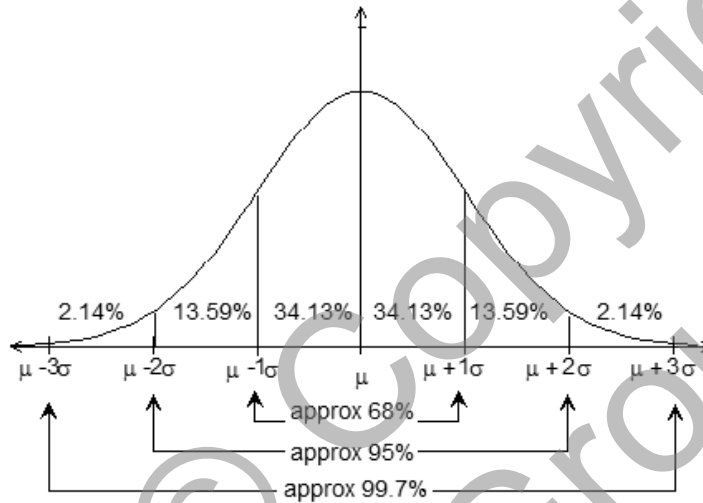


Chapter 14 – Normal Distribution

There are an infinite number of normal distributions, one for each selection of mean and standard deviation. In this course, we study standard normal distribution, where a normal random variable has a mean of zero and a standard deviation of one.

Normal distributions have a symmetrical, bell shaped curve where the area under the curve represents the probability.



Standardization Notation and Calculating Probability

A standardised score or a 'Z' score is the normal random variable of a standard normal distribution. When calculating the probability of a normal distribution it must be transformed from the standard normal distribution using the formula;

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Where

- 'Z' is the standardised score
- 'X' is the normal random variable score
- ' μ ' is the mean of X, and
- ' σ ' is the standard deviation or $\sqrt{\text{variance}}$ of X.

The area under the curve represents a probability of 1. This function gives the probability to the left of the value Z. It can be used to calculate;

$P(Z > a)$ the probability that a standard normal random variable falls between a given value 'a' and plus infinity.

$P(Z < a)$ the probability that a standard normal random variable falls between minus infinity and a given value 'a'.

$P(a < Z < b)$ the probability that a standard normal random variable falls between two given values 'a' and 'b'.

Determining Probability Using the Empirical Rule

The empirical rule states, that for a normal distribution:

- Approximately 68% of scores lie within one standard deviation of the mean.
- Approximately 95% of scores lie within two standard deviations of the mean.
- Approximately 99.7% of scores lie within three standard deviations of the mean.

Worked Example 1

If $X \sim N(5, 2^2)$, find $P(X < 7)$.

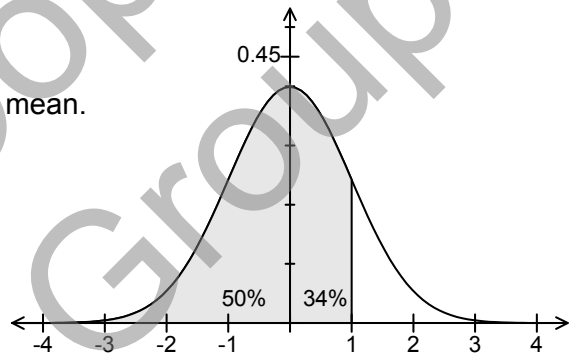
$$\mu = 5 \quad \sigma = 2$$

If we standardise a score of 7 we get;

$$Z = \frac{7-5}{2} = 1$$

i.e. a score of 7 is 1 standard deviation above the mean.

$$\begin{aligned} P(X < 7) &= P(Z < 1) \\ &= 50\% + 34\% \\ &= \underline{84\% \text{ approx.}} \end{aligned}$$



Worked Example 2

If $X \sim N(5, 2^2)$, find $P(3 < X < 9)$.

