Exercise walk-through

2.8
Distributions – The Normal Distribution

GAC Mathematics

Statistics: collecting and displaying data

- A **population** is everyone you want to describe; a **sample** is who you actually measure.
- To generalise, the **sample** must be **representative**: chosen so that every part of the **population** has a fair chance of being included.
- **Qualitative data** are categories; **quantitative data** measure amounts, either **discrete** (counted) or **continuous** (measured).
- **Bias** enters through who you ask, who answers and how the question is worded. The *Literary Digest* poll had all three problems: its lists favoured the wealthy, and the 24% who replied were a self-selected group.

$$D_1 = \frac{20}{10} = 2 \qquad D_2 = \frac{20}{20} = 1$$

Statistics: measures of variability

- The **range** is the largest value minus the smallest, so one extreme value can change it a lot.
- The **interquartile range (IQR)** is $Q_3 - Q_1$, the spread of the middle half. A **box plot** shows the minimum, quartiles, **median** and maximum.
- The **standard deviation** measures spread using squared distances from the **mean**. For a whole **population**, $\sigma = \sqrt{\sum (x-\mu)^2/n}$; for a sample estimate, $s = \sqrt{\sum (x-\bar{x})^2/(n-1)}$. It has the same units as the data.
- Two data sets can share a mean and be completely different: a city with every day at 20 °C and one swinging between 0 °C and 40 °C have the same mean temperature.

$$\text{range} = 12 - 2 = 10 \qquad IQR = 8 - 4 = 4$$

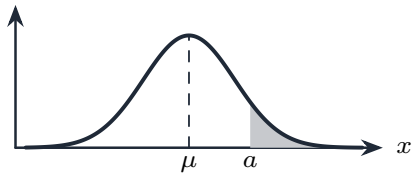
Distributions: the normal distribution

- The **normal distribution** is a symmetric, bell-shaped model described by its **mean** μ and **standard deviation** σ. Many measurements, such as adult heights, are close to normal.
- About 68% of values lie within one standard deviation of the mean, 95% within two, and 99.7% within three.
- A **z-score** says how many standard deviations a value is from the mean: $z = \dfrac{x - \mu}{\sigma}$. It lets you compare values on different scales.
- Use the cumulative area $\Phi(z) = P(Z \le z)$ and symmetry: $\Phi(-z) = 1 - \Phi(z)$.

$$z = \frac{60 - 50}{10} = 1 \qquad P(X > 60) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

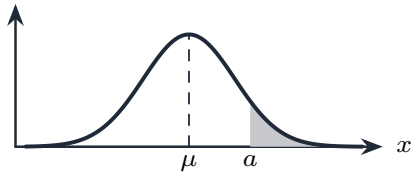
Exercise 2.1

2.1 A normal model has mean 100 and standard deviation 15. Find the z-score of 130. [2]



Exercise 2.1

2.1 A normal model has mean 100 and standard deviation 15. Find the z-score of 130. [2]



2.1

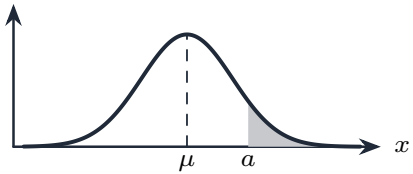
Standardise using the model mean and standard deviation.

$$z = \frac{x - \mu}{\sigma} = \frac{130 - 100}{15} = 2$$

Z-score 2.

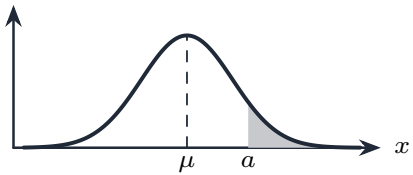
Exercise 2.2

2.2 Using the supplied cumulative values, find the probability that a standard-normal value is greater than 1. [2]



Exercise 2.2

2.2 Using the supplied cumulative values, find the probability that a standard-normal value is greater than 1. [2]



2.2

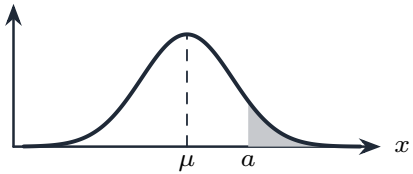
Use the complement of the area at or below 1.

$$P(Z > 1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

Probability 0.1587.

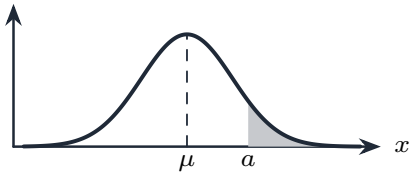
Exercise 2.3

2.3 Use symmetry and the supplied values to find the probability that a standard-normal value lies between negative 1 and 1. [3]



Exercise 2.3

2.3 Use symmetry and the supplied values to find the probability that a standard-normal value lies between negative 1 and 1. [3]



2.3

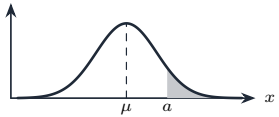
Symmetry gives $\Phi(-1) = 1 - 0.8413 = 0.1587$. Subtract cumulative areas.

$$P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826$$

Probability 0.6826.

