a stronger linear association? Explain.

4. [3, p. 650] In a research project to determine the amount of a drug that remains in the bloodstream after a given dosage, the concentrations y (in mg of drug / dL of blood) were recorded after t hours, as shown:

t	y
1.0	7.6
2.0	7.2
4.0	6.1
8.0	3.8
10.0	2.9
12.0	2.0
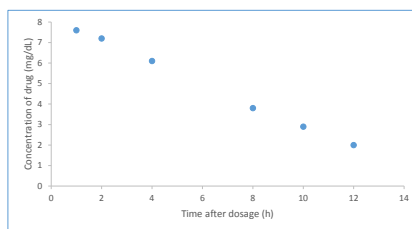
   (a) Find the equation of the best-fit line.

   (b) Find the residual for the point $(4.0, 6.1)$.

   (c) Find the correlation coefficient and the coefficient of determination.

   (d) What percentage of the variation in y is accounted for by the best-fit line?

   (e) What drug concentration does the best-fit line predict at 6 hours?

   (f) Why should we not use the data to predict the drug concentration 24 hours after the dosage is given?

   (g) According to the best-fit line, what value of t corresponds to $y = 5.0$ mg/dL?

5. [2, p. 341] The following are measurements of the air velocity and evaporation coefficient of burning fuel droplets in an impulse engine:

$$x = \text{air velocity (cm/s)}$$
$$y = \text{evaporation coefficient (mm}^2\text{/s)}$$

x	y
20	0.18
60	0.37
100	0.35
140	0.78
180	0.56
220	0.75
260	1.18
300	1.36
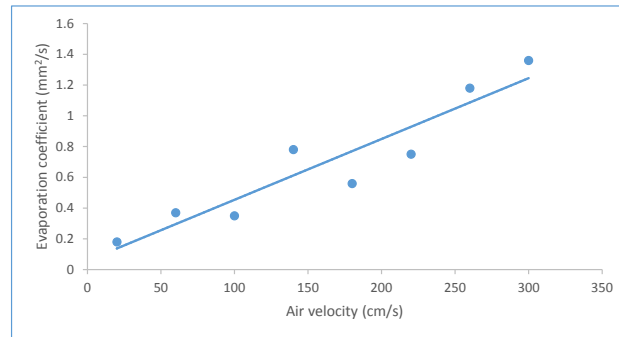
The equation of the best-fit line is $\hat{y} = 0.058 + 0.004t$ and the coefficient of determination is 0.8793.

   (a) Is the linear association positive or negative?

   (b) Find the correlation coefficient.

   (c) What percentage of the variation in y is accounted for by the best-fit line?

   (d) Interpret the slope of $\hat{y}$.

Appendix A   Histograms in Excel 2013

To create a histogram for a data set like

$$90, 93, 94, 95, 98, 100, 101, 102, 103, 105, 108, 108, 109, 110, 110, 114, 115, 115, 119, 120$$

Step 1 Set up Excel for Data Analysis

FILE $\rightarrow$ Options $\rightarrow$ Add-Ins

At the bottom of the window, under *Manage:* select Excel Add-ins and click Go...

A new window will open. Select Analysis ToolPak and click OK

Step 2 Enter your data points in column A

Step 3 Enter your class endpoints in column B

Excel calls classes "bins" and uses the right endpoint of each class as the class mark. Your classes will be $(-\infty, B1]$, $(B1, B2]$, $(B2, B3]$, ...

For the given example, a good choice of classes is $(-\infty, 90]$, $(90, 95]$, $(95, 100]$, $\cdots$, $(115, 120]$ so column B should consist of the list $90, 95, 100, 105, 110, 115, 120$

Step 4 Create the histogram

DATA $\rightarrow$ Data Analysis $\rightarrow$ Histogram $\rightarrow$ OK

A new window will open.

- For *Input Range:* enter the cell range of your data set. For this example it is A1:A20
- For *Bin Range:* enter the cell range of your class endpoints. For this example it is B1:B7
- Select New Workbook
- Select Chart Output
- Click OK

A new Excel workbook will open, displaying the frequency distribution table and the histogram. Click on the histogram to activate a menu for editing options.

Note The *More* class includes any data points greater than your largest class endpoint.