$$\mu = 5 \quad \sigma = 2$$

If we standardise a score of 3 we get;

$$Z = \frac{3-5}{2} = -1$$

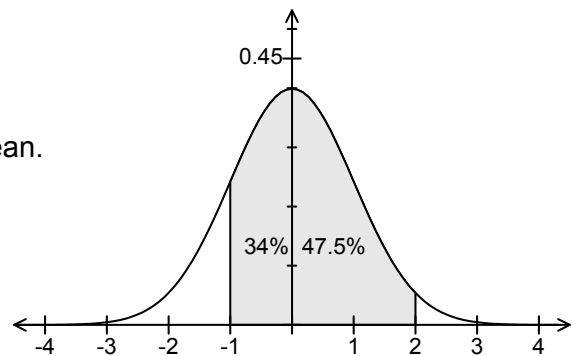
i.e. a score of 3 is 1 standard deviation below the mean.

If we standardise a score of 9 we get

$$Z = \frac{9-5}{2} = 2$$

i.e. a score of 9 is 2 standard deviations above the mean.

$$\begin{aligned} P(X < 7) &= P(-1 < Z < 2) \\ &= 34\% + 47.5\% \\ &= \underline{81.5\% \text{ approx.}} \end{aligned}$$



Determining Probability Using Technology

Worked Example 1

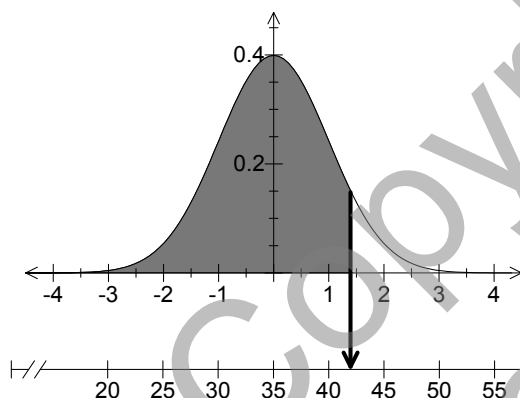
The wingspans of a particular type of beetle were normally distributed with a mean of 35 mm and a standard deviation of 5 mm. What proportion of beetles will have wing span less than 42 mm in length?

The first step is to find a standardized score corresponding to $X = 42$ mm.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{42 - 35}{5}$$

$$Z = 1.4$$



The second step is to find the cumulative probability associated with a Z value less than or equal to 1.4. The shaded (grey) area under the curve is represents the probability.

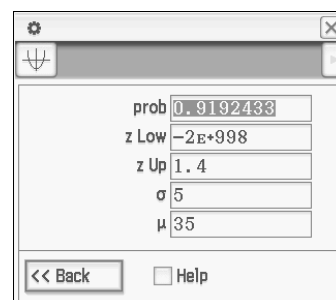
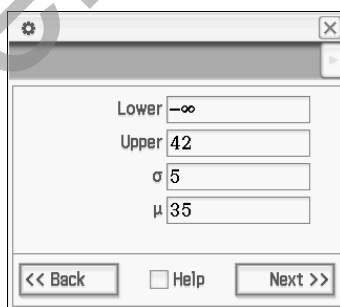
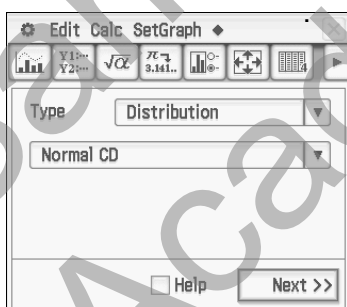
$$P(Z \leq 1.4) = 0.9192$$

This value can be found by utilising your calculator in one of two ways.

Option 1:



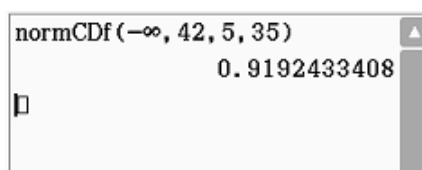
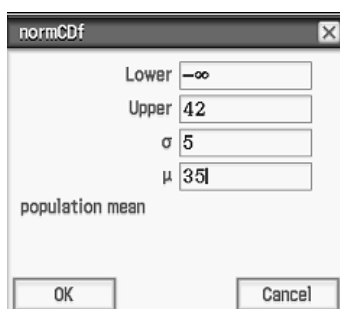
Calc > Distribution



Option 2:



Interactive > Distribution > Continuous > normCDF



The probability is 0.9192 (4 d.p.)

