Probability Distributions for Continuous Variables Review

Learning Goals

Section	After this section, I can
7.1	• distinguish between discrete variables and continuous variables • work with sample values for situations that can take on continuous values • represent a probability distribution using a mathematical model • represent a sample of values of a continuous random variable using a frequency table, a frequency histogram, and a frequency polygon
7.2	• determine the theoretical probability for a continuous random variable over a range of values • determine the mean and standard deviation of a sample of values • calculate and explain the meaning of a *z*-score • solve real-world probability problems involving normal distributions
7.3	• recognize the general characteristics of a normal distribution • use technology to simulate a normal distribution in order to investigate its properties • determine probabilities for a normal distribution
7.4	• distinguish among the meanings of common confidence levels such as 90%, 95%, and 99% • determine the margin of error for a population mean estimated using a sample • determine the upper and lower limits of the confidence interval
7.5	• make connections between a normal distribution and a binomial distribution • make connections between a normal distribution and a hypergeometric distribution • recognize the role of the number of trials in these connections

MHR Page 372 #s 1 - 7 & 9 -12

Solutions

1. Advanced scuba divers sometimes breathe enriched air called nitrox. Nitrox contains 32% or 36% oxygen rather than the usual 21% found in ordinary air. Nitrox is toxic to humans if breathed for too long. Before using a nitrox-filled tank, the diver must verify and record the actual percent of oxygen in the tank. The table shows the oxygen content of a sample of 15 tanks.

a) Can you determine from the table whether the distribution is uniform? Explain your answer.

b) Devise a plan to determine whether the distribution is uniform. Carry out your plan, and draw a conclusion.

Percent Oxygen				
32.1	32.2	31.8	32.2	32.1
32.0	31.8	32.1	32.0	31.9
32.0	31.9	31.9	32.2	31.8

a) The raw data is difficult to analyse in this form, so we can't really determine if the distribution is uniform.

b) To help analyse the data we can put it into a tally/frequency table. By looking at the table we can see that the distribution is uniform, because each percentage of oxygen has the same frequency.

Percent Oxygen	Frequency
31.7–31.8	3
31.8–31.9	3
31.9–32.0	3
32.0–32.1	3
32.1–32.2	3

2. A sporting goods company produces custom wooden arrows. Any arrows with a length less than 69.2 cm or greater than 73.0 cm are rejected. The remaining arrows follow an approximately uniform distribution.

a) What is the height of the probability distribution?

b) What is the probability that an arrow has a length less than 71.1 cm?

c) What is the probability that an arrow has a length between 70.6 cm and 71.6 cm?

a) The area under the probability distribution is always equal to one.

base x height = Area

3.8h = 1

h = 1 / 3.8

h = 0.263

base = 73.0 - 69.2

base = 3.8

The height of the probability distribution is about 0.263

b) The area under the graph for a given region is equal to the probability of it occurring: $P(69.2 < x < 71.1)$ = area

Area = base x height

Area = 1.9 x 0.263

Area = 0.4997

base = 71.1 - 69.2

base = 1.9

The probability of getting an arrow less than 71.1 cm is 0.4997

c) The area under the graph for a given region is equal to the probability of it occurring: $P(70.6 < x < 71.6)$ = area

Area = base x height

Area = 1.0 x 0.263

Area = 0.263

base = 71.6 - 70.6

base = 1.0

The probability of getting an arrow between 70.6 cm and 71.6 cm is 0.263

3. Ryan is raising fruit flies as a science project. The table shows the frequencies of the lifetimes of the flies.

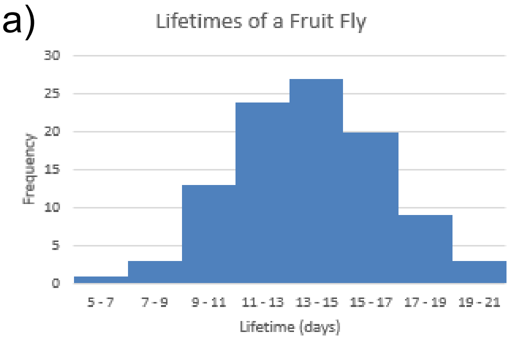
a) Sketch a frequency histogram for these data.

b) Sketch a frequency polygon for these data.

