SPECIALIST MATHEMATICS

UNITS 3 & 4
Contents

About this series iii
About this book viii
About the authors x

Complex numbers 1
1.1 Complex numbers 2
Exam prep 1.1 10
Exam practice 1.1 11
1.2 The Argand diagram 13
Exam prep 1.2 17
Exam practice 1.2 19
1.3 Modulus and argument 21
Exam prep 1.3 29
Exam practice 1.3 30
1.4 Graphs in the complex plane 33
Exam prep 1.4 39
Exam practice 1.4 40
1.5 De Moivre’s theorem 44
Exam prep 1.5 47
Exam practice 1.5 47
1.6 Roots of unity 49
Exam prep 1.6 53
Exam practice 1.6 53
1.7 Factorising polynomials over the complex field 55
Exam prep 1.7 60
Exam practice 1.7 61
1.8 Solving polynomial equations 62
Exam prep 1.8 69
Exam practice 1.8 69
Summary 74

Complex numbers examination 1 79
Complex numbers examination 2 85

Rational and circular functions 93
2.1 Rational functions 94
Exam prep 2.1 102
Exam practice 2.1 103
2.2 Graphing rational functions 104
Exam prep 2.2 110
Exam practice 2.2 110
2.3 Quotient functions 113
Exam prep 2.3 117
Exam practice 2.3 118

2.4 The absolute value function 120
Exam prep 2.4 127
Exam practice 2.4 129
2.5 The reciprocal circular functions 133
Exam prep 2.5 140
Exam practice 2.5 141
2.6 The compound angle formulas 143
Exam prep 2.6 148
Exam practice 2.6 149
2.7 Trigonometric identities 150
Exam prep 2.7 151
Exam practice 2.7 152
2.8 The inverse circular functions 153
Exam prep 2.8 160
Exam practice 2.8 160
Summary 164

Functions examination 1 169
Functions examination 2 171

Vectors 177
3.1 Operation with vectors 178
Exam prep 3.1 188
Exam practice 3.1 189
3.2 Linear dependence and independence of vectors 193
Exam prep 3.2 196
Exam practice 3.2 196
3.3 Resolving vectors 198
Exam prep 3.3 199
3.4 Scalar product of vectors 201
Exam prep 3.4 205
Exam practice 3.4 206
3.5 Parallel and perpendicular vectors 211
Exam prep 3.5 212
Exam practice 3.5 213
3.6 Vector proofs of geometric results 215
Exam prep 3.6 216
Exam practice 3.6 216
Summary 221
**Differentiation**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The product, quotient and chain rules</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.1</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.1</td>
<td>230</td>
</tr>
<tr>
<td>4.2</td>
<td>Differentiating circular functions</td>
<td>232</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.2</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.2</td>
<td>234</td>
</tr>
<tr>
<td>4.3</td>
<td>Differentiating the inverse circular functions</td>
<td>236</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.3</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.4</td>
<td>239</td>
</tr>
<tr>
<td>4.4</td>
<td>Differentiating exponential and logarithmic functions</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.4</td>
<td>242</td>
</tr>
<tr>
<td>4.5</td>
<td>The second derivative</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.5</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.5</td>
<td>245</td>
</tr>
<tr>
<td>4.6</td>
<td>Applying the second derivative</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.6</td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.6</td>
<td>250</td>
</tr>
<tr>
<td>4.7</td>
<td>The chain rule and related rates of change</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.7</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.7</td>
<td>253</td>
</tr>
<tr>
<td>4.8</td>
<td>Implicit differentiation</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>Exam prep 4.8</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Exam practice 4.8</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>260</td>
</tr>
</tbody>
</table>

**Integration**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Statics</td>
<td>266</td>
</tr>
<tr>
<td></td>
<td>Exam prep 5.1</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td>Exam practice 5.1</td>
<td>272</td>
</tr>
<tr>
<td>5.2</td>
<td>Statics involving gravitational force</td>
<td>275</td>
</tr>
<tr>
<td></td>
<td>Exam prep 5.2</td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>Exam practice 5.2</td>
<td>282</td>
</tr>
<tr>
<td>5.3</td>
<td>Objects in motion</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>Exam prep 5.3</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>Exam practice 5.3</td>
<td>296</td>
</tr>
<tr>
<td>5.4</td>
<td>Speed and distance from resultant force</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>Exam prep 5.4</td>
<td>315</td>
</tr>
<tr>
<td></td>
<td>Exam practice 5.4</td>
<td>316</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>322</td>
</tr>
</tbody>
</table>

**Exam prep**

- 4.1
- 4.2
- 4.3
- 4.4
- 4.5
- 4.6
- 4.7
- 4.8
- 5.1
- 5.2
- 5.3

**Exam practice**

- 4.1
- 4.2
- 4.3
- 4.4
- 4.5
- 4.6
- 4.7
- 4.8
- 5.1
- 5.2
- 5.3

**Summary**

- 4
- 5
- 6
**Differential equations**  373

7.1  Differential equations  374  
Exam prep 7.1  379  
Exam practice 7.1  379

7.2  Differential equations of the form \( \frac{dy}{dx} = f(x) \)  382  
Exam prep 7.2  385  
Exam practice 7.2  386

7.3  Differential equations of the form \( \frac{d^2y}{dx^2} = f(x) \)  388  
Exam prep 7.3  390  
Exam practice 7.3  390

7.4  Differential equations of the form \( \frac{dy}{dx} = f(y) \)  392  
Exam prep 7.4  400  
Exam practice 7.4  401

7.5  Separation of variables \( \frac{dy}{dx} = f(y)g(x) \)  409  
Exam prep 7.5  412  
Exam practice 7.5  412

7.6  Slope fields  414  
Exam prep 7.6  418  
Exam practice 7.6  418

7.7  Euler’s method  423  
Exam prep 7.7  429  
Exam practice 7.7  429

7.8  Variable acceleration applications  472  
Exam prep 8.4  478  
Exam practice 8.4  478

**Areas and volumes of integration**  485

9.1  Areas under curves  486  
Exam prep 9.1  495  
Exam practice 9.1  496

9.2  The graph of the anti-derivative  501  
Exam prep 9.2  504  
Exam practice 9.2  505

9.3  Volumes of solids of revolution  507  
Exam prep 9.3  513  
Exam practice 9.3  514

9.4  Arc lengths of curves  521  
Exam prep 9.4  525  
Exam practice 9.4  525

**Summary**  482  
Exam prep 7.5  412  
Exam practice 7.5  412

**Vector functions**  543

10.1  Position vectors  544  
Exam prep 10.1  549  
Exam practice 10.1  549

10.2  Differentiating and integrating vectors  552  
Exam prep 10.2  554  
Exam practice 10.2  555

10.3  Applying vectors to motion  559  
Exam prep 10.3  563  
Exam practice 10.3  563

**Summary**  566  
Mechanics and vectors examination 1  568  
Mechanics and vectors examination 2  571

**Kinematics**  435

8.1  Straight-line motion  436  
Exam prep 8.1  445  
Exam practice 8.1  446

8.2  Velocity–time graphs  452  
Exam prep 8.2  458  
Exam practice 8.2  458

8.3  The equations of kinematics  463  
Exam prep 8.3  468  
Exam practice 8.3  469
Random variables and statistical inference  579

11.1 Linear combinations of random variables  580
Exam prep 11.1  587
Exam practice 11.1  588

11.2 Independent normal random variables  590
Exam prep 11.2  592
Exam practice 11.2  593

11.3 Sample means and simulations  595
Exam prep 11.3  601
Exam practice 11.3  602

11.4 Confidence intervals for the population mean  605
Exam prep 11.4  612
Exam practice 11.4  613

11.5 Hypothesis testing related to the mean  615
Exam prep 11.5  622
Exam practice 11.5  623

11.6 Errors in hypothesis testing  626
Exam prep 11.6  628
Exam practice 11.6  628

Summary  630

Probability and statistics examination 1  635
Probability and statistics examination 2  637

Solutions  643
Glossary and index  721
CHAPTER 11

RANDOM VARIABLES AND STATISTICAL INFERENCE

11.1 Linear combinations of random variables
   - The mean of $aX + b$
   - The variance of $aX + b$
   - The mean and variance of a linear combination of two independent random variables

11.2 Independent normal random variables
   - Using CAS: Finding normal probabilities

11.3 Sample means and simulations
   - The sampling distribution of the sample means
   - Using CAS: The distribution of the sample means by simulation
   - Larger samples and the sampling distribution of the sample means
   - The standard error for the sample means

11.4 Confidence intervals for the population mean
   - Confidence intervals of a normally distributed sample
   - The central limit theorem
   - Using CAS: The confidence interval for the population mean
   - Confidence intervals and the margin of error

11.5 Hypothesis testing related to the mean
   - The null and alternative hypotheses
   - p-values for hypothesis testing related to the mean
   - Significance levels for making decisions
   - One-tailed and two-tailed tests
   - Critical values for a given significance level
   - Using CAS: Finding the critical z-value

11.6 Errors in hypothesis testing
   - Summary
Random variables have a probability distribution that can be used to predict the likelihood of a particular outcome occurring. The total value of homes that a real estate agent sells in a month, for example, will have a probability distribution which can be used to predict the probability that the agent will sell more than a particular value in a given month. If the total monthly commission the agent earns is a linear function of the value of homes sold, then properties of the distribution like the mean and variance of the commission can also be determined. The monthly commission is a linear combination of the random variable, the value of homes sold.

The mean of $aX + b$

The mean or expected value of a random variable $X$ is the long-term average value of $X$. If the random variable $X$ is the height of preschool-aged girls in Victoria, then $E(X)$ is the average height of all preschool-aged girls in Victoria.

There are two formulas for the expected value of the random variable $X$, depending on whether the random variable is discrete or continuous. (This is covered in detail in Mathematical Methods Unit 4, Chapters 7 and 11).

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(X) = \sum x \cdot p(x)$</td>
<td>$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) , dx$</td>
</tr>
</tbody>
</table>
Consider a random variable \( X \) with an expected value \( E(X) \).

A random variable \( Y \) is a linear transformation of \( X \) if \( Y = aX + b \).

The expected value of \( Y \) is

\[
E(aX + b) = a \ E(X) + b
\]

### Worked example 1

The probability distribution of a discrete random variable \( X \) is shown below.

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pr(X = x) )</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**a** Find the mean of \( X \).