Worked Example 2

A packet of cooking chocolate contains 500g of chocolate. Suppose the actual weight (X) of these packets is normally distributed with a mean of 512 grams and a standard deviation of 8 grams.

- a) What is the probability of picking a packet between 504 and 520g?

or

$$X \sim N(512, 8^2), \text{ find the } P(504 \leq X \leq 520)$$

The probability is 0.6827 (4 d.p.)

- b) What are the minimum and maximum values of the middle 60%?

The minimum and maximum values are 505.5 and 518.5 respectively (to the nearest 0.5)

- c) If $X \sim N(512, 64)$, determine the 23rd percentile.

The 23rd percentile is 506.09.

Resource Free Questions

1. A manufacturing company approximates the length of its matchsticks using a normal distribution model. The matchsticks have a mean of 4.5 cm with a standard deviation of 1 mm.

For a randomly selected matchstick, use the empirical rule to determine the probability that it measures;

- a) between 4.4cm and 4.6 cm?
 - b) between 4.3cm and 4.7 cm?
 - c) between 4.2cm and 4.8 cm?
2. 600 scores are normally distributed with a mean of 135 and a standard deviation of 10. Use the empirical rule to determine each of the following.
- a) Approximately what percentage of scores lie within the interval 125 – 145?
 - b) What interval about the mean includes the middle 95% of the data?
 - c) Approximately what percentage of scores lie within the interval 125 – 155?
 - d) Approximately how many scores lie within the interval 125 – 145?

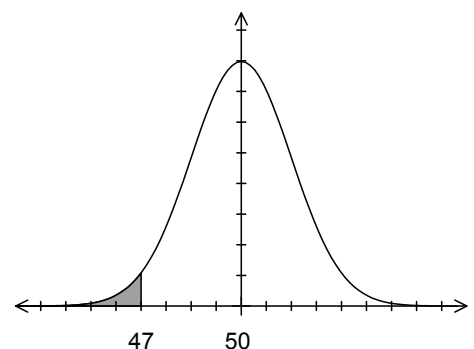
3. The standby time on a new model of smartphone is normal distributed. The mean standby time is 500 hours with a standard deviation of 18 hours. Use the empirical rule to what percentage of this particular model of smart phone has a standby time longer than 536 hours?

4. X is a normally distributed variable with a mean of 56 and a standard deviation of 4. Find, using the empirical rule;

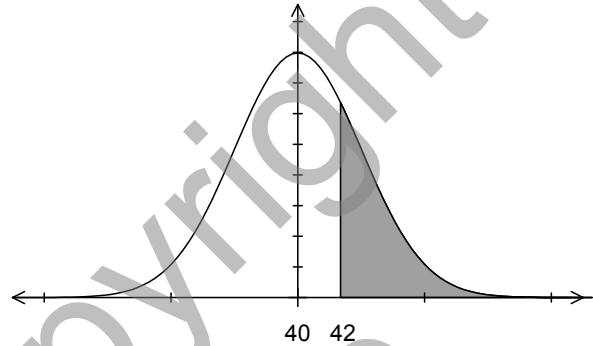
- a) $P(x < 60)$
- b) $P(x > 48)$
- c) $P(52 < x < 64)$

5. The shaded area in the diagram shown opposite is 0.3. If the mean is 50,

- a) what percentile is 47?
- b) to which percentile does 53 belong?



6. A set of scores is normally distributed with a mean of 14 and a standard deviation of 4.45. If the 25th percentile is 11, determine the interquartile range of the scores.
7. In the diagram shown on the right, the shaded area is 0.25. Determine the interquartile range of this set of scores.



