The Normal Distribution and z-Scores

Lesson objectives

- I can determine the theoretical probability for a continuous random variable over a range of values
- I can determine the mean and standard deviation of a sample of values
- I can calculate and explain the meaning of a z-score
- I can solve real-world problems involving normal distributions

1.1

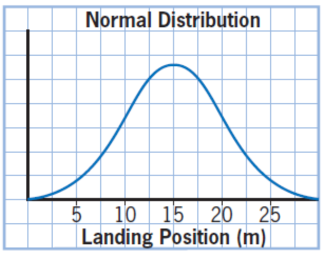
Lesson objectives | Teachers' notes | Lesson notes

MHR Page 341 #s 1 - 5, 9, 10 & 13

Warm up

In a spot landing contest, airplane pilots take turns trying to land as close as possible to a line painted on the runway. Most will land close to the line, with fewer and fewer as the distance from the line increases. This kind of probability distribution is referred to as a **normal distribution**.

Suppose that the landing zone is 30 m long, with the target line at 15 m. The graph of a normal distribution for the spot landing contest would look something like this:



Suggest two or three other situations that might reasonably be expected to form a normal distribution.

Study the shape of the normal distribution. What features can you see?

Definition

Normal Distribution

- A probability distribution around a central value, dropping off symmetrically to the right and left, forming a bell-like shape

Mathematically, the two "tails" of the normal distribution continue forever. In a real situation, the probability of finding a datum far from the central peak is essentially zero.

If you collect sample values of a variable that is expected to follow a normal distribution, the sample data will cluster around a central peak to form a bell-like shape, or "bell curve."

For a discrete distribution, you can calculate probabilities using counting techniques. For a continuous distribution, you can calculate probabilities by determining the area under a graph for a range of values. In this section, you will learn how to determine probabilities if you know that the distribution is a normal distribution.

Example 1

Use a Frequency Distribution to Estimate Probabilities

Fruityfizz Soft Drinks bottles its products in containers marked 500 mL. The table shows the frequency distribution for a sample of 200 bottles.

Volume (mL)	Frequency, f
490–492	0
492–494	0
494–496	2
496–498	11
498–500	43
500–502	81
502–504	48
504–506	14
506–508	1
508–510	0

a) Add a relative frequency column to the table.
b) Use the table to determine the probability that a given bottle will contain less than 500 mL of soft drink.
c) Use the table to determine the probability that a given bottle will contain between 498 mL and 502 mL of soft drink.
d) Is it possible to determine the probability that a given bottle will contain exactly 500 mL of soft drink using the table? Explain your answer.
e) Sketch a scatter plot of the frequency versus the volume. For the horizontal axis, use the midpoint of each interval.
f) Sketch a scatter plot of the relative frequency versus the volume. For the horizontal axis, use the midpoint of each interval. Sketch a smooth curve through the points. How does the shape of the graph of the frequency data compare to the shape of the graph of the relative frequency data?
g) Could you use the area under the relative frequency graph to answer parts b) and c)? Explain your answer.
h) Interpret the answers to parts b) and c) in relation to areas under a probability density graph.

Example 1

Use a Frequency Distribution to Estimate Probabilities

Fruityfizz Soft Drinks bottles its products in containers marked 500 mL. The table shows the frequency distribution for a sample of 200 bottles.

Volume (mL)	Frequency, f	Relative Frequency, $rf = \frac{f}{200}$
490–492	0	0.000
492–494	0	0.000
494–496	2	0.010
496–498	11	0.055
498–500	43	0.215
500–502	81	0.405
502–504	48	0.240
504–506	14	0.070
506–508	1	0.005
508–510	0	0.000

a) Add a relative frequency column to the table.
b) Use the table to determine the probability that a given bottle will contain less than 500 mL of soft drink.
c) Use the table to determine the probability that a given bottle will contain between 498 mL and 502 mL of soft drink.
d) Is it possible to determine the probability that a given bottle will contain exactly 500 mL of soft drink using the table? Explain your answer.
e) Sketch a scatter plot of the frequency versus the volume. For the horizontal axis, use the midpoint of each interval.
f) Sketch a scatter plot of the relative frequency versus the volume. For the horizontal axis, use the midpoint of each interval. Sketch a smooth curve through the points. How does the shape of the graph of the frequency data compare to the shape of the graph of the relative frequency data?
g) Could you use the area under the relative frequency graph to answer parts b) and c)? Explain your answer.
h) Interpret the answers to parts b) and c) in relation to areas under a probability density graph.