Appendix B   Statistics on the SHARP EL-531X

To find the **mean** and **standard deviation** of a list of numbers like

$$5, 2, 6, 4, 7$$

Step 1 Set up the calculator

$\boxed{\text{MODE}} \rightarrow \boxed{1}$ *to select STAT* $\rightarrow \boxed{0}$ *to select SD*

Step 2 Enter your data points

For the given example, this looks like:

$\boxed{5} \rightarrow \boxed{\text{M+}} \rightarrow \boxed{2} \rightarrow \boxed{\text{M+}} \rightarrow \boxed{6} \rightarrow \boxed{\text{M+}} \rightarrow \boxed{4} \rightarrow \boxed{\text{M+}} \rightarrow \boxed{7} \rightarrow \boxed{\text{M+}}$

At the end of this sequence, the screen should read *DATA SET = 5* to indicate that you've entered 5 data points.

Step 3 Calculate statistics on your data set

You do not need to re-enter the data set between calculations.

- For population mean, μ, or sample mean, $\bar{x}$:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{4}$ *to select $\bar{x}$*
  For the given example, your answer should be 4.8

- For sample standard deviation, s:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{5}$ *to select sx*
  For the given example, your answer should be $1.9234\cdots$

- For population standard deviation, σ:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{6}$ *to select σx*
  For the given example, your answer should be $1.7204\cdots$

Step 4 Clear the data set

$\boxed{\text{2nd F}} \rightarrow \boxed{\text{ALPHA}} \rightarrow \boxed{0}$ *to select MEM* $\rightarrow \boxed{0}$ *to select CLR_MEMORY?*

To find the **mean** and **standard deviation** of a probability distribution like

x	$P(x)$
0	0.1
1	0.5
2	0.4

Step 1 Set up the calculator

$\boxed{\text{MODE}} \rightarrow \boxed{1}$ *to select STAT* $\rightarrow \boxed{0}$ *to select SD*

Step 2 Enter your data points

For the given example, this looks like:

$\boxed{0} \rightarrow \boxed{\text{STO}}$ to select $(x, y) \rightarrow \boxed{\cdot} \rightarrow \boxed{1} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{1} \rightarrow \boxed{\text{STO}}$ to select $(x, y) \rightarrow \boxed{\cdot} \rightarrow \boxed{5} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{2} \rightarrow \boxed{\text{STO}}$ to select $(x, y) \rightarrow \boxed{\cdot} \rightarrow \boxed{4} \rightarrow \boxed{\text{M+}}$

At the end of this sequence, the screen should read *DATA SET = 3* to indicate that you've entered 3 data points.

Step 3 Calculate statistics on your probability distribution

You do not need to re-enter the data set between calculations.

- For mean, μ, aka expected value:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{4}$ *to select $\bar{x}$*
  For the given example, your answer should be 1.3

- For standard deviation, σ:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{6}$ *to select σx*
  For the given example, your answer should be $0.64031\cdots$

Step 4 Clear the data set

$\boxed{\text{2nd F}} \rightarrow \boxed{\text{ALPHA}} \rightarrow \boxed{0}$ *to select MEM* $\rightarrow \boxed{0}$ *to select CLR_MEMORY?*

Note: This process can also be used to find the mean and SD of a data set given as a frequency distribution table.

To find the **linear regression line** and **correlation coefficient** of a bivariate data set like

x	y
1	3
2	6
3	6
4	8
5	12

Step 1 Set up the calculator

$\boxed{\text{MODE}} \rightarrow \boxed{1}$ *to select STAT* $\rightarrow \boxed{1}$ *to select LINE*

Step 2 Enter your data points

For the given example, this looks like:

$\boxed{1} \rightarrow \boxed{\text{STO}}$ *to select (x,y)* $\rightarrow \boxed{3} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{2} \rightarrow \boxed{\text{STO}}$ *to select (x,y)* $\rightarrow \boxed{6} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{3} \rightarrow \boxed{\text{STO}}$ *to select (x,y)* $\rightarrow \boxed{6} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{4} \rightarrow \boxed{\text{STO}}$ *to select (x,y)* $\rightarrow \boxed{8} \rightarrow \boxed{\text{M+}} \rightarrow$

$\boxed{5} \rightarrow \boxed{\text{STO}}$ *to select (x,y)* $\rightarrow \boxed{1} \rightarrow \boxed{2} \rightarrow \boxed{\text{M+}}$

At the end of this sequence, the screen should read *DATA SET = 5* to indicate that you've entered 5 data points.

Step 3 Calculate statistics on your data set

You do not need to re-enter the data set between calculations.

- For the linear regression equation, $\hat{y} = a + bx$:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{(}$ *to select a* $\rightarrow \boxed{=}$
  For the given example, your answer should be 1
  $\boxed{\text{ALPHA}} \rightarrow \boxed{)}$ *to select b* $\rightarrow \boxed{=}$
  For the given example, your answer should be 2
  Together, the answer for this example is $\hat{y} = 1 + 2x$.