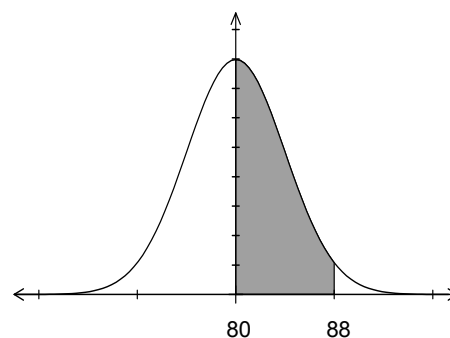
8. 95% of players at a football club weigh between 52 and 80 kg. Assuming that the data is normally distributed, which of the following is most likely to be the mean and standard deviation of scores? Justify your answer.

- a) Mean = 66 kg, Std Dev = 14 kg b) Mean = 66 kg, Std Dev = 7 kg

9. If $X \sim N(12.5, 3^2)$, find, using the empirical rule;

- a) $P(X < 15.5)$
 b) $P(6.5 < X < 9.5)$

10. A set of scores is normally distributed with a mean of 80. If the shaded area on the diagram below is approximately 0.475, what is the standard deviation for this set of scores?



Resource Rich Questions

11. A set of scores is normally distributed with a mean of 18 and a standard deviation of 2.6. If the 25th percentile is 16.25, determine the interquartile range of the scores.
12. The contents of a particular brand of tinned soup is normally distributed with a mean of 250 mL and standard deviation of 8 mL. Determine the probability that a randomly selected container tin contains between 250 and 260 mL of soup.
13. A set of scores is normally distributed with a mean of 85 and a standard deviation of 6. Determine the interquartile range of the scores.
14. In an international collaboration between universities, 7030 candidates were randomly selected to sit a test. Justin achieved a score of 930. The mean test score was 860 with a standard deviation of 90. Assuming that the test scores are normally distributed, estimate how many students had a higher score than Justin.
15. The time taken to complete a Public Sector Recruitment Test is normally distributed with a mean of 80 minutes and a standard deviation of 10 minutes.
 - a) What is the probability of a candidate completing the test in one hour or less?
 - b) What is the probability of a candidate taking between 65 and 75 minutes to complete the test.
 - c) Sam completed the test in 55 minutes. What percentage of candidates completed the test faster than Sam?
 - d) What is the interquartile range for completion times?
16. A coffee machine claims to dispense 150 mL of the selected beverage. The volumes of the drinks dispensed are found to be normally distributed with a mean of 149.5mL and a standard deviation of 1.4mL.
 - a) What proportion of drinks will be between 149 and 151 mL?
 - b) What is the least that can be expected in the largest 5% of beverages?

17. A weights of 9 gram sachets of saffron are normally distributed with a mean of 9.12 grams and a standard deviation of 0.15 grams.
- What is the probability that a randomly selected sachet of saffron will weigh 9.2 grams or more?
 - What are the first and last deciles for this distribution?
18. A new experimental drug undergoes its first round of clinical trials. The time taken for patients to feel its effects are found to be normally distributed with a mean of 20 minutes and standard deviation of 4 minutes.
- What percentage of patients takes longer than 15 minutes to feel the effects of the drug?
 - What percentage of patients takes between 15 and 25 minutes to feel the effects of the drug?

The clinical trial found that some patients were “immune” to the experimental drug and experienced no effect after taking the drug. It was determined that if a patient did not react to the drug after 35 minutes, then they are deemed to be “immune” to its effects.

- According to the data, how many patients in 100 000 will be immune to the effects of this drug?
19. Julia recently completed her end of year exams. Julia finds out from her languages teachers that the Italian exam had a mean of 48 and a standard deviation of 12 and the French exam had a mean of 62% and a standard deviation of 8.
- Julia scored 66 in French. Convert this result to a z-score.
 - Julia scored 71 in Italian. Convert this result to a z-score.
 - In which subject did Julia attain the “better” result? Justify your answer.
20. Consider the table below displaying Daniel's scores in three different science subjects.

Subject	Daniel's Score	Mean	Std. Dev.
Biology	66	67	7
Chemistry	54	48	5
Physics	61	56	8

Compare the overall distribution of the scores to determine which of the three subjects Daniel performed best in.