**b** Find \( E(3X + 2) \).

**Working**

\[
E(X) = \sum x \cdot p(x) = 1.0
\]

The mean or \( E(X) = 1 \).

\[
E(3X + 2) = 3 \ E(X) + 2 = 3 \times 1 + 2 = 5
\]
The probability density function for a continuous random variable \( X \) is given by

\[
f(x) = \begin{cases} 
4x^3, & 0 \leq x \leq 1 \\
0, & \text{elsewhere}
\end{cases}
\]

Find

a  \( E(X) \)

b  \( E(10X - 4) \)

**Working**

a  1  Use the formula

\[
E(X) = \int_{-\infty}^{\infty} x f(x) \, dx
\]

and simplify.

2  Evaluate the integral.

\[
E(X) = \left[ \frac{4x^5}{5} \right]_0^1 = \frac{4(1)^5}{5} - 0 = \frac{4}{5}
\]

b  Substitute into the formula

\[
E(aX + b) = aE(X) + b \quad \text{with} \quad a = 10, \ b = -4.
\]

\[
E(10X - 4) = 10 \times E(X) - 4 = 10 \times \frac{4}{5} - 4 = 4
\]

**The variance of \( aX + b \)**

The variance and standard deviation of a random variable \( X \) measure the spread of the variable about the mean.

The formulas for the variance and standard deviation are:

\[
\text{Var}(X) = E(X^2) - \mu^2 \quad \text{SD}(X) = \sqrt{\text{Var}(X)}
\]

where \( \mu = E(X) \)

There are two formulas for \( E(X^2) \), depending on whether the random variable is discrete or continuous.

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(X^2) = \sum x^2 \cdot p(x) )</td>
<td>( E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) , dx )</td>
</tr>
</tbody>
</table>
A random variable \( Y \) is a linear transformation of \( X \) if \( Y = aX + b \).

The variance of \( Y \) is

\[
\text{Var}(aX + b) = a^2 \text{Var}(X)
\]

In the formula \( Y = aX + b \), the addition of \( b \) increases all the data values by \( b \) but does not influence the spread.

**Worked example 3**

The probability density function for a continuous random variable \( X \) is given by

\[
f(x) = \begin{cases} 
4x^3, & 0 \leq x \leq 1 \\
0, & \text{elsewhere}
\end{cases}
\]

Find the variance and hence find \( \text{Var}(5X + 10) \).

1. Write the formula for the variance and calculate \( E(X) \) and \( E(X^2) \). The mean \( E(X) \) was calculated in Worked example 2.

   \[
   E(X) = \int_0^1 x \times 4x^3 \, dx = \frac{4}{5}
   \]

   \[
   E(X^2) = \int_0^1 x^2 \times 4x^3 \, dx
   \]

   \[
   = \int_0^1 4x^5 \, dx
   \]

   \[
   = \left[ \frac{4x^6}{6} \right]_0^1
   \]

   \[
   = \frac{2(1)^6}{3} - 0
   \]

   \[
   = \frac{2}{3}
   \]

2. Substitute in the variance formula from Step 1.

   \[
   \text{Var}(X) = \frac{2}{3} - \left( \frac{4}{5} \right)^2
   \]

   \[
   = \frac{2}{3} - \frac{16}{25}
   \]

   \[
   = \frac{2}{3} \times \frac{2}{75}
   \]

3. Substitute into the formula

   \[
   \text{Var}(aX + b) = a^2 \text{Var}(X)
   \]

   \[
   \text{Var}(5X + 10) = 5^2 \text{Var}(X)
   \]

   \[
   = 25 \times \frac{2}{75}
   \]

   \[
   = \frac{2}{3}
   \]
The mean and variance of a linear combination of two independent random variables

Jenny builds a fence with a mean time of 2 hours and a variance of 0.5 hours. Abrar paints a fence with a mean time of 3 hours and a variance of 0.3 hours. Both the time for building the fence and the time to paint the fence are independent random variables. In this section, we will examine the mean and variance of combinations of two independent random variables.

For two independent random variables \(X\) and \(Y\),

\[
E(aX + bY) = aE(X) + bE(Y)
\]

\[
\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)
\]

At Franken’s Furniture, the time taken to build a dining table, \(X\) hours, is a continuous random variable with a mean time of 15 hours and a variance of 9 hours and the time taken to stain and polish the table, \(Y\) hours, is a continuous random variable with a mean time of 6 hours and a variance of 4 hours. Find the mean and variance of the time taken to complete a dining table.

**Working**

1. The total time taken is the time to build the table and the time to stain and polish the table.

2. Use \(E(aX + bY) = aE(X) + bE(Y)\) with \(a = 1, b = 1\) to find the mean time.

\[
E(1X + 1Y) = 1E(X) + 1E(Y) = E(X) + E(Y) = 15 + 6 = 21 \text{ hours}
\]

3. Use \(\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)\) with \(a = 1, b = 1\) to find the variance.

\[
\text{Var}(1X + 1Y) = 1^2 \text{Var}(X) + 1^2 \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y) = 9 + 4 = 13 \text{ hours}
\]
Two independent random variables $X$ and $Y$ have means of 30 and 20 and variances of 5 and 8 respectively. If $Z = 4X + 2Y$, find the mean and variance of $Z$.

1. To find the mean of $Z$, use the formula
   
   $E(aX + bY) = a \ E(X) + b \ E(Y)$. 

   $E(Z) = E(4X + 2Y)$
   $= 4 \ E(X) + 2 \ E(Y)$
   $= 4 \times 30 + 2 \times 20$
   $= 160$

2. To find the variance of $Z$, use the formula
   
   $Var(aX + bY) = a^2 \ Var(X) + b^2 \ Var(Y)$. 

   $Var(Z) = Var(4X + 2Y)$
   $= 4^2 \ Var(X) + 2^2 \ Var(Y)$
   $= 4^2 \times 5 + 2^2 \times 8$
   $= 112$

Two independent discrete random variables $X$ and $Y$ have the probability distributions shown.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$Pr(X = x)$</th>
<th>$y$</th>
<th>$Pr(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>15</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>20</td>
<td>0.1</td>
</tr>
</tbody>
</table>

If $Z = X + Y$, find

a. $E(Z)$

b. $Var(Z)$
1. Find the mean and variance of the discrete random variable $X$ using the formulas

\[ E(X) = \sum x \cdot p(x) \]
\[ E(X^2) = \sum x^2 \cdot p(x) \]
\[ \text{Var}(X) = E(X^2) - \mu^2 \]

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
<th>$x \cdot p(x)$</th>
<th>$x^2 \cdot p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.6</td>
<td>1.8</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>1.2</td>
<td>4.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2.4</td>
<td>7.6</td>
</tr>
</tbody>
</table>

$E(X) = \mu = 2.4$
$E(X^2) = 7.6$
$\text{Var}(X) = E(X^2) - \mu^2$
$= 7.6 - 2.4^2$
$= 1.84$

2. Find the mean and variance of the discrete random variable $Y$.

<table>
<thead>
<tr>
<th>$y$</th>
<th>$p(y)$</th>
<th>$y \cdot p(y)$</th>
<th>$y^2 \cdot p(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.3</td>
<td>1.5</td>
<td>7.5</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>1.0</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>0.1</td>
<td>1.5</td>
<td>22.5</td>
</tr>
<tr>
<td>20</td>
<td>0.1</td>
<td>2.0</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>6.0</td>
<td>80</td>
</tr>
</tbody>
</table>

$E(Y) = \mu = 6.0$
$E(Y^2) = 80$
$\text{Var}(Y) = E(Y^2) - \mu^2$
$= 80 - 6.0^2$
$= 44$

3. Find the expected value and variance of $Z$ using the formulas

\[ E(aX + bY) = a \ E(X) + b \ E(Y) \]
\[ \text{Var}(aX + bY) = a^2 \ \text{Var}(X) + b^2 \ \text{Var}(Y) \]

For $Z = X + Y$,

$E(Z) = E(X) + E(Y)$
$= 2.4 + 6.0$
$= 8.4$

$\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$
$= 1.84 + 44$
$= 45.84$
Linear combinations of random variables

**Prep 1**  
**WORKED EXAMPLE 1**

The discrete random variable $X$ has the probability distribution shown.

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr($X = x$)</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Find

a  $E(X)$  

b  $E(10X + 3)$

**Prep 2**  
**WORKED EXAMPLE 2**

A continuous random variable $X$ has a probability density function given by

$$f(x) = \begin{cases} 
\frac{1}{288} (12x - x^2), & 0 \leq x \leq 12 \\
0, & \text{otherwise}
\end{cases}$$

Find

a  the expected value of $X$  

b  the expected value of $\frac{1}{3}X + 2$.

**Prep 3**  
**WORKED EXAMPLE 3**

A continuous random variable $X$ has a probability density function given by

$$f(x) = \begin{cases} 
\frac{x}{8}, & 0 \leq x \leq 4 \\
0, & \text{otherwise}
\end{cases}$$

Find

a  $\text{Var}(X)$  

b  $\text{Var}(18X + 14)$

**Prep 4**  
**WORKED EXAMPLE 4**

When Eddy cleans his house, he dusts and vacuums. The time that he takes to dust the house is a continuous random variable with a mean of 30 minutes and a variance of 25 minutes. The time that Eddy takes to vacuum the house is another continuous random variable with a mean of 40 minutes and a variance of 64 minutes. Find the mean and the variance of the time that Eddy takes to clean his house.

**Prep 5**  
**WORKED EXAMPLE 5**

Two independent random variables $X$ and $Y$ have means of 50 and 15 and variances of 10 and 20 respectively. If $Z = 3X + 4Y$, find the mean and variance of $Z$. 
Two independent discrete random variables $X$ and $Y$ have the probability distributions shown.

<table>
<thead>
<tr>
<th>$x$</th>
<th>Pr$(X = x)$</th>
<th>$y$</th>
<th>Pr$(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>5</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>10</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>15</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>20</td>
<td>0.1</td>
</tr>
</tbody>
</table>

If $Z = 3X + Y$, find

a. $E(Z)$

b. $Var(Z)$

---

**Linear combinations of random variables**

**Question 1**

$X$ is a random variable with a mean of 45 and a variance of 24.
The mean of the variable $Y = 5X - 2$ is

A. 225  
B. 120  
C. 223  
D. 118  
E. 69

**Question 2**

$X$ is a random variable with a mean of 12 and a variance of 9.
The variance of the variable $Y = 4X + 12$ is

A. 144  
B. 204  
C. 192  
D. 156  
E. 36

**Question 3**

$X$ is a random variable with a mean of 36 and a standard deviation of 4.
The variance of the variable $Y = 2X - 5$ is

A. 8  
B. 64  
C. 16  
D. 3  
E. 67

**Question 4**

Two independent random variables $X$ and $Y$ have means of 15 and 10 and variances of 4 and 2 respectively.
The mean and variance of $5X + 2Y$ is

A. 95 and 24  
B. 415 and 108  
C. 25 and 6  
D. 95 and 6  
E. 95 and 108
The following information refers to Questions 5 & 6.