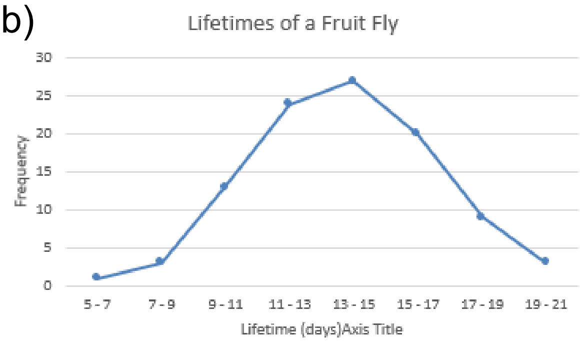
c) Estimate the mean life of the fruit flies.

Lifetime (days)	Midpoint	Frequency	Mid x Freq
5 - 7	6	1	6
7 - 9	8	3	24
9 - 11	10	13	130
11 - 13	12	24	288
13 - 15	14	27	378
15 - 17	16	20	320
17 - 19	18	9	162
19 - 21	20	3	60

a)



b)



c) Mean life = Total of (Mid x Freq) ÷ Total Frequency

= 1368 ÷ 100

= 13.68

The mean life is about 14 days.

4. Refer to the fruit fly data in #3.

a) Add a column of relative frequencies to the table.

b) What is the probability that a given fruit fly will die before the end of the first week?

c) What is the probability that a fruit fly will live from 11 to 17 days?

Lifetime (days)	Frequency	Relative Freq
5 - 7	1	0.010
7 - 9	3	0.030
9 - 11	13	0.130
11 - 13	24	0.240
13 - 15	27	0.270
15 - 17	20	0.200
17 - 19	9	0.090
19 - 21	3	0.030

b) $P(x < 7) = 0.01$

The probability that a given fruit fly will die before the end of the first week is 0.01

c) $P(11 < x < 17) = P(11 < x < 13) + P(13 < x < 15) + P(15 < x < 17)$

$= 0.240 + 0.270 + 0.200$

$= 0.71$

The probability that a fruit fly will live from 11 to 17 days is 0.71

5. Triple Q Farms grows soybeans. A farmer is testing a new strain of plant. After 3 months, 28 seeds produced plants with heights as shown.

a) Determine the mean height.

b) Determine the standard deviation of the heights.

c) Sketch a frequency histogram for these data.

d) Do the heights appear to be normally distributed? Explain.

Soybean Height (cm)	x^2
25.1	630.0
39.7	1576.1
44.2	1953.6
32.0	1024.0
37.6	1413.8
38.7	1497.7
42.9	1840.4
48.8	2381.4
40.2	1616.0
49.5	2450.3
36.8	1354.2
34.4	1183.4
38.8	1505.4
34.9	1218.0
41.0	1681.0
41.1	1689.2
36.9	1361.6
36.1	1303.2
44.6	1989.2
32.0	1024.0
43.3	1874.9
47.4	2246.8
41.6	1730.6
30.0	900.0
37.1	1376.4
45.8	2097.6
44.0	1936.0
33.8	1142.4

a) Mean = Total of heights ÷ Number of plants

= 1098.3 ÷ 28

= 39.225

The mean height is 39.225 cm

b) Standard deviation $s = \sqrt{\dfrac{\sum x^2 - n(\overline{x})^2}{n-1}}$

$\sum x^2$ = 43997.27   n = 28 $n(\overline{x})^2$ = 28(39.225^2)

= 43080.8175

$$s = \sqrt{\frac{43997.27 - 43080.8175}{28 - 1}}$$

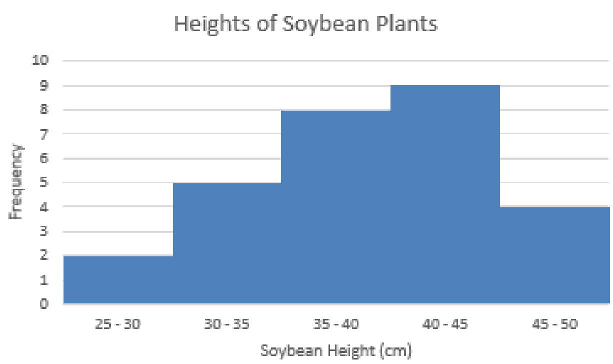
s = 5.8260   The standard deviation is about 5.8260