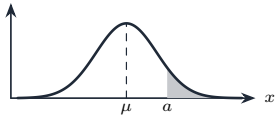
Exercise 2.4

2.4 A normal model has mean 70 and standard deviation 8. Find the raw values corresponding to z-scores negative 1 and 1.5. Give the probability above the latter value. [3]



Exercise 2.4

2.4 A normal model has mean 70 and standard deviation 8. Find the raw values corresponding to z-scores negative 1 and 1.5. Give the probability above the latter value. [3]



2.4

Invert standardisation: $x = \mu + z\sigma$.

$$x_1 = 70 + (-1)(8) = 62$$

Lower raw value.

$$x_2 = 70 + 1.5(8) = 82$$

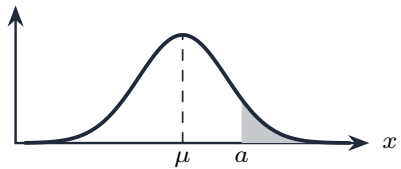
Upper raw value.

$$P(X > 82) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$$

Tail probability.

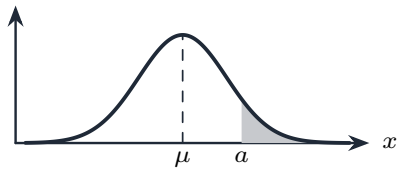
Exercise 3.1

3.1 A stated normal model for completion time has mean 40 minutes and standard deviation 5 minutes. (a) Find the probability of completion between 35 and 50 minutes. (b) For 200 independently modelled tasks, find the expected number exceeding 50 minutes. Explain why that expectation is not a guaranteed observed count. [6]



Exercise 3.1

3.1 A stated normal model for completion time has mean 40 minutes and standard deviation 5 minutes. (a) Find the probability of completion between 35 and 50 minutes. (b) For 200 independently modelled tasks, find the expected number exceeding 50 minutes. Explain why that expectation is not a guaranteed observed count. [6]



3.1

(a) Standardise both bounds.

$$z_L = \frac{x_L - \mu}{\sigma} = \frac{35 - 40}{5} = -1$$

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 3.1

3.1

(a) Standardise both bounds.

$$z_L = \frac{x_L - \mu}{\sigma} = \frac{35 - 40}{5} = -1$$

$$z_U = \frac{x_U - \mu}{\sigma} = \frac{50 - 40}{5} = 2$$

Both bounds correct. Subtract cumulative areas.

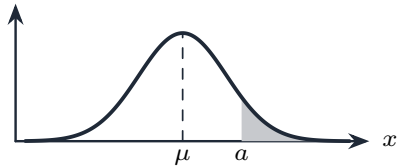
$$p = \Phi(2) - \Phi(-1) = 0.9772 - 0.1587 = 0.8185$$

Interval probability.

(b) Tail probability is $1 - 0.9772 = 0.0228$, so expected count $E = Np = 200(0.0228) = 4.56$. The random observed count is an integer and can vary; 4.56 is an expectation over repeated modelled samples.

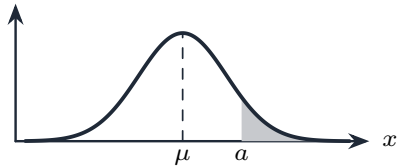
Exercise 3.2

3.2 Two stated normal score models are A with mean 60 and standard deviation 10, and B with mean 75 and standard deviation 5. A learner scores 80 on A and 85 on B. (a) Find both z-scores and compare relative position. (b) Use the supplied table to find each upper-tail probability. (c) For 500 modelled scores on each scale, find each expected count above those thresholds. (d) Explain two limits of using these calculations to claim equal ability on different tests. [8]



Exercise 3.2

3.2 Two stated normal score models are A with mean 60 and standard deviation 10, and B with mean 75 and standard deviation 5. A learner scores 80 on A and 85 on B. (a) Find both z-scores and compare relative position. (b) Use the supplied table to find each upper-tail probability. (c) For 500 modelled scores on each scale, find each expected count above those thresholds. (d) Explain two limits of using these calculations to claim equal ability on different tests. [8]



3.2

(a) Standardise relative to each model.

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 3.2

3.2

(a) Standardise relative to each model.

$$z_A = \frac{80 - 60}{10} = 2$$

A z-score.

$$z_B = \frac{85 - 75}{5} = 2$$

B z-score. Both scores are two model standard deviations above their respective means.

(b) For each model, $p = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$.

(c) Expected count on each scale is $E = Np = 500(0.0228) = 11.4$.

(d) Normality and representative parameters must be justified for the actual data. Equal standardised relative position does not establish equivalent test content, difficulty, reliability or learner ability; separate evidence is needed.