- For the correlation coefficient, r:
  $\boxed{\text{ALPHA}} \rightarrow \boxed{\div}$ *to select r* $\rightarrow \boxed{=}$
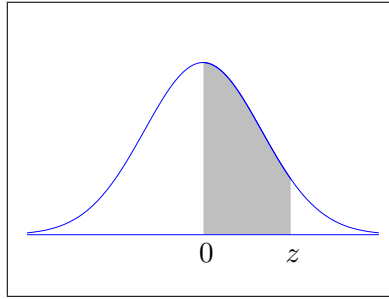 For the given example, your answer should be $0.9534\cdots$

- You can also calculate quantities like $\bar{x}$, $\bar{y}$, $\sum x$, $\sum x^2$, $\sum xy, \cdots$

Step 4 Clear the data set

$\boxed{\text{2nd F}} \rightarrow \boxed{\text{ALPHA}} \rightarrow \boxed{0}$ *to select MEM* $\rightarrow \boxed{0}$ *to select CLR_MEMORY?*

Standard Normal Distribution Table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998

t-Distribution Table
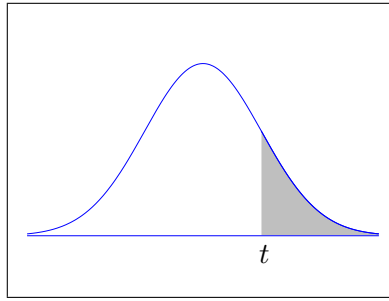
The shaded area is equal to α for $t = t_\alpha$.

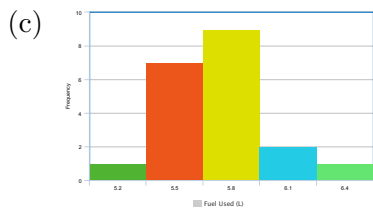| df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

Answers to Exercises

1 Collection and Representation of Data

1. (a) 5C3=10

   (b) {1,2,4}, {1,2,8}, {1,2,9}, {1,4,8}, {1,4,9},{1,8,9}, {2,4,8}, {2,4,9}, {2,8,9}, {4,8,9}

2. Select a simple random sample of 20 machines. The selected machines will be Group 1; the unselected machines will be Group 2.

3. 426 cans of regular pop and 174 cans of diet pop

4. (a) 1-in-50 systematic sample

   (b) simple random sample

   (c) stratified random sample

   (d) cluster sample

5. (a)

fuel used (L)	frequency
5.1-5.3	1
5.4-5.6	7
5.7-5.9	9
6.0-6.2	2
6.3-6.5	1

   (b)

fuel used (L)	relative frequency
5.1-5.3	$1/20 = 0.05$
5.4-5.6	$7/20 = 0.35$
5.7-5.9	$9/20 = 0.45$
6.0-6.2	$2/20 = 0.1$
6.3-6.5	$1/20 = 0.05$

   (c)

2 Summarizing Data

1. $\mu = 72$; median $= 74$

2. $\bar{x} = 25.4$ g ; median $= 26$ g

3. 86

4. $\mu = 74.7$

5. $\bar{x} = 37.31\,°C$; median $= 37.4\,°C$

6. $\bar{x} = \$910,000$; median $= \$800,000$

7. $\sigma^2 = 15.6$; $\sigma = 3.9$

8. $s^2 = 7$ g^2; $s = 2.6$ g

9. (a) Data sets are equally spread out.

   (b) Set 2 is more spread out.

10. (a) Mean and median both increase (by $\$5,000$); SD stays the same.

    (b) Mean, median and SD all increase (by 10%)

    (c) Mean increases. Median stays the same. SD decreases.

3 Probability

1. (a) $\frac{4}{8}$

   (b) $\frac{3}{8}$

2. $\frac{4}{10}$

3. (a) $\frac{3}{16}$

   (b) $\frac{5}{16}$

   (c) $\frac{13}{16}$

4. (a) $\frac{1}{6}$

   (b) $\frac{3}{6}$

   (c) $\frac{3}{6}$

5. $\frac{14}{30}$

6. (a) $\frac{96}{231}$

   (b) $\frac{176}{231}$

   (c) $\frac{55}{231}$

7. 0.949 or 94.9%

8. $\frac{18}{62}$

9. (a) 7.2 million

   (b) 400,000

(c) 6.48 million

(d) 1.08 million

10. (a) 0.505

(b) 0.032

4 Discrete Random Variables

1. (a) 0.52

(b) $\mu = E(X) = 11.007$

(c) $\sigma^2 = 3.972$

(d) $\sigma = 1.993$

(e) 0.93

2. (a) 0.6

(b) $\mu = E(Y) = 0.1$

(c) $\sigma^2 = 23.3$

(d) $\sigma = 4.83$

(e) 0.7

3. (a) \$7,500

(b) \$20,000

(c) 11,250,000 \$^{2}

(d) 100,000,000 \$^{2}

(e) Project B

4. (a)