c) Create a grouped frequency table

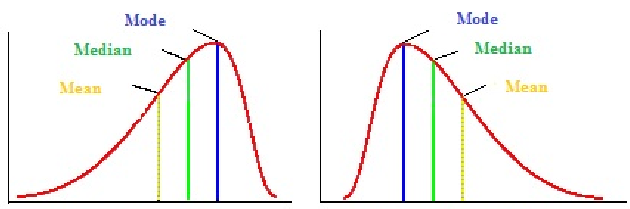
Lowest value = 25.1 cm

Highest value = 49.5 cm

Soybean Height (cm)	Frequency
25 - 30	2
30 - 35	5
35 - 40	8
40 - 45	9
45 - 50	4



d) The distribution certainly has a bell shaped look to it, however the data does appear to be more negatively skewed when compared to a true normal distribution.



Left-Skewed (Negative Skewness)   Right-Skewed (Positive Skewness)

6. Current engineering graduates earn a mean starting salary of $62 000 in Canada, with a standard deviation of $2500. Assuming that the salaries are normally distributed, what is the probability that a graduate will find a job with a starting salary of more than $65 000?

Find the z-score for $65,000 and then look up its associated probability in the table.

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869

x = 65,000 $\bar{x}$ = 62,000 s = 2500

$$z = \frac{x - \bar{x}}{s}$$

z = (65,000 - 62,000) / 2500

z = 1.2 ⟶ P(x < 65,000) = 0.8849

However, we want P(x > 65,000)

P(x > 65,000) = 1 - P(x < 65,000)

= 1 - 0.8849

= 0.1151

The probability of a graduate getting a job with a starting salary greater than $65,000 is 0.1151

7. A police radar unit is set up to monitor vehicles on a stretch of highway with a speed limit of 100 km/h. Long-term records for this location show that speeds vary normally with a mean of 105 km/h and a standard deviation of 7 km/h. Drivers who exceed the speed limit by 20 km/h accumulate demerit points as well as pay a fine.

a) What percent of the drivers will accumulate demerit points?

b) What is the probability that a given vehicle has a speed between 99 km/h and 101 km/h?

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495
1.7	0.9554	0.9564	0.9573	0.9582	0.9591
1.8	0.9641	0.9649	0.9656	0.9664	0.9671
1.9	0.9713	0.9719	0.9726	0.9732	0.9738
2.0	0.9772	0.9778	0.9783	0.9788	0.9793
2.1	0.9821	0.9826	0.9830	0.9834	0.9838

a) Find the z-score for 120 km/h and then look up its associated probability in the table.

x = 120, $\bar{x}$ = 105, s = 7

$$z = \frac{x - \bar{x}}{s}$$

z = (120 - 105) / 7

z = 2.14 ⟶ P(x < 120) = 0.9838

However we want P(x > 120) = 1 - P(x < 120)

= 1 - 0.9838

= 0.0162

1.6% of the drivers are expected to accumulate demerit points.

b) Find the z-scores for 101 km/h and 99km/h then look up their associated probabilities in the table. Finally subtract the probabilities.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0[illegible]15	0.0[illegible]15
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0[illegible]21	0.0[illegible]21
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0[illegible]29	0.0[illegible]28
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0[illegible]39	0.0[illegible]38
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0[illegible]52	0.0[illegible]51
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0[illegible]69	0.0[illegible]68
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0[illegible]91	0.0[illegible]89
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0[illegible]19	0.0[illegible]16
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0[illegible]54	0.0[illegible]50
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0[illegible]97	0.0[illegible]92
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0[illegible]50	0.0[illegible]44
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0[illegible]14	0.0[illegible]07
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0[illegible]92	0.0[illegible]84
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0[illegible]85	0.0[illegible]75
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0[illegible]94	0.0[illegible]82
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0[illegible]21	0.0[illegible]08
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0[illegible]69	0.0[illegible]53
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1[illegible]38	0.1[illegible]20
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1[illegible]30	0.1[illegible]10
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1[illegible]46	0.1[illegible]23
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1[illegible]85	0.1[illegible]60
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1[illegible]22
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2[illegible]06
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2[illegible]14
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.287[illegible]	0.2843