The mean price per punnet of strawberries and blueberries is $1.50 and $2.50 and the variance is $0.60 and $0.80 respectively.

**Question 5**

The mean price of four punnets of strawberries and five punnets of blueberries is
A $4.00  B $18.50  C $1.40  D $17.50  E $86.50

**Question 6**

The variance of the price of four punnets of strawberries and five punnets of blueberries is
A $6.40  B $29.60  C $18.50  D $1.40  E $5.40

**Question 7**

A random variable $X$ has the probability density function

$$f(x) = \begin{cases} \frac{x}{2}, & \text{if } x \in [0,2] \\ 0, & \text{otherwise} \end{cases}$$

Find

a $E(X)$  b $E(3X + 6)$  c $\text{Var}(X)$  d $\text{Var}(3X + 6)$

**Question 8**

A random variable $Y$ has probability density function

$$f(y) = \begin{cases} \frac{3}{32}(4y - y^2), & 0 \leq y \leq 4 \\ 0, & \text{otherwise} \end{cases}$$

Find the variance of $Z = 3Y - 11$.

**Question 9**

Each school day, Sam walks from his house to the bus stop and then catches the bus to school.

The time taken to walk to the bus stop is a continuous random variable $X_1$ with a mean of 10 minutes and a variance of 1 minute. The time taken by the bus is a continuous random variable $X_2$ with a mean of 20 minutes and a variance of 5 minutes. If $X_1$ and $X_2$ are independent, find

a the mean and variance of the total time to get to school
b the mean and variance of the total travel time for the school week (assume that the times for the return trip from school to home have the same mean and variance).

**Question 10**

Two independent random variables $X$ and $Y$ have means of 11 and 7 and variances of 6 and 4 respectively. Find the mean and variance of $9X - 3Y$. 

---

9780170386449
A linear combination of two independent *normal* random variables also has a normal distribution.

If $X$ is normally distributed with a mean of $\mu_X$ and a variance of $(\sigma_X)^2$, and $Y$ is an independent normally distributed random variable with a mean of $\mu_Y$ and a variance of $(\sigma_Y)^2$, then $aX + bY$ is also normally distributed with

**Mean:** $E(aX + bY) = a\mu_X + b\mu_Y$

**Variance:** $\text{Var}(aX + bY) = a^2(\sigma_X)^2 + b^2(\sigma_Y)^2$

### Worked example 7

A Year 10 class sits a Maths test and an English test. The scores on the Maths test are normally distributed with a mean of 65 and a standard deviation of 5, and the scores on the English test are normally distributed with a mean of 60 and a standard deviation of 10.

If the test scores are independent, find

**a** the mean and variance of the total score of the two tests

**b** the probability, correct to three decimal places, that a student will score better on the Maths test than on the English test.

**Working**

1. Let $X =$ the Maths test score and $Y =$ the English test score
2. Use $E(aX + bY) = a\mu_X + b\mu_Y$ to find the mean of the total score.
3. Use $\text{Var}(aX + bY) = a^2(\sigma_X)^2 + b^2(\sigma_Y)^2$ to find the variance of the total score.
b 1 Write a probability equation. \[ \Pr(X > Y) = \Pr(X - Y > 0) \]

2 Find the mean and variance of \(X - Y\) by using the formulas for mean and variance mentioned in part a.

\[
E(X - Y) = \mu_X + (-1)\mu_Y \\
= 65 - 60 \\
= 5 \\
\Var(X - Y) = (\sigma_X)^2 + (-1)^2(\sigma_Y)^2 \\
= 5^2 + 10^2 \\
= 125
\]

3 Find the standard deviation.

Standard deviation = \(\sqrt{125} = 5\sqrt{5}\)

4 Calculate the normal probability using CAS (see below).

\[ \Pr(X - Y > 0) \approx 0.673 \]

### Using CAS | Finding normal probabilities

The probability for the normal distribution in Worked example 7 is calculated below.

#### TI-NSPIRE CAS

Press \(\text{menu}, 5: \text{Probability}, 5: \text{Distributions}, 2: \text{NormalCdf}\).

Enter 0 for the lower bound, \(\infty\) as the upper bound, 5 as the mean and \(5\sqrt{5}\) as the standard deviation, then press \(\text{enter}\).

\[ \Pr(X - Y > 0) = 0.673 \]

#### CLASSPAD

**STEP 1**

In \(\sqrt{\text{Main}}\), tap \(\text{Interactive, Distribution/Inv. Dist, Continuous, normCDF}\).

**STEP 2**

Enter 0 for the lower bound, \(\infty\) as the upper bound, \(5\sqrt{5}\) as the standard deviation and 5 as the mean.

\[ \Pr(X - Y > 0) = 0.673 \]
A factory produces nuts and bolts. The weight of the bolts is normally distributed with a mean of 15 g and a standard deviation of 0.4 g. The weight of the nuts is normally distributed with a mean of 5 g and a standard deviation of 0.2 g. Find

a the mean and the standard deviation of the combined mass of a nut and a bolt 
b the probability, correct to three decimal places, that the combined mass is less than 20.5 g.

The following information refers to questions 1 and 2.

Two independent normal random variables $X$ and $Y$ have means of 14 and 24 and standard deviations of 3 and 6 respectively.

**Question 1**
The mean of $X + Y$ is

A 9  B 10  C 38  D 24  E 47

**Question 2**
The standard deviation of $X + Y$ is

A 9  B $3\sqrt{5}$  C 45  D 38  E 3

**Question 3**
Two independent normal random variables $X$ and $Y$ have standard deviations of 10 and 4 respectively. The standard deviation of $X - Y$ is

A $\sqrt{6}$  B $2\sqrt{21}$  C 6  D 84  E $2\sqrt{29}$

**Question 4**
$X$ is a normal random variable with a mean of 125 and a standard deviation of 10, and $Y$ is a normal random variable with a mean of 25 and a standard deviation of 5. If $X$ and $Y$ are independent, the mean and standard deviation of $2X - 4Y$ is

A 350 and $20\sqrt{2}$  B 150 and $5\sqrt{5}$  C 150 and 0
D 150 and $20\sqrt{2}$  E 135 and 30
Question 6

Two independent normal random variables $X$ and $Y$ have means of 20 and 40 and standard deviations of 4 and 12 respectively. $\Pr(X + Y < 55)$ is closest to

- A 0.346
- B 0.415
- C 0.384
- D 0.106
- E 0.654

Question 7

Each weekday morning, Felicity drives to the station and then catches the train to work. The time taken for the car trip has a mean of 20 minutes and a standard deviation of 4 minutes. The time taken for the train trip has a mean of 15 minutes and a standard deviation of 2 minutes. The time to drive to work and the time to travel by train are independent variables.

Find

- a the mean and standard deviation of the total travel time to work
- b the probability, correct to three decimal places, that Felicity takes less than 37 minutes to get to work.

Question 8

A deluxe car wash comprises a wash and vacuum. The times for the wash and vacuum are independent normally distributed variables. If the time for the car wash has a mean of 18 minutes and a standard deviation of 3 minutes, and the time for the vacuum has a mean of 12 minutes with a standard deviation of 2 minutes, find the probability, correct to three decimal places, that 3 cars can be washed and vacuumed in less than 100 minutes.
Now we will examine the properties of the sampling distribution of the sample means. We can use CAS to generate random samples from a normally distributed population. This will ensure that the population mean is known and we can investigate properties of the sample mean.

The sampling distribution of the sample means

One hundred random samples of 20 integers are taken from the set \{1, 2, 3, \ldots, 100\} and the mean of each sample is calculated. The mean of the population is 50.5.

Each of the 100 samples has a different mean and these means are shown in the dot plot.

The means of the samples are symmetrically spread about the centre, which is the mean of all the sample means.

In this simulation, the mean of the sample means is 51.2, which is approximately equal to the mean of the population (50.5).
Using CAS

**The distribution of the sample means by simulation**

Generate 100 samples of size 30 from a population that is normally distributed with a mean of 40 and a standard deviation of 5.

Calculate the mean of each sample and plot the distribution of the sample means.

**TI-NSpire CAS**

**STEP 1**

On a Lists & Spreadsheets page, label column A 'mean30'.

In cell A1, press `=`.

From page 1, select mean(.)

Press `b`, 3: Data, 5: Random, 4: Normal.

Enter 40,5,30)) and press `.`.

This will calculate the mean of a sample of 30 normal random variables with a mean of 40 and a standard deviation of 5.

Position the cursor in cell A1.

Press `b`, 3: Data, 3: Fill.

Use the touch pad to fill down to cell A100 and press `.`.

**Syntax:** \( \text{randnorn}(\text{mean, standard deviation, sample size}) \)

**STEP 2**

Press `c` and add a Data & Statistics page.

On the horizontal axis, use the touch pad to select 'mean30' from 'click to add variable'.

This graph shows the 100 sample means from the samples of size 30.

**STEP 3**

Select the Lists & spreadsheets page. Position the cursor back in cell A1.


The mean of the sample means is close to 40.
CLASSPAD

STEP 1
From \[\text{Menu}\], select Spreadsheet menu \[\text{Spreadsheet}\].
In cell A1, press \[\text{Keyboard}\].
Tap \[\text{Catalogue}\], \[\text{M}\] and select \text{mean}(, double tap.
Tap \[\text{R}\], select \text{randNorm}(, double tap.
Enter \(5, 40, 30\), tap \[\text{EXE}\].
Syntax: \text{randNorm(standard deviation, mean, sample size)}

\[\text{Your numbers and graph will vary.}\]

STEP 2
Place the cursor in cell A1, and tap \[\text{Edit}, \text{Fill, Fill Range}\].
Complete \(A1 : A100\) and tap \[\text{OK}\]. Be patient, it takes a little while.

STEP 3
To graph the distribution, select cell A1 and tap \[\text{Edit}, \text{Select Range}\].
Enter the range \(A1 : A100\) and tap \[\text{OK}\].
Tap \[\text{Graph, Histogram}\].