b) Add up the relative frequencies

= 0 + 0 + 0.010 + 0.055 + 0.215 = 0.280

The probability that a given bottle will have less than 500 mL of soft drink is 0.280

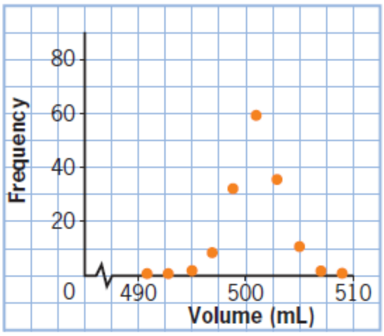
c) Again, add up the relative frequencies

= 0.215 + 0.405 = 0.620

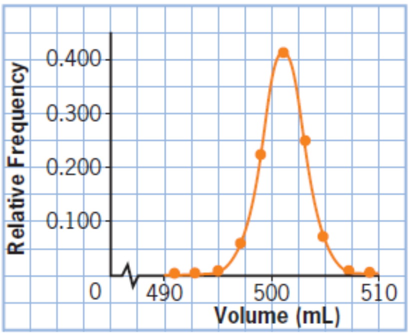
The probability that a given bottle will have between 498 mL and 502 mL of soft drink is 0.620

d) No. We don't have individual frequencies. Also, the bin width would be zero, so we cannot find the area under a probability distribution graph.

e)



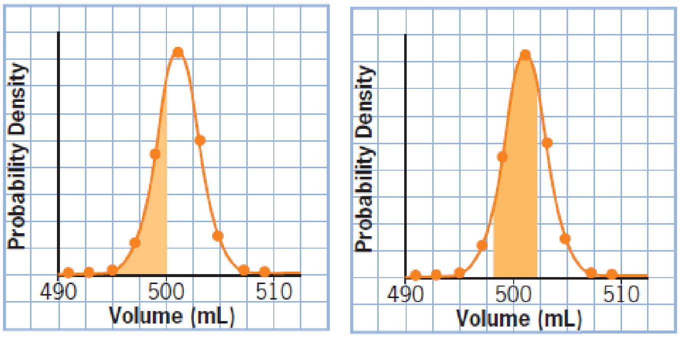
f)



They both have a similar shape to them.

g) Despite having the same general shape as a probability density graph, the area under a relative frequency graph does not equal one, so it cannot be used.

h)



P(x < 500) = area under the graph from far left to x = 500

P(498 < x < 502) = area under the graph from x = 498 to x = 502

Your Turn

A men's clothing store kept records of the waist sizes of a sample of their customers, measured to the nearest quarter of an inch. The table shows the relative frequencies.

Waist Size (in.)	Relative Frequency
26–28	0.000
28–30	0.025
30–32	0.175
32–34	0.295
34–36	0.300
36–38	0.160
38–40	0.045
40–42	0.000

What is the probability that a customer has a waist size

a) between 30 and 32 in.?

b) of more than 36 in.?

c) between 30 and 36 in.?

d) of exactly 38 in.?

a) P(30 < x < 32) = 0.175

The probability of a waist size between 30 and 32 inches is 0.175

b) P(x > 36) = 0.160 + 0.045 + 0 = 0.205

The probability of a waist size more than 36 inches is 0.205

c) P(30 < x < 36) = 0.175 + 0.295 + 0.300 = 0.770

The probability of a waist size between 30 and 36 inches is 0.770

d) We cannot find the probability for an exact value because we don't have the individual waist sizes. Also, the width of the bin would be zero on a probability distribution graph.