Lower limit: x = 99, $\overline{x}$ = 105, s = 7

$$z = \frac{x - \overline{x}}{s}$$

z = (99 - 105) / 7

z = -0.86 ⟶ P(x < 99) = 0.1949

Upper limit: x = 101, $\overline{x}$ = 105, s = 7

$$z = \frac{x - \overline{x}}{s}$$

z = (101 - 105) / 7

z = -0.57 ⟶ P(x < 101) = 0.2843

So the P(99 < x < 101) = P(x < 101) - P(x < 99)

= 0.2843 - 0.1949

= 0.0894

The probability of a given vehicle having a speed between 99 km/h and 101 km/h is 0.0894

9. Marketing analysts for a soft drink manufacturer conducted a survey in a large town. They determined that 42% of the 150 people surveyed regularly bought the soft drink.

a) Determine the margin of error at a 95% confidence limit.

b) Determine the confidence interval for the market share for the soft drink.

$$E = z\sqrt{\frac{p(1-p)}{n}}$$

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

a) A confidence level of 95% has a z-score of 1.960.

z = 1.960, p = 0.42, n = 150

$$E = 1.960\sqrt{\frac{0.42(1 - 0.42)}{150}}$$

E = 0.078985811

The margin of error at a confidence level of 95% is about 7.9%

b) Confidence Interval = Mean ± Margin of Error

Lower limit = 42 - 7.9 = 34.1%

Upper limit = 42 + 7.9 = 49.9%

The 95% confidence interval for a market share of 42% is 34.1% to 49.9%

10. The mean lifetime expected from a model of automobile follows a normal distribution with a standard deviation of 9500 km. A sample of 100 cars showed a mean lifetime of 190 000 km.

a) Determine the margin of error at a 90% confidence limit.

b) Determine the confidence interval for the mean.

$$E = z\frac{\sigma}{\sqrt{n}}$$

σ = standard deviation of the mean

Confidence Level	*z*-Score
90%	1.645
95%	1.960
99%	2.576

a) A confidence level of 90% has a z-score of 1.645.

$z = 1.645$, $n = 100$, $\sigma = 9500$

$E = 1.645(9500 / \sqrt{100})$

$E = 1.562.75$

The margin of error at a 90% confidence level is 1562.75 km

b) Confidence Interval = Mean ± Margin of Error

Lower limit = 190,000 - 1562.75 = 188437.25

Upper limit = 190,000 + 1562.75 = 191,562.75

The 90% confidence interval for the mean lifetime is from about 188,437.25 km to 191,562.75 km

11. A bus company has records showing that its buses arrive on time 95% of the time. Suppose the company operates 65 bus trips each day. The CEO has asked for the probability that 60 or more of these buses arrive on time.

a) Use the binomial distribution to determine the probability that 60 or more buses arrive on time.

b) Could this distribution be reasonably modelled using a normal approximation? Give reasons for your answer.

c) Determine the mean and standard deviation of the normal approximation.

d) Use the normal approximation to determine the probability that 60 or more buses arrive on time.

e) How does the answer to part d) compare with the answer to part a)?

p = P(on time) = 0.95, q = P(not on time) = 0.05, n = 65

a) Recall that for a binomial distribution $P(X) = {}_nC_r(p)^r(q)^{n-r}$

$P(X \geq 60) = P(60) + P(61) + P(62) + P(63) + P(64) + P(65)$

$= {}_{65}C_{60}(0.95)^{60}(0.05)^5 + {}_{65}C_{61}(0.95)^{61}(0.05)^4 + {}_{65}C_{62}(0.95)^{62}(0.05)^3 + {}_{65}C_{63}(0.95)^{63}(0.05)^2 + {}_{65}C_{64}(0.95)^{64}(0.05)^1 + {}_{65}C_{65}(0.95)^{65}(0.05)^0$

$= 0.1189 + 0.1852 + 0.2270 + 0.2054 + 0.1220 + 0.0356$

$= 0.8941$

The probability of 60 or more buses arriving on time is 0.8941

b) Check the restrictions for a binomial distribution to be considered to behave like a normal approximation...