STEP 4
Tap \[\text{Resize}\] to fill the whole screen.
To calculate the summary statistics, tap the column heading A, select the range \(A1 : A100\).
Tap \[\text{Calc, One-variable}\].
The mean of the sample means is close to 40.
The sampling distribution of the sample means contains the means ($\bar{x}$) of all possible samples of size $n$ from a population.

The mean of the sample means ($\mu_\bar{x}$) is equal to the mean of the population ($\mu$).

$$\mu_\bar{x} = \mu$$

There is some variability in the sample means and the standard deviation is a measure of the variability. The standard deviation of the sample means is the standard error of the sample mean.

**Worked example 8**

A random sample is taken from a population that is normally distributed with a mean of 80 and a standard deviation of 10. Using simulation, 100 of these random samples is generated and the results are shown in the dot plot.

Find the probability that a sample contains a mean greater than 82.

Count the number of sample means greater than 82. The probability is a fraction of the total number of samples.

**Working**

10 of the samples contain sample means greater than 82.

Probability of a mean greater than 82 = $$\frac{10}{100} = \frac{1}{10}$$
The amount of time spent in the supermarket shopping for groceries is normally distributed with a mean of 30 minutes and a standard deviation of 6 minutes.

a. Simulate 50 samples of size 40 and calculate the sample means for each sample. Display the results using CAS.

b. Find the average of the sample means for the samples of size 40.

**Working**

a. Generate the means of the samples by CAS using the randNorm command.

**TI-Nspire CAS**

Cell formula
\[ \text{=mean(randnorm}(30,6,40)) \]
Copy for 50 cells.

**CLASSPAD**

Cell formula
\[ \text{=mean(randNorm}(6,30,40)) \]
Copy for 50 cells.

b. Calculate one-variable statistics for the data in column A. Answer may vary slightly.

Mean of \( \bar{x} \) = 29.95

Mean of \( \bar{x} \) = 29.85

**Larger samples and the sampling distribution of the sample means**

When a random sample is taken from a population, the reliability of the sample statistics is influenced by the sample size. This effect can be seen when the mean is calculated with samples of different sizes.

For example, in the situation below, samples are taken from a normally distributed population with a mean of 30 and a standard deviation of 5.

Sample means with a sample size of 20

Sample means with a sample size of 200
The effect of a larger sample size:

- The sample mean is closer to the population mean.
- There is less variability in the means.

A larger sample size produces less error and a more reliable estimate of the population mean.

### Sample size

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Mean of $\bar{x}$</th>
<th>Standard deviation of $\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>29.96</td>
<td>1.174</td>
</tr>
<tr>
<td>200</td>
<td>30.01</td>
<td>0.346</td>
</tr>
</tbody>
</table>

Random samples of size $n$ are taken from a population that is normally distributed with a mean of 160 and a standard deviation of 20. Complete 100 simulations for sample sizes of 30 and 150 and determine the mean and standard deviation of the sample means.

#### Working

1. Generate the means of the samples by CAS using the randNorm command.

   - **TI-Nspire CAS**
     
     Cell formula: `=mean(randNorm(160,20,n))`
     
     Copy for 100 cells.

   - **ClassPad**
     
     Cell formula: `=mean(randNorm(20,160,n))`
     
     Copy for 100 cells.

2. Answers may vary slightly.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Mean of $\bar{x}$</th>
<th>Standard deviation of $\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>159.50</td>
<td>3.385</td>
</tr>
<tr>
<td>150</td>
<td>159.77</td>
<td>1.621</td>
</tr>
</tbody>
</table>
The standard error for the sample means

The standard deviation of the sample means is called the standard error of a sample mean. This standard error of \( \bar{x} \) is written as \( SE(\bar{x}) \) or \( \sigma_{\bar{x}} \). The standard error indicates how much the sample mean \( \bar{x} \), for a particular sample size, deviates from the population mean \( \mu \).

The standard error for a sample mean is given by

\[
SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}
\]

where \( \sigma \) is the population standard deviation and \( n \) is the sample size.

We are unlikely to know the population standard deviation, so the sample standard deviation is used as an approximation for the population standard deviation.

**Worked example 11**

A random sample of 200 pears is taken at a fruit cannery. The weight of the pears is normally distributed with a mean weight of 70 g and a standard deviation of 8 g. Find, correct to 3 decimal places, the standard error of the sample mean.

**Working**

Use the formula \( SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} \).

\[
SE(\bar{x}) = \frac{8}{\sqrt{200}} = 0.566
\]

---

**Sample means and simulations**

**Prep 1**

A random sample is taken from a population that is normally distributed with a mean of 80 and a standard deviation of 10. Using simulation, 100 of these random samples is generated and the results are shown in the dot plot.

Find the probability that a sample contains a mean less than 78.

**Prep 2**

The amount of time a person spends standing, in a normal working day, is normally distributed with a mean of 400 minutes and a standard deviation of 30 minutes.

a  Simulate 50 samples of size 80 and calculate the sample means for each sample. Display the results using CAS.

b  Find the average of the sample means for samples of size 80.
Random samples of size $n$ are taken from a population that is normally distributed with a mean of 250 and a standard deviation of 50. Complete 100 simulations for sample sizes of 20 and 200 and determine the mean and standard deviation of the sample means.

Copy and complete the table below.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Mean of $\bar{x}$</th>
<th>Standard deviation of $\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A random sample of 100 eggs is taken at a free-range farm. The weight of the eggs is normally distributed with a mean weight of 12 g and a standard deviation of 1.5 g. Find the standard error in the sample mean.

**Sample means and simulations**

*The following information refers to questions 1 and 2.*

The weight of hatched chicks is normally distributed with a mean weight of 70 g and a standard deviation of 3 g. A simulation of 120 samples of size 60 is completed and the sample means are shown.

**Question 1**

The probability of a sample having a sample mean less than 69.2 is

A $\frac{1}{20}$  B $\frac{3}{70}$  C $\frac{1}{3}$  D $\frac{1}{40}$  E $\frac{3}{100}$

**Question 2**

The probability of a sample having a sample mean greater than $x$ is $\frac{1}{10}$.

The value of $x$ is

A 70.8  B 70.4  C 70.6  D 70  E 70.2
The following information refers to Questions 3, 4 & 5.

Random samples of size 50 are taken from a population that is normally distributed with a mean of \(a\) and a standard deviation of \(b\), and the mean of the samples is calculated. This process is repeated by simulation \(x\) times and the one-variable statistics for the sample means is shown.

**Question 3**
The number of simulations is
- A 750
- B 50
- C 150
- D 100
- E 80

**Question 4**
The value of the population mean, \(a\), is
- A 750
- B 80
- C 50
- D 3.5
- E 741

**Question 5**
The standard deviation of the population is
- A 20
- B 3.52
- C 25
- D 3.5
- E 31.5

**Question 6**
A random sample of size 64 is taken from a normally distributed population with a mean of 40 and a standard deviation of 8. The standard error for the sample means is
- A 5
- B 8
- C 1
- D 4
- E 2

**Question 7**
A random sample of 50 items is taken from a normally distributed population.
The sample mean is 29.95 and the standard deviation of the sample means of size 50 is 0.566.
Find, to the nearest integer,
\(a\) the population mean
\(b\) the population standard deviation.
Question 8

The birth weight of babies is normally distributed with a mean of $\mu$ and a standard deviation of $\sigma$. A random sample of 100 babies is taken and the birth weights are recorded. If the sample mean is 3.49 kg and the standard deviation of the sample means of size 100 is 0.05 kg, find, correct to one decimal place,

a the population mean

b the population standard deviation.

Question 9

The volume of soft drink in a can is normally distributed with a mean of 345 mL and a standard deviation of 2 mL.

a Simulate 50 sample means by taking random samples of size 60. Present the summary as a dot plot or histogram.

b Find the mean of the sample means and the standard deviation of the sample means.
A parameter is a characteristic value of a particular population, such as the mean. A statistic is an estimate of a parameter found from a sample. In most cases, you cannot be absolutely certain of the value of a parameter because you cannot obtain values from a whole population, so you must use a statistic instead.

A point estimate of a parameter is a single value obtained from a sample. The statistic used is called an estimator of the parameter.

An interval estimate of a parameter is an interval that is likely to include the value of the parameter.

Suppose we need to find the amount of time Year 12 students spend on social media per day and we find the mean daily social media use from a sample of 50 Year 12 students. This sample mean is a point estimate of the population mean. An interval estimate for the daily social media use of all Year 12 students is an interval that is likely to contain the population mean.

The observed interval will vary from sample to sample. The frequency with which the interval contains the population parameter being estimated is given by the confidence level.
For example, if we take 100 samples of size $n$ from a population to estimate the population mean $\mu$ and calculate 95% confidence intervals for the sample mean for each sample, then about 95 of these intervals will contain the population mean.

For a **confidence interval** symmetrical about the mean in a distribution:

- the **confidence level** is equivalent to the probability that the population parameter being estimated lies within the confidence interval
- the **margin of error** is the distance of the ends of the interval from the mean. It is half of the confidence interval
- the **tails** are the sections above and below the confidence interval

In the confidence interval $(a, b)$, $a$ is the **lower quantile** and $b$ is the **upper quantile**.

The confidence level of a confidence interval is equal to the proportion of values in the distribution that lie between the upper and lower quantiles. If a confidence interval of $(a, b)$ has a confidence level of 95%, then 95% of the values in the distribution will be included in this interval.

For a confidence interval $(a, b)$ with a confidence level of $c$:

$$\Pr(a < X < b) = c$$

The **tails** represent the values in the distribution that are less than $a$ or greater than $b$.

$$\Pr(X < a) = \Pr(X > b) = \frac{1-c}{2}$$
Confidence intervals of a normally distributed variable

The inverse cumulative normal distribution is used to determine the confidence interval at any confidence level for a normally distributed variable.

**Worked example 12**

The time, in minutes, that Year 12 students spend on social media is a normally distributed random variable with a mean of 80 minutes and a standard deviation of 9 minutes. Find a 95% confidence interval, correct to one decimal place, for this variable.

**Working**

1. Determine the percentage of values in each tail.
   
   The confidence interval contains 95% of the sample.
   