x	$P(x)$
70,000	0.35
-10,000	0.65

(b) \$18,000

(c) \$38,000

(d) Project Alpha

5. (a)

x	$P(x)$
$m - 4000$	0.013
m	0.987

(b) \$112

5 Binomial and Poisson Distributions

1. (a) 0.33

(b) 0.78

(c) 0.99

2. (a) 0.18

(b) 0.51

(c) 0.06

3. (a) 0.82

(b) 0.91

4. (a) 0.05

(b) 0.06

5. 0.008

6 Continuous Random Variables

1. (a) 0

(b) 0.5

(c) 0.5

(d) 0.375

(e) 0.125

(f) $\mu = \frac{4}{3}$, $\sigma = 0.80$

2. (a) $k = \frac{5}{32}$

(b) $\frac{5}{3}$

(c) 0.28

3. (a) 0

(b) 0.26

(c) 0.62

(d) 0.22

4. (a) $f(x) = \begin{cases} \frac{1}{9} & 2 < x < 11 \\ 0 & \text{otherwise} \end{cases}$

(b) $\frac{5}{9}$

(c) $\frac{4}{9}$

(d) $\frac{2}{9}$

5. (a) 0.18

(b) 0.82

6. (a) 0.21

(b) 10 months

7 The Normal Distribution

1. (a) 0.3849

(b) 0.4975

(c) 0.9758

(d) 0.0563

(e) 0.0270

(f) 0.0069

(g) 0.0901

2. 0.9974

3. (a) 0.9987

   (b) 0.6280

   (c) The smallest 2% are under 742.5 μm

4. 0.0475

5. (a) 45.84%

   (b) Fastest 10% are above 135.8 km/h

   (c) 92.65%

   (d) $a = 21.484$

6. 4.082 ounces

8 Central Limit Theorem

1. 0.8594

2. (a) 0.2420

   (b) 0.0392

3. (a) 0.2546

   (b) 0.0885

   (c) 0.6569

4. 0.2005

5. 355.8 mL

6. 0.0392

9 Confidence Intervals

1. $76.5 < \mu < 79.5$

2. (a) $\mu < 28.77\,°C$

   (b) $\mu < 28.96\,°C$

3. (a) $\mu > 10.99$ inches

   (b) $\mu > 10.98$ inches

4. 58

5. 163

6. (a) margin of error decreases

   (b) margin of error increases

   (c) margin of error increases

   (d) margin of error doesn't change

7. $21.4 < \mu < 28.6$ ppm

8. $66.4 < \mu < 71.8$ pounds

9. (a) $\mu < 52$ miles per gallon

   (b) $\mu < 52$ miles per gallon

10. (a) $\mu > 355.4$ mL

    (b) $\mu > 355.0$ mL

11. $0.19\,°C$

10 Linear Regression

1. (a) negative

   (b) nonlinear

   (c) positive

   (d) zero

2. (a) 64%

   (b) 81%

3. Data Set B because $|r_B| > |r_A|$

4. (a) $\hat{y} = 8.16 - 0.52t$

   (b) 0.02

   (c) $r = -0.999$, $r^2 = 0.998$

   (d) 99.8%

   (e) 5.04 mg/dL

   (f) 24 is outside the data set range $1.0 \le t \le 12.0$

   (g) 6.1 h

5. (a) positive

   (b) 0.9377

   (c) 87.93%

   (d) As the air velocity increases by 1 cm/s, the evaporation coefficient increases on average by 0.004 mm^2/s

References

[1] Devore, J., Farnum, N., & Doi, J. (2014). *Applied Statistics For Engineers and Scientists* (3rd ed.). Stamford, CT: Cengage Learning.

[2] Johnson, R. (2005) *Miller and Freund's Probability and Statistics for Engineers* (7th ed.) Upper Saddle River, NJ: Pearson Prentice Hall.

[3] Washington, A. & Boue, M. (2010) *Basic Technical Mathematics with Calculus: SI Version* (10th ed.) Toronto: Pearson.