$np = 65(0.95) = 61.75$

$nq = 65(0.05) = 3.25$

Since both of these are not > 5 then it is not reasonable to model this distribution using a normal approximation.

c) p = P(on time) = 0.95, q = P(not on time) = 0.05, n = 65

$\mu = np$
= 65(0.95)
= 61.75

$\sigma = \sqrt{npq}$
$= \sqrt{65(0.95)(0.05)}$
= 1.757

The mean is 61.75 and the standard deviation is about 1.757

d) For 60 or more buses to arrive on time, we need to use z-scores and the table. Applying the continuity correction we need P(59.5 ≤ x ≤ 65.5)

x = 59.5, $\bar{x}$ = 61.75, s = 1.757

$z = \frac{x - \bar{x}}{s}$

z = (59.5 - 61.75) / 1.757

z = -1.28 ⟶ P(x < 59.5) = 0.1003

x = 65.5, $\bar{x}$ = 61.75, s = 1.757

$z = \frac{x - \bar{x}}{s}$

z = (65.5 - 61.75) / 1.757

z = 2.13 ⟶ P(x < 65.5) = 0.9901

P(59.5 < x < 65.5) = P(x < 65.5) - P(x < 59.5)
= 0.9901 - 0.1003
= 0.8898

The probability of 60 or more buses arriving on time is 0.8898

e) The answer to part (a) is very slightly higher (0.43%) compared to the answer to part (d).

12. A newspaper reports that 350 of the 3500 people living in a small town have the flu. The data management class at the local high school has 25 students.

a) Do you expect this to be a representative sample? Explain.

b) Could you reasonably model this distribution using a normal approximation? Give reasons for your answer.

c) Determine the mean and standard deviation of the normal approximation.

d) Use the normal approximation to determine the probability that at least 5 of the students have the flu.

e) Use technology to compare the answer in part **d)** to that calculated from the hypergeometric distribution.

a) No this is not a representative sample because it only has high school students in it.

b) This is a hypergeometric distribution because once someone is selected, they can't be reselected. For a normal approximation to be reasonable the sample size needs to be less than 10% of the population. This is true.

c) $\mu = np$
= 25(0.1)
= 2.5

$\sigma = \sqrt{npq\left(\frac{NP-n}{NP-1}\right)}$

$= \sqrt{25(0.1)(0.9)\frac{3500 - 25}{3500 - 1}}$

= 1.4948

The mean is 2.5 and the standard deviation is about 1.4948

d) For at least 5 students to have flu, we need to use z-scores and the table. Applying the continuity correction we need $P(x \geq 4.5)$

$x = 4.5$, $\bar{x} = 2.5$, $s = 1.4948$

$$z = \frac{x - \bar{x}}{s}$$

$z = (4.5 - 2.5) / 1.4948$

$z = 1.34 \longrightarrow P(x < 4.5) = 0.9009$

However we want $P(x \geq 4.5) = 1 - P(x < 4.5)$

$= 1 - 0.9009$

$= 0.0991$

The probability of 5 or more students having flu is 0.0991

e) Using the hypergeometric formula from lesson 4.50 as well as the indirect method we can calculate $P(x \geq 5)$.

x	n	r	a	n-a	r-x	aCx	n-aCr-x	nCr	P(x)
0	3500	25	350	3150	25	1	1.68132E+62	2.3645E+63	0.071105977
1	3500	25	350	3150	24	350	1.34462E+60	2.3645E+63	0.199033046
2	3500	25	350	3150	23	61075	1.03201E+58	2.3645E+63	0.266565525
3	3500	25	350	3150	22	7084700	7.5883E+55	2.3645E+63	0.227364712
4	3500	25	350	3150	21	614597725	5.33534E+53	2.3645E+63	0.138678669
								SUM =	0.902747929

$P(x \geq 5) = 1 - P(x < 5)$

$= 1 - 0.902747929$

$= 0.0973$

The probability of 5 or more students having flu is 0.0973

The normal approximation gives a slightly higher answer of 9.9% versus 9.7%