   The proportion in each tail \( \frac{1-0.95}{2} = 0.025 \)

2. Use the inverse cumulative normal distribution to determine the quantiles of the confidence interval.

   \( \Pr(X < \text{lower quantile}) = 0.025 \)
   \( \Pr(X < \text{upper quantile}) = 1 - 0.025 = 0.975 \)

3. Use CAS to determine the quantiles of the confidence interval.

**TI-Nspire CAS**

On a calculator page, press \( \text{menu}, 5: \text{Probability}, 5: \text{Distributions}, 3: \text{Inverse Normal} \). For the lower quantile, enter 0.025 for the area, 80 for the mean and 9 for the standard deviation. Repeat for the upper quantile, entering 0.975 for the area.

95% confidence interval = (62.4, 97.6)

**CLASSPAD**

In \( \sqrt{\text{Main}}, \) tap \( \text{Interactive, Distribution/Inv.Dist, Inverse, invNormCDF} \).

For the lower quantile, select ‘left’ for the tail, enter 0.025 for the area, 9 for the standard deviation, 80 for the mean, and tap \( \text{OK} \).

For the upper quantile, repeat, selecting ‘right’ for the tail.

95% confidence interval = (62.4, 97.6)
The central limit theorem

The Tasmanian Farmers Fruit Growers’ Association wants to determine the mean weight of Golden Delicious apples grown in their state in this year’s harvest. They take random samples of 100 apples from randomly-selected orchards and use this sample to estimate the mean weight of every Golden Delicious apple in this year’s Tasmanian harvest.

The sampling distribution of the sample means of size 100 is the set of every possible sample mean and is represented by \( \bar{X} = \{ \bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k \} \). In this set, \( \bar{x}_1 \) is the mean of sample 1, \( \bar{x}_2 \) is the mean of sample 2 and so on. The standard deviation of the set of sample means of size 100, written as \( \sigma_{\bar{X}} \), is a measure of the spread of the sample means about the mean. The properties of this sampling distribution allow predictions to be made about the population mean \( \mu \).

The central limit theorem states that the distribution of the sample means taken from any kind of distribution is approximately normally distributed. The larger the sample size, \( n \), the closer the approximation to the normal distribution. \( n = 30 \) is generally considered large enough to assume the sampling distribution is normal.

\[
\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}
\]

If \( X \) is a normally distributed random variable, then the sampling distribution of sample means, of any size \( n \), is also normal.

Worked example 13

A random sample of size 50 is taken from a population with a standard deviation of 5. If the mean of the sample is 75, find a 90% confidence interval for the population mean (answer correct to 2 decimal places).

1. Estimate the sampling distribution standard deviation.

\[
\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{50}} = \frac{5}{\sqrt{2}} = \frac{\sqrt{2}}{2}
\]

2. Since \( n \geq 30 \), you can use the normal distribution.

The sampling distribution is approximately normal because \( n = 50 \geq 30 \).
3 Determine the percentage of values in each tail.

The confidence interval contains 90% of the sample.
The proportion in each tail \( \frac{1 - 0.90}{2} = 0.05 \)

4 Use the inverse cumulative normal distribution to determine the quantiles of the confidence interval.

\[ \Pr(X < \text{lower quantile}) = 0.05 \]
\[ \Pr(X < \text{upper quantile}) = 1 - 0.05 = 0.95 \]

5 Use CAS to determine the quantiles of the confidence interval.
Enter the sample mean as the mean of the sampling distribution.

The 90% confidence interval is (73.84, 76.16)

Using CAS The confidence interval for the population mean

A CAS can be used to find the confidence interval for the population mean from a sample of a given size, such as for Worked example 13 above.

**TI-Nspire CAS**

**STEP 1**

On a calculator page, press \( \text{menu} \), 6: Statistics, 6: Confidence Intervals, 1: z Interval.

Select Stats for the Data Input Method and then tap OK.

**STEP 2**

Complete the dialogue box as shown and press \( \text{enter} \).

The 90% confidence interval is (73.84, 76.16).
Confidence intervals and the margin of error

The confidence interval for the population mean is \( \left( \bar{x} - \frac{z \sigma}{\sqrt{n}}, \bar{x} + \frac{z \sigma}{\sqrt{n}} \right) \) when the population standard deviation \( \sigma \) is known.

When \( \sigma \) is not known, the confidence interval is \( \left( \bar{x} - \frac{s}{\sqrt{n}}, \bar{x} + \frac{s}{\sqrt{n}} \right) \),

where \( s \) is the sample standard deviation, \( \bar{x} \) is the sample mean and \( z \) is the appropriate quantile for the standard normal distribution.

The margin of error is \( M = z \frac{s}{\sqrt{n}} \).

The \( z \)-values vary for different confidence levels and can be calculated from an inverse cumulative standard normal distribution \( Z \sim N(0, 1) \).

The most common confidence levels are 90%, 95% and 99%.

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Proportion in each tail</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>0.005</td>
<td>2.58</td>
</tr>
<tr>
<td>95%</td>
<td>0.025</td>
<td>1.96</td>
</tr>
<tr>
<td>90%</td>
<td>0.05</td>
<td>1.64</td>
</tr>
</tbody>
</table>

For example, the 95% confidence interval for the population mean is \( \left( \bar{x} - 1.96 \frac{s}{\sqrt{n}}, \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right) \).
At an egg farm, the mass of a random sample of 100 eggs is recorded. The sample has a mean of 5.2 g and a standard deviation of 1 g. Find a 95% confidence interval for the mean mass of eggs at the farm (correct to 3 decimal places).

**Working**

1. Estimate the standard error for the sampling distribution using the sample standard deviation \( s \) as an estimate of the population standard deviation \( \sigma \).

   Use \( s = 1 \) to find the standard error of the sample means.

   \[
   S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1}{\sqrt{100}} = \frac{1}{10}
   \]

2. Since \( n \geq 30 \), you can use the normal distribution.

3. Determine the percentage of values in each tail and the \( z \)-value for this proportion.

   The sampling distribution is approximately normal as \( n = 100 \geq 30 \).

   The confidence interval contains 95% of the sample.

   The proportion in each tail = \( \frac{1-0.95}{2} = 0.025 \).

   The \( z \)-value for this proportion is 1.96.

4. Substitute into the formula for the confidence interval.

   \[
   \left( \bar{x} - z \frac{s}{\sqrt{n}}, \bar{x} + z \frac{s}{\sqrt{n}} \right)
   \]

   \[
   = \left( 5.2 - 1.96 \times \frac{1}{10}, 5.2 + 1.96 \times \frac{1}{10} \right)
   \]

   \[
   = (5.004, 5.396)
   \]

   The 95% confidence interval for the mean of eggs at the farm is (5.004, 5.396).

---

A population has a standard deviation of 20. Determine the sample size, \( n \), needed so that the 95% confidence level for the population mean is within 5 of the sample mean.

**Working**

Use the margin of error formula to calculate \( n \).

The margin of error = 5.

Use the population standard deviation 20 for \( s \) and \( z = 1.96 \) for a 95% confidence level.

\[
M = z \frac{s}{\sqrt{n}}
\]

\[
5 = 1.96 \times \frac{20}{\sqrt{n}}
\]

\[
\sqrt{n} = \frac{1.96 \times 20}{5}
\]

\[
n = 61
\]
Confidence intervals for the population mean

Prep 1

WORKED EXAMPLE 12

The time, in minutes, that customers spend at the art gallery is found to be a normally distributed random variable with a mean of 60 minutes and a standard deviation of 15 minutes.

a Find a 95% confidence interval, correct to one decimal place, for this variable.

b Find the 99% confidence interval for the normal variable \( X \) with mean 55 and standard deviation 18.

Prep 2

WORKED EXAMPLE 13

a A random sample of size 100 is taken from a population with a standard deviation of 16. If the mean of the sample is 45, find a 90% confidence interval for the population mean (correct to 2 decimal places).

b The mean of a sample of size 50 is 28. Determine a 95% confidence interval for the population mean if the standard deviation of the population is 9.

Prep 3

WORKED EXAMPLE 14

At an apple orchard, the weight of a random sample of 50 apples is recorded. The sample has a mean of 85 g and a standard deviation of 6 g. Find a 95% confidence interval for the mean weight of apples at the farm (correct to 1 decimal place).

Prep 4

WORKED EXAMPLE 15

a A population has a standard deviation of 30. Determine the sample size needed so that the 95% confidence level of the population mean is within 7 of the sample mean.

b A population has a standard deviation of 20. Determine the sample size needed so that the 90% confidence level of the population mean is within 4 of the sample mean.
Confidence intervals for the population mean

**Question 1**

A normally distributed population has a mean of 62 and a standard deviation of 13. The 95% confidence interval, correct to 2 decimal places, is

A (40.62, 83.38)  
B (35.48, 84.52)  
C (36.00, 88.00)  
D (36.52, 87.48)  
E (49.00, 75.00)

**Question 2**

Samples of size 50 were taken from a normally distributed population with a mean of 55 and a standard deviation of 10. The standard deviation of the sampling distribution is

A \(\frac{\sqrt{2}}{2}\)  
B 5  
C \(\sqrt{2}\)  
D \(\sqrt{10}\)  
E \(\frac{11\sqrt{2}}{2}\)

**Question 3**

In a fishing competition at a large lake, 60 trout are caught and weighed. The trout have a mean weight of 1.1 kg and a standard deviation of 0.2 kg. The 95% confidence interval for the mean weight of trout in the lake is

A (1.05, 1.15)  
B (0.9, 1.3)  
C (0.7, 1.5)  
D (1.06, 1.14)  
E (0.95, 1.25)

**Question 4**

A population has a standard deviation of 32. Determine the sample size needed such that the 95% confidence level of the population mean is within 4 of the sample mean.

A 125  
B 246  
C 61  
D 50  
E 342

**Question 5**

A random sample of 100 is taken from a population. The sample is normally distributed with a mean of 18 and a standard deviation of 3. The margin of error in a 95% confidence interval is

A 0.059  
B 0.54  
C 3.528  
D 0.424  
E 0.588

**Question 6**

At HyperMart Hardware, the amount of money spent by each customer on the weekend is a random variable with a standard deviation of $45. The mean of a random selection of 50 customer receipts is $125. Find the 90% confidence interval for the average amount of money spent by all the customers on the weekend. Express the quantiles to the nearest dollar.
Question 7

A random sample of 100 solar lights produced by a company were tested to determine the number of hours of illumination after a full day of charging. The mean number of hours was 5.5 with a standard deviation of 0.5. Find a 95% confidence interval for the average number of hours of illumination for all the lights produced by the company.