Example 2

Spot Landing Contest

Forty aircraft participated in a spot landing contest at the Wainfleet Ring Aerodrome. The landing zone was 30 m long, with the target line at the 15 m mark. The table shows the touchdown position of each aircraft along the landing zone. The data in the table are expected to follow a normal distribution.

a) Determine the mean and the standard deviation of the spot landing data.

b) What is the z-score for a pilot who lands her plane at a position of 18.3 m?

c) What is the probability that a pilot lands at a position of 18.3 m or less?

d) What is the probability that a pilot lands at a position of more than 18.3 m?

e) What is the probability that a pilot lands at a position between 12.2 m and 18.3 m?

Landing Zone Position (m)			
10.6	18.9	17.7	22.9
12.2	11.9	12.2	10.6
17.0	13.4	14.0	14.6
14.0	15.5	18.9	13.4
11.6	12.5	18.0	9.8
11.3	16.2	10.6	14.6
18.0	13.1	11.9	12.5
14.6	17.4	15.2	11.9
19.8	22.3	14.6	15.5
10.6	17.7	12.2	15.5

$\sum x^2$ = total(a^2)

a) Enter the data in desmos (a = [data,data,...,data]) mean(a) = 14.63 m

$$s = \sqrt{\frac{\sum x^2 - n(\bar{x})^2}{n-1}}$$

This is a sample because more than 40 aircraft are likely to use the aerodrome and therefore could have been selected.

$$= \sqrt{\frac{8975.68 - 40(14.63)^2}{40-1}}$$

standard deviation ≈ 3.259 metres

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

z = (18.3 - 14.63) / 3.259

z ≈ 1.12

The z-score for landing an aircraft at 18.3 m is about 1.12

c) Calculate the z-scores and then use the table on pages 480-481.

P(x ≤ 18.3) = 1.12 ⟶ P = 0.8686

z	0.00	0.01	0.02
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985
0.6	0.7257	0.7291	0.7324
0.7	0.7580	0.7611	0.7642
0.8	0.7881	0.7910	0.7939
0.9	0.8159	0.8186	0.8212
1.0	0.8413	0.8438	0.8461
1.1	0.8643	0.8665	0.8686
1.2	0.8849	0.8869	0.8888

These probabilities are very much like percentiles in that they tell us the probability for that value or less.

Note that P(x ≤ 18.3) is the same as P(x < 18.3) because we cannot find P(x = 18.3).

d) P(x > 18.3) = 1 - P(x ≤ 18.3)

= 1 - 0.8686

= 0.1314

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

z = (12.2 - 14.63) / 3.259

z ≈ -0.75 ⟶ P = 0.2266

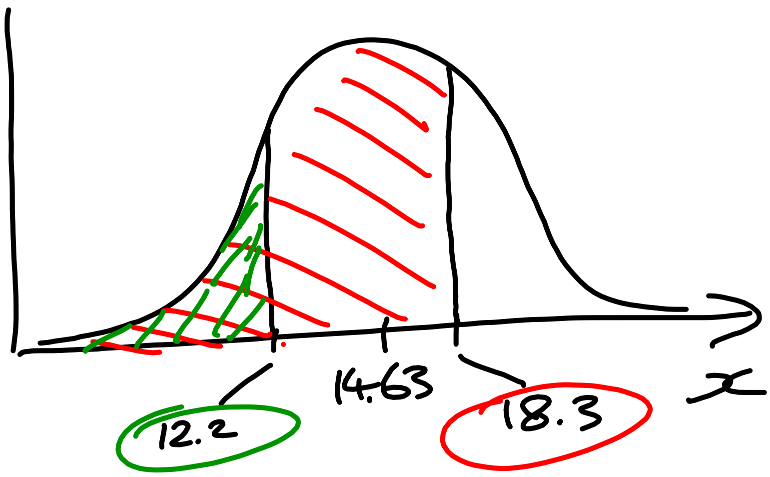
P(12.2 ≤ x ≤ 18.3) = P(x ≤ 18.3) - P(x ≤ 12.2)

= 0.8686 - 0.2266

= 0.6420

The probability of the pilot landing between 12.2 m and 18.3 m is about 0.6420

z	0.00	0.01	0.02	0.03	0.04	0.05
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054
–2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122
–2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158
–2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202
–1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256
–1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322
–1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401
–1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495
–1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606
–1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735
–1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885
–1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056
–1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251
–1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469
–0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711
–0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977
–0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266



Your Turn

Kunal is trying out for the school football team. He wants to know how far he can kick the ball for a field goal. The table shows the data from 20 trials.