Solutions

1. a) *approx. 68%* b) *approx. 95%* c) *approx. 99.7%*

2. a) *approx. 68%* b) $115 - 155$
c) $34 + 47.5 = \textit{approx. 81.5\%}$ d) $\frac{68}{100} \times 600 = 408$

3. $P(X > 536) = 100 - (50 + 47.5)$
 $P(X > 536) = 2.5\% \textit{ approx}$

4. a) $50 + 34 = 84\% \textit{ approx}$
b) $50 + 47.5 = 97.5\% \textit{ approx}$
c) $34 + 47.5 = 81.5\% \textit{ approx}$

5. a) 30^{th} b) 70^{th}

6. *25th percentile is 11, then 75th percentile is 17*
 $\therefore IQR = 17 - 11 = 6$

7. $42 - 38 = 4$

8. Option B. $66 \pm 2(7) = 80 \textit{ and } 52$

9. a) $50 + 34 = 84\%$
b) $47.5 - 34 = 13.5\%$

10. $0.475 \times 2 = 0.95 \textit{ or } 95\%$

88 is two standard deviations above the mean

$\therefore \textit{ the standard deviation is 4.}$

11. *25th percentile is 16.25, then 75th percentile is 19.75*

invNormCDF("L", 0.75, 2.6, 18)	▲
	19.75367335

$\therefore IQR = 19.75 - 16.25 = 3.5$

12. 0.3944 (4 dp.)

normCDF(250, 260, 8, 250)
0.3943502263

13. $IQR = 89.05 - 80.95 = 8.09$

x_1 InvN	80.953061
x_2 InvN	89.046939
prob	0.5
σ	6
μ	85

14. Approximately 1535 candidates.

normCdf(930, ∞ , 90, 860)	0.2183500154
0.2183500154 \times 7030	1535.000608

15. a) 0.4975 (4 dp.) b) 151.80 mL (2 dp.)

normCdf(149, 151, 1.4, 149.5)	0.4975191832
invNormCdf("L", 0.95, 1.4, 149.5)	151.8027951

16. a) 0.0228 (4 dp.)
 b) 0.2417 (4 dp.)
 c) 0.0062 (4 dp.)
 d) $IQR = 13.49$ (2 dp.)

normCdf($-\infty$, 60, 10, 80)	0.02275013195
normCdf(65, 75, 10, 80)	0.2417303375
normCdf($-\infty$, 55, 10, 80)	6.209665326E-3
invNormCdf("L", 0.25, 10, 80)	73.2551025
invNormCdf("L", 0.75, 10, 80)	86.7448975
86.7448975 - 73.2551025	13.489795

17. a) normCdf(9.6, ∞ , 0.15, 9.2) = 0.0912
 b) 1st Decile 8.93 (2 dp.)
 9th Decile 9.31 (2 dp.)

x_1 InvN	8.9277673
x_2 InvN	9.3122327
prob	0.8
σ	0.15
μ	9.12

18. a) 0.8943 (4 dp.)
 b) 0.7887 (4 dp.)
 c) 8.84 ~ 9 patients per 100 000

```
normCDF(15, ∞, 4, 20)
0.8943502263
normCDF(15, 25, 4, 20)
0.7887004527
normCDF(35, ∞, 4, 20)
8.84172852E-5
8.84172852E-5×100000
8.84172852
□
```

19. a) $Z_{(French)} = \frac{66-48}{12}$ b) $Z_{(Italian)} = \frac{71-62}{8}$
 $Z_{(French)} = 1.5$ $Z_{(Italian)} = 1.13$ (2 dp.)
- c) In comparison to the mean of each subject, Julia attained a better result in French, achieving a score 1.5 standard deviations above the mean.

20. $z_{(biology)} = \frac{66-67}{7} = -0.1\dot{6}$ $z_{(chemsitry)} = \frac{54-48}{5} = 1.2$ $z_{(physics)} = \frac{61-56}{14} = 0.625$

Daniel performed best in Chemistry where his score was 1.2 standard deviations above the mean.

Sample © Copyright
 Academic Group