Question 8

A random sample of 49 Victorian Year 11 students were surveyed to find the number of hours of part-time work they completed each week during a school term. If the sample results have a mean of 5.5 hours and a standard deviation of 1.5 hours, determine the margin of error in a 90% confidence interval correct to two decimal places.

Question 9

The average growth of children between the ages of 2 years and 3 years is thought to be 9 cm with a standard deviation of 1 cm. How many results would need to be obtained to establish this with a margin of error of 2 mm at a 95% confidence level?

Question 10

Measurable human traits are normally distributed. A sample of 16 people are found to have an average height of 170 cm with a standard deviation of 8 cm. What is the 95% confidence level for human height?
Statistical inference is the process of drawing conclusions about a population based on the properties of a sample.

A test given to all Year 7 students in Victoria has a mean score of 30 and a standard deviation of 6. Researchers claim that students who watch video tutorials achieve better than average. A random sample of 50 Year 7 students who watched video tutorials was tested and the mean score on the test is found to be 33. Is this a significant difference or could it be due to chance?

In this section, we will examine how we can test the researchers’ claim.

The null and alternative hypotheses

Hypothesis testing is a systematic way to test a claim about a parameter in the population using data measured in a sample. There are two hypotheses.

**H$_0$**: The null hypothesis

This is a statement about a population parameter (the population mean) that is assumed to be true.

The null hypothesis is the starting point. We test to see if the parameter stated in the null hypothesis is likely to be true.

**H$_1$**: The alternative hypothesis

This statement reflects the claim being made and is the opposite of the null hypothesis. The value of the population parameter is either less than, greater than or not equal to the stated value in the null hypothesis.

The null and alternative hypotheses must be mutually exclusive.
For the example of the Victorian Year 7 students who watch video tutorials.

\( H_0: \mu = 30 \)

The null hypothesis states that the mean test result for Victorian Year 7 students who watched video tutorials will be 30 (that is, 'no effect').

\( H_1: \mu > 30 \)

The alternative hypothesis states the mean test result for Victorian Year 7 students who watched video tutorials will be greater than 30 (that is, 'positive effect'). This hypothesis always relates to the claim being made.

**Worked example 16**

The monthly weight gain of turkeys has a mean of 1.5 kg and a standard deviation of 0.2 kg. A poultry farm claims that their turkeys gain more weight on their new high protein feed. Write the null and alternative hypotheses.

**Working**

1. The null hypothesis is a statement about the population that is assumed to be true.  
   \( H_0: \mu = 1.5 \)

2. The alternative hypothesis is a statement about the population mean that contradicts the null hypothesis.  
   The claim that the turkeys gain more weight indicates that the direction of the alternative hypothesis must be \( \mu > \) the population mean.  
   \( H_1: \mu > 1.5 \)

**Worked example 17**

*Excite* bubble gum claims that its gum holds its flavour for an average of 30 minutes. Write the null and alternative hypotheses.

**Working**

1. The null hypothesis is a statement about the population mean that is assumed to be true.  
   \( H_0: \mu = 30 \)

2. The alternative hypothesis is a statement about the population mean that contradicts the null hypothesis. The alternative hypothesis is \( \mu \neq \) the population mean.  
   \( H_1: \mu \neq 30 \)
The procedure for a hypothesis test

- Identify the population of interest and the parameters: population mean (μ) and population standard deviation (σ).
- Identify the sample and the statistics: sample size (n), sample mean (x̄) and sample standard deviation (s).
- Write the null hypothesis and alternative hypothesis.
- Find the z- and p-values for the sample mean.
- Make a decision to either reject or fail to reject the null hypothesis.
- Write the conclusion.

p-values for hypothesis testing related to the mean

The null hypothesis is the statement being tested. Usually, the null hypothesis is a statement of 'no effect' or 'no difference'.

If the null hypothesis is disproved (rejected), then the alternative hypothesis is 'supported' and can be used to describe the population parameter.

Based on the sample data, the testing method will help determine whether to reject the null hypothesis or fail to reject the null hypothesis. The p-value is used to make that determination.

The p-value is the probability of obtaining the value in the sample or a more extreme value, assuming that the null hypothesis is true.

We use a one-sample z test to compare the mean in one sample to the expected norm, which in our case is the population mean. The z test can only be used when the population is normally distributed or the size of the population is large and the standard deviation is known.

When calculating the p-value with a one-sample z test, a cumulative normal distribution is used.

Mean = the population mean μ, or \( \mu \)

Standard deviation (standard error) \( \sigma_x = \frac{\sigma}{\sqrt{n}} \)

If \( \sigma \) is not known, then use \( s_x = \frac{s}{\sqrt{n}} \) in place of \( \sigma_x \).

Significance levels for making decisions

The decision to reject the null hypothesis is made using the p-value.

The value at which the null hypothesis is rejected is called the significance level and is represented by the Greek letter α.
If the \( p \)-value is less than \( \alpha \), we reject the null hypothesis and support the alternative hypothesis. The low \( p \)-value indicates that the observed data are extremely unlikely if the null hypothesis is true.

If the \( p \)-value is greater than \( \alpha \), we fail to reject the null hypothesis.

Common levels used for \( \alpha \) are 0.05 and 0.01:

- \( p \)-value > 0.05 \( \rightarrow \) the observed difference is ‘not significant’
- \( p \)-value < 0.05 \( \rightarrow \) the observed difference is ‘significant’ (to support the alternative hypothesis)
- \( p \)-value < 0.01 \( \rightarrow \) the observed difference is ‘highly significant’ (to support the alternative hypothesis)

If the significance level \( \alpha \) is not given, use 0.05.

The number of hours of sleep an adult has each night is normally distributed with a mean of 7.7 hours and a standard deviation of 0.5 hours. The number of hours of sleep for 50 randomly selected adults who are snorers produced a mean of 7.5 hours.

Test the hypothesis that snorers sleep less than the general population.

Test at the 5% level of significance.

1. Write the null hypothesis. 
   \( H_0: \mu = 7.7 \). The mean amount of sleep that adult snorers have is 7.7 hours.

2. Write the alternative hypothesis. 
   \( H_1: \mu < 7.7 \). The mean amount of sleep that adult snorers have is less than 7.7 hours.

3. Find the \( p \)-value.
   Assume that \( H_0 \) is true.
   Find the probability of obtaining the sample mean or a value more extreme in the direction of the alternative hypothesis.
   As \( H_1 \) is \( \mu < 7.7 \) and the observed mean in the sample is 7.5, the observed value or a more extreme value is \( \bar{x} < 7.5 \).

4. Complete a one-sample \( z \) test.
   The mean used is the population mean.
   The \( z \) test can be completed using CAS.

\[ p = Pr(\bar{x} < 7.5) \]

\[ \mu = 7.7 \]
### TI-Nspire CAS

Set the data input method to stats.
Complete the dialogue box as shown.

### CLASSPAD

In **[menu]**, tap **13: Statistics**.
Tap Calc, Test, One-Sample Z-test, Variable then tap **NEXT>>**.
Enter the values as shown, then tap **NEXT>>**.

Write the $p$-value.

5. **Make a decision about the null hypothesis for 5% level of significance.**
   - Reject if $p < 0.05$.
   - Fail to reject if $p > 0.05$.

6. **Make a conclusion about the claim.**

   \[
   \Pr(\bar{x} < 7.5) = p = 0.0023
   \]

   The chance of getting a mean less than 7.5 is 0.0023 if $H_0$ is true. Since $p < 0.05$, we reject the null hypothesis and therefore the result is significant.

   The mean amount of sleep that adult snorers have per night is significantly less than 7.7 hours. ($z = -2.828$, $p = 0.0023$, $\bar{x} = 7.5$, $\sigma = 0.5$, $n = 50$)
   There is enough evidence to support the claim that snorers sleep less than the general population.
One-tailed and two-tailed tests

In the previous example, a one-tailed test was performed to test a 'less than' claim, but a two-tailed test is required for a 'not equal to' claim.

One-tailed test
A 'less than' (<) or 'greater than' (>) sign in the alternative hypothesis indicates that a one-tailed test is required.

In this one-tailed test,
- the null hypothesis \( H_0 \): \( \mu = \mu_1 \)
- the alternative hypothesis \( H_1 \): \( \mu < \mu_1 \)
- \( p \)-value = \( \Pr(\bar{X} < x) \)

where \( \bar{x} \) is the sample mean and \( \mu_1 \) is the mean of the population.

Two-tailed test
A two-tailed test looks for any change in a parameter.
The presence of a 'not equal to' sign (\( \neq \)) in the alternative hypothesis indicates that a two-tailed test is required.

In this two-tailed test,
- the null hypothesis \( H_0 \): \( \mu = \mu_1 \)
- the alternative hypothesis \( H_1 \): \( \mu \neq \mu_1 \)
- \( p \)-value = \( 2 \times \Pr(\bar{X} > x) \), for \( \bar{x} > \mu_1 \) or
- \( p \)-value = \( 2 \times \Pr(\bar{X} < x) \), for \( \bar{x} < \mu_1 \)

where \( \bar{x} \) is the sample mean and \( \mu_1 \) is the mean of the population.
A fruit juice company sets their machines up to dispense 350 mL of juice in each bottle. The mean and standard deviation of the volume of juice in each bottle are 350 mL and 5 mL respectively. A consumer group claims that there is not consistently 350 mL of juice in each bottle. They then test a sample of 100 bottles and find the average volume to be 348 mL. Test the consumer group’s claim at the 1% significance level.

**Working**

1. Write the null hypothesis.
   
   \[ H_0 : \mu = 350 \]

2. Write the alternative hypothesis.
   
   The alternative hypothesis is that there is not 350 mL in a bottle.
   
   \[ H_1 : \mu \neq 350 \]

3. Assume that \( H_0 \) is true.
   
   Find the probability of obtaining the sample mean or a value more extreme in the direction of the alternative hypothesis.
   
   As \( H_1 \) is \( \mu \neq 350 \), a two-tailed test is required.
   
   We first calculate the probability of obtaining the observed value of \( \bar{x} = 348 \), or a more extreme value (\( \bar{x} < 348 \)).
   
   \[ p = 2 \times \Pr(\bar{x} < 348) \]

4. Complete a two-sample \( z \) test.
   
   The mean used is the population mean.
   
   The \( p \)-value is found using CAS.
   
   Input the data on a CAS.

5. Make a decision about the null hypothesis.
   
   When \( p \)-value < 0.01, the observed difference is highly significant.