Field Goal Distance (yd)			
17	27	31	25
25	44	35	24
31	48	42	48
45	34	41	38
40	43	45	21

a) Determine the mean and standard deviation of the data.

b) What is the probability that Kunal kicks a distance of less than 30 yd?

c) What is the probability that Kunal kicks a distance of 20 yd to 40 yd?

a) Enter the data in desmos (a = [data,data,...,data]) mean(a) = 35.2 yd

$$s = \sqrt{\frac{\sum x^2 - n(\overline{x})^2}{n-1}}$$

$$s = \sqrt{\frac{26520 - 20(35.2)^2}{(20 - 1)}}$$

standard deviation ≈ 9.567 metres

b) $z = \dfrac{x - \overline{x}}{s}$

z = (30.0 - 35.2) / 9.567

z ≈ -0.54

P(x < 30) = -0.54 ⟶ P = 0.2946

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
–2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
–2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
–2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
–1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
–1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
–1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
–1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
–1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
–1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
–1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
–1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
–1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
–1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
–0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
–0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
–0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
–0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
–0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776

z	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517
0.2	0.5793	0.5832	0.5871	0.5910
0.3	0.6179	0.6217	0.6255	0.6293
0.4	0.6554	0.6591	0.6628	0.6664
0.5	0.6915	0.6950	0.6985	0.7019
0.6	0.7257	0.7291	0.7324	0.7357

c) $z = \dfrac{x - \overline{x}}{s}$

z = (40.0 - 35.2) / 9.567

z ≈ 0.50

P(x ≤ 40) = 0.50 ⟶ P = 0.6915

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

z = (20.0 - 35.2) / 9.567

z ≈ -1.59

P(x ≤ 20) = -1.59 ⟶ P = 0.0559

P(20 ≤ x ≤ 40) = P(x ≤ 40) - P(x ≤ 20)

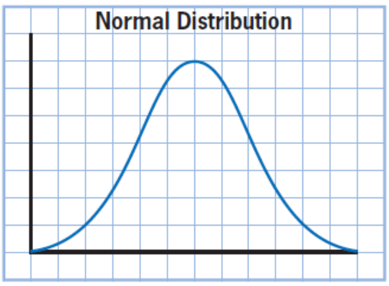
= 0.6915 - 0.0559

= 0.6356

The probability of Kunal kicking between 20 yd and 40 yd is about 0.6356

Key Concepts

- The frequency polygon approximates the shape of the frequency distribution.
- You must use a range of values to determine the theoretical probability for a continuous random variable.
- The probability that a continuous random variable takes on any single value is zero.
- A normal distribution is a probability distribution around a central value, dropping off symmetrically to the right and left, forming a bell-like shape.
- You can determine the probability that a variable will lie within a range of values by finding an area under the normal distribution.
- You can use z-scores to determine probabilities, either from a table or by using technology.



R1. Consider the field goal distances in the Your Turn following Example 2. As Kunal practises and gains more skill, would you expect the mean and standard deviation to remain the same? Explain your answer.

As Kunal practices and gains more skill, I expect the mean and standard deviation to change. As he improves and becomes more consistent, I expect the mean to increase and the standard deviation to decrease.

R2. Relate to the graph of a normal distribution to explain why $P(z < -5)$ is almost zero.

The z-scores for a normal distribution follow a normal distribution themselves. They have a mean of 0 and a standard deviation of 1. From the graph, P(z < -5) is located to the very left at the end of the tail. So, the probability that far from the central peak, is essentially zero.