6. Make a conclusion about the claim.
   
   Since the null hypothesis \( H_0 \) is rejected and the alternative hypothesis is \( H_1 : \mu \neq 350 \), we need to decide if \( \mu > 350 \) or \( \mu < 350 \).
   
   The sample mean \( \bar{x} = 348 \) is less than 350, therefore we choose \( \mu < 350 \).
   
   The mean volume of juice in the bottle is significantly less than 350 mL. (\( z = -4 \), \( p = 0.000 \, 063 \), \( \bar{x} = 348 \), \( \sigma = 5 \), \( n = 100 \))
   
   There is enough evidence to support the claim that the average volume of juice in the bottle is less than 350 mL.
Critical values for a given significance level

In hypothesis testing, there are two ways to determine if there is enough evidence to reject or fail to reject the null hypothesis.

The first way is to calculate the p-value and compare it with the level of significance $\alpha$, which we have covered in detail in the previous section.

The second way is to compare the test statistic ($z$) with a critical value determined from the level of significance $\alpha$. If the test statistic is more extreme than the critical value, the null hypothesis is rejected.

When $\sigma$ is known, the $z$-value for the sample statistic is given by

$$ z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} $$

where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

When $\sigma$ is not known, use $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ in place of $\sigma_{\bar{x}}$.

**Exam hack**

Memorise the critical $z$-values for one- and two-tailed tests at the 0.01, 0.05 and 0.1 levels of significance, shown in the above tables. They may be required in exam 1 in questions on hypothesis testing.

The inverse cumulative normal distribution can be used to find the critical $z$-value.

Let the critical $z$-value = $c$.

For a significance level $\alpha = 0.05$ with a right one-tailed test, $Pr(z > c) = 0.05$. 

Using CAS  Finding the critical $z$-value

A CAS can be used to find the critical $z$-value for a one-tailed test with a significance level of 0.05.

**TI-Nspire CAS**


The area is the proportion below the critical value.

Area = $1 - 0.05 = 0.95$

For the standard normal variable $z$, the mean $\mu = 0$, the standard deviation $\sigma = 1$.

\[
Pr(z > c) = 0.05
\]

$c = 1.64485$

**ClassPad**

On the main menu $\sqrt{\text{Menu}}$, tap Interactive, Distributions, Inverse, InvNormCDF.

Select ‘Right’ for Tail setting and type in 0.05 for prob, then tap OK.

For the standard normal variable $z$, the mean $\mu = 0$, the standard deviation $\sigma = 1$.

\[
Pr(z > c) = 0.05
\]

$c = 1.64485$
Yogood sells yoghurt in containers with an advertised mean volume of 250mL. Their competitors claim that on average the containers contain less yoghurt than advertised. They take a random sample of 16 containers and find the mean volume is 248 mL. The volume of yoghurt in all containers is normally distributed with a standard deviation of 20 mL.

**a** State appropriate null and alternative hypotheses for the volume.

**b** The $p$-value for this test is given by the expression $\Pr(Z < z)$, where $Z$ has the standard normal distribution. Find the value of $z$ and hence determine whether the null hypothesis should be rejected at the 0.05 level of significance.

**Working**

**a** 1 Write the null hypothesis. 
$H_0: \mu = 250$

2 Write the alternative hypothesis. 
$H_1: \mu < 250$

**b** 1 Find the value of the test statistic $z$. 
Use $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ where $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Substitute $\mu = 250$, $\sigma_{\bar{x}} = 5$

\[
z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{248 - 250}{5} = -0.4
\]

2 Compare the test statistic with the critical $z$-value for the significance level of 0.05. 
For $\alpha = 0.05$, the critical $z$-value = −1.645 and the test statistic $z$ is −0.4.

3 Make a decision about the null hypothesis. 
The test statistic $z = -0.4$ is not less than the critical $z$-value = −1.645 and lies outside the rejection region.

We fail to reject the null hypothesis $\mu = 250$ mL at the 0.05 significance level.

$(z = -0.4, \bar{x} = 248, \sigma = 20, n = 16)$
A normally distributed population has a mean of 45 and a standard deviation of 5.

Consider the hypotheses $H_0: \mu = 45$, $H_1: \mu \neq 45$.

a. Find the critical $z$-values at the 0.04 significance level, correct to three decimal places.

b. A random sample of size 25 is taken from this population. The random variable $\bar{X}$ represents the sampling distribution of the sample means. Find the critical values of $\bar{X}$ at which the null hypothesis would be rejected.

**Working**

**A**

1. Draw the standard normal distribution, label the critical $z$-values and the associated probabilities.

A two-tailed test is required for $\mu \neq 45$.

2. Use an inverse cumulative normal distribution on a CAS to find $z$.

In the two-tailed test, we reject the null hypothesis $\mu = 45$ if the test statistic is in the rejection region shaded in the standard normal distribution.

**B**

1. Find the value of $\sigma_{\bar{X}}$.

2. $\bar{X}$ is the sampling distribution of the sample means of size 25 with a mean of 45 and a standard error of 1.

The value of $c$ corresponds to the critical $z$-value of 2.054.

Substitute into $z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$.

The value of $d$ corresponds to the critical $z$ value of $-2.054$.

Substitute into $z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$.

Reject the null hypothesis if the sample mean is less than 42.946 or greater than 47.054.
Hypothesis testing related to the mean

**Prep 1**

**WORKED EXAMPLE 16**

a. The average time spent waiting at a medical centre is 40 minutes. The centre introduces a new system of warning patients by SMS in advance if the doctors are running late and they claim that this has reduced the average waiting time. Write the null and alternative hypotheses.

b. The average mark for a secondary school entrance exam is 75. A private company runs entrance exam preparation courses and they claim that students who do their courses perform on average better on the entrance exam. Write the null and alternative hypotheses.

**Prep 2**

**WORKED EXAMPLE 17**

a. The average mass of Milko’s chocolate bar is 250 g. The factory purchased a new machine and the supervisor claims the mean mass of the chocolate bars has changed. Write the null and alternative hypotheses.

b. The residents of a small town want to evaluate changes in the pH of the water in a river that runs behind their homes. For the last 35 years, the average pH has been 5.5, however over the last few years, a lot of construction has been occurring in the area and the residents believe the mean pH of the river has changed. Write the null and alternative hypotheses.

**Prep 3**

**WORKED EXAMPLE 18**

Globalite produces light bulbs that, when tested, have a life with a mean of 3000 hours and a standard deviation of 500 hours. A large retail store that uses the bulbs believe they burn on average for less than 3000 hours. The store tests a random sample of 100 bulbs and finds the mean of the sample is 2800 hours.

a. Write the null and alternative hypotheses.

b. Calculate the $z$- and $p$-values.

c. What is your conclusion about the retailer’s claim?

**Prep 4**

**WORKED EXAMPLE 19**

A manufacturer produces drill bits with a mean life of 580 hours and a standard deviation of 30 hours. A construction company that uses the drill bits disputes the manufacturer’s claim that the bits will last an average of 580 hours. The company draws a sample of 100 bits and finds the mean is 577 hours.

a. Write the null and alternative hypotheses.

b. Calculate the $z$- and $p$-values.

c. What is your conclusion about the construction company’s claim?
Pies 'R Squared makes meat pies with a mean weight of 175 g. A consumer organisation claims that on average these pies are smaller than their advertised weight. It takes a random sample of 25 pies and finds that its mean weight is 170 g. The weight of all meat pies produced by Pies 'R Squared is normally distributed with a standard deviation of 10 g.

a State appropriate null and alternative hypotheses for the weight.

b The $p$-value for this test is given by the expression $\Pr(Z < z)$, where $Z$ has the standard normal distribution. Find the value of $z$ and hence determine whether the null hypothesis should be rejected at the 0.05 level of significance.

A normally distributed population has a mean of 80 and a standard deviation of 10.
Consider the hypotheses $H_0: \mu = 80$, $H_1: \mu \neq 80$.

a Find the critical $z$-values at the 0.02 significance level correct to three decimal places.

b A random sample of size 25 is taken from this population. The random variable $\bar{X}$ represents the sampling distribution of the sample means. Find the critical values of $\bar{X}$ at which the null hypothesis would be rejected.

Hypothesis testing related to the mean

The following information refers to Questions 1 & 2.
The mass of redfin in a large lake is normally distributed with a mean of 1.5 kg and a standard deviation of 0.4 kg. An angling club believes that pollution in the lake has reduced the size of the fish. They take a random sample of 80 fish and find the average mass of the fish is 1.4 kg.

**Question 1**
The null hypothesis is

A $H_0: \mu = 1.4$

D $H_0: \mu = 1.5$

**Question 2**
The alternative hypothesis is

D $H_1: \mu < 1.4$

E $H_1: \mu \neq 1.5$
The following information refers to Questions 3 & 4.
The drying time of a certain brand of paint has a mean of 90 minutes and a standard deviation of 10 minutes. A painting contractor claims the addition of an agent will reduce the average drying time. The drying time of 60 modified paint tins was found to have a mean of 87 minutes.

Question 3
The null and alternative hypotheses are
A  \( H_0: \mu < 90, H_1: \mu = 90 \)  \hspace{1cm} B  \( H_0: \mu = 90, H_1: \mu = 87 \)  \hspace{1cm} C  \( H_0: \mu = 90, H_1: \mu < 87 \)
D  \( H_0: \mu = 87, H_1: \mu < 87 \)  \hspace{1cm} E  \( H_0: \mu = 90, H_1: \mu < 90 \)

Question 4
The \( p \)-value for the hypothesis test is
A  0.0101  \hspace{1cm} B  0.0011  \hspace{1cm} C  0.0201  \hspace{1cm} D  0.05  \hspace{1cm} E  0.00023

The following information refers to Questions 5 & 6.
The time taken for the express train to travel between two major cities is a continuous random variable with a mean of 95 minutes and a standard deviation of 4 minutes. The transport authority makes some modifications to the crossings on the route and would like to know if the travel time has changed. They take a random sample of 50 trips and find the mean travel time is 93 minutes.

Question 5
The null and alternative hypotheses are
A  \( H_0: \mu < 95, H_1: \mu = 93 \)  \hspace{1cm} B  \( H_0: \mu = 93, H_1: \mu = 93 \)  \hspace{1cm} C  \( H_0: \mu = 95, H_1: \mu < 93 \)
D  \( H_0: \mu = 95, H_1: \mu 
eq 95 \)  \hspace{1cm} E  \( H_0: \mu = 95, H_1: \mu < 95 \)

Question 6
The \( p \)-value for the hypothesis is
A  0.0015  \hspace{1cm} B  0.0092  \hspace{1cm} C  0.0184  \hspace{1cm} D  0.0001  \hspace{1cm} E  0.0231

Question 7
A health centre finds that the weight loss for people who complete their ten-week program is normally distributed with a mean of 0.65 kg per week and a standard deviation of 0.05 kg per week. A group of 45 randomly selected people who complete the ten-week program also use a diet app and they find their average weight loss is 0.67 kg per week. They claim the use of the app improves their weight loss.

a  Write the null and alternative hypotheses.
b  Calculate the \( z \) and \( p \)-values.
c  State the decision.
d  What is your conclusion about the claim made about the effectiveness of the app?
Question 8

The resting pulse rate of primary school-age children taken from data over a 30-year period has a mean of 65 beats per minute and a standard deviation of 10 beats per minute.

The medical association believes that changes in lifestyle have affected the rest pulse of children. They take a random sample of 100 primary school-age children and find their average rest pulse is 68 beats per minute.

a Write the null and alternative hypotheses.

b Calculate the $z$- and $p$-values.

c State the decision.

d What is your conclusion?

Question 9

The scores on an IQ test are normally distributed with a mean of 100 and a standard deviation of 15. An organisation claims that their courses will on average improve candidate's scores on an IQ test. They take a random sample of 40 people who have completed the course and find their average IQ is 101. Test the claim made by the organisation.

Question 10

A manufacturer produces pistons that have a diameter that are normally distributed with a mean of 750 mm and a standard deviation of 0.5 mm. A mechanic believes that the pistons supplied to him do not regularly have a diameter of 750 mm, so he measures 20 randomly selected pistons and finds their average diameter is 750.2 mm. Test the claim made by the mechanic.
No hypothesis test is 100% certain. As the test is based on probabilities, there is always a chance of making an error. There are two types of errors.

A **type I error** occurs when we reject the null hypothesis when it is true. The probability of making this type of error is $\alpha$, as this is the level at which the null hypothesis is rejected.

A **type II error** occurs when we fail to reject the null hypothesis when it is false.

A type I error is also called a *false positive* and a type II error a *false negative*.
A car that uses unleaded petrol has a fuel consumption of 10 L/100 km. A fuel company claims that their additive will improve the fuel consumption.

a Write the null and alternative hypotheses.

b Describe a type I error and its impact in this situation.

c Describe a type II error and its impact in this situation.

---

**Working**

1 Write the null and alternative hypotheses.

   \( H_0: \mu = 10 \text{ L/100 km} \)
   
   The additive will not improve the fuel consumption.

   \( H_1: \mu < 10 \text{ L/100 km} \)
   
   The additive will improve the fuel consumption.

2 A type I error occurs when we reject the null hypothesis when it is true.

   A type I error occurs if \( H_0: \mu = 10 \text{ L/100 km} \) is rejected when it is true. This would mean that you conclude the additive would improve the fuel consumption when it really does not. Consumers will therefore be wasting their money on the product.

3 A type II error occurs when we fail to reject the null hypothesis when it is false.

   A type II error occurs if the null hypothesis is not rejected when it is false. This would mean that you conclude the additive would have not improved the fuel consumption in the car when in fact it would. Therefore, consumers are missing out on an effective product.
Errors in hypothesis testing

Prep 1

A potato farmer supplies potatoes for making chips that have a mean length of 12 cm with a standard deviation of 1 cm. An inspector takes a random sample of potatoes from the truck and measures the length of the potatoes. The chip company will reject the truck load of potatoes if the average length is less than 11.5 cm.

i Write the null and alternative hypotheses.

ii Describe a type I error and its impact in this situation.

iii Describe a type II error and its impact in this situation.

b In Brisbane, the response time for an ambulance has a mean of 13.5 minutes and a standard deviation of 3.2 minutes. The director of emergency response initiates some changes to the call centre to improve the ambulance response.

i Write the null and alternative hypotheses.

ii Describe a type I error and its impact in this situation.

iii Describe a type II error and its impact in this situation.

Question 1

A type I error can occur

A when the null hypothesis is true but is not rejected
B when the alternative hypothesis is false but is not rejected
C when the null hypothesis is true and the alternative hypothesis is false
D when the null hypothesis is true but is rejected
E when null hypothesis is false but is rejected.

Question 2

A type II error can occur

A when the null hypothesis is false but is not rejected
B when the alternative hypothesis is false but is not rejected
C when the null hypothesis is false and the alternative hypothesis is true
D when the null hypothesis is true but is rejected
E when the null hypothesis is false but is rejected.
Question 3

A tyre company produces a steel belted radial tyre designed for normal road use. The life of the tyre in kilometres has a mean of 40 000 and a standard deviation of 5000. The RACV claims that the mean life of the tyres will be improved if the maximum speed is reduced from 110 km/h to 90 km/h. They select a random sample of 100 drivers who reduce their maximum speed to 90 km/h and find the mean life of their tyres is 41 000 km.

a  State the null and alternative hypotheses.
b  Find the \( z \)- and \( p \)-values.
c  Describe a type I error and its impact in this situation.
d  Describe a type II error and its impact in this situation.
e  What is the decision?
f  What is the conclusion about the claim made by the RACV?
g  What error do you run the risk of making?

Question 4

The battery life of a brand of smartphone is normally distributed with a mean of 7 hours and a standard deviation of 1 hour. The phone's software receives an update that claims to improve the battery life of the phone. A random sample of 50 people record the battery life of the phone and the average is found to be 7.5 hours.

a  State the null and alternative hypotheses.
b  Find the \( z \)- and \( p \)-values.
c  Describe a type I error and its impact in this situation.
d  Describe a type II error and its impact in this situation.
e  What is the conclusion about the claim made about the phone's software update?
Random variables and statistical inference

Linear combinations of random variables

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(X) = \sum x \cdot p(x) )</td>
<td>( E(X) = \int_{-\infty}^{\infty} x \cdot f(x) , dx )</td>
</tr>
<tr>
<td>( E(X^2) = \sum x^2 \cdot p(x) )</td>
<td>( E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) , dx )</td>
</tr>
</tbody>
</table>

A random variable \( Y \) is a linear transformation of \( X \) if \( Y = aX + b \).

\[ E(aX + b) = a \cdot E(X) + b \]

- The formulas for the variance and standard deviation are:
  \[ \text{Var}(X) = E(X^2) - \mu^2 \]
  \[ \text{SD}(X) = \sqrt{\text{Var}(X)} \]
  where \( \mu = E(X) \)

- For two independent random variables \( X \) and \( Y \),
  \[ E(aX + bY) = aE(X) + bE(Y) \]
  \[ \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) \]

Independent normal random variables

- If \( X \) is normally distributed with a mean of \( \mu_X \) and a variance of \( (\sigma_X)^2 \), and \( Y \) is an independent normally distributed random variable with a mean of \( \mu_Y \) and a variance of \( (\sigma_Y)^2 \), then \( aX + bY \) is also normally distributed with
  \[ \text{Mean:} \quad E(aX + bY) = a\mu_X + b\mu_Y \]
  \[ \text{Variance:} \quad \text{Var}(aX + bY) = a^2(\sigma_X)^2 + b^2(\sigma_Y)^2 \]

The sampling distribution of the sample means

- The sampling distribution of the sample means contains the means \( (\bar{x}) \) of all possible samples of size \( n \) from a population.

- The mean of the sample means \( (\mu_\bar{x}) \) is equal to the mean of the population \( (\mu) \):
  \[ \mu_\bar{x} = \mu \]

- The standard deviation of the sample means is the standard error of the sample mean.
• A larger sample size produces less error and a more reliable estimate of the population mean: the sample mean is closer to the population mean and there is less variability in the means.

• The standard error indicates how much the sample mean \( \bar{x} \), for a particular sample size, deviates from the population mean \( \mu \).

• The standard error for a sample mean is \( \text{SE}(\bar{x}) = \frac{\sigma}{\sqrt{n}} \), where \( \sigma \) is the population standard deviation and \( n \) is the sample size.

Confidence intervals for the population mean

• A **point estimate** of a parameter is a single value obtained from a sample. The statistic used is called an **estimator** of the parameter.

• An **interval estimate** of a parameter is an interval that is likely to include the value of the parameter. An interval estimate for a population parameter is called a **confidence interval**.

For a confidence interval symmetrical about the mean in a distribution:

• the **confidence level** is equivalent to the probability that the population parameter being estimated lies within the confidence interval

• the **margin of error** is the distance of the ends of the interval from the mean. It is half of the confidence interval.

The central limit theorem

• If \( X \) is a random variable with mean \( \mu \) and standard deviation \( \sigma \), then the sampling distribution of sample means \( \bar{X} \), of size \( n \), will be approximately normal with mean \( E(\bar{X}) \) and standard deviation \( \text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} \), provided \( n \) is sufficiently large (\( n \geq 30 \)).

• For a sample with mean \( \bar{x} \) and standard deviation \( s \), \( \bar{x} \) is an estimator of \( E(\bar{X}) \).

  When \( \sigma \) is not known, \( s \) is used as an estimator of \( \sigma \).

  The standard deviation of the sampling distribution is \( s_{\bar{x}} = \frac{s}{\sqrt{n}} \).

Confidence intervals and the margin of error

• The **confidence interval** for the population mean is \( \left[ \bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right] \) when the population standard deviation \( \sigma \) is known.

• When \( \sigma \) is not known, the confidence interval is \( \left[ \bar{x} - z \frac{s}{\sqrt{n}}, \bar{x} + z \frac{s}{\sqrt{n}} \right] \), where \( z \) is the sample standard deviation, \( \bar{x} \) is the sample mean and \( z \) is the appropriate quantile for the standard normal distribution.

• The **margin of error** is \( M = z \frac{s}{\sqrt{n}} \).
**Hypothesis testing**

- Hypothesis testing is a systematic way to test a claim about a parameter in the population using data measured in a sample.

- \( H_0: \) The null hypothesis is a statement about a population parameter (the population mean) that is assumed to be true.

  We test to see if the parameter stated in the null hypothesis is likely to be true.

- \( H_1: \) The alternative hypothesis is a statement that reflects the claim being made and is the opposite of the null hypothesis. It states that the value of the population parameter is less than, greater than or not equal to the stated value in the null hypothesis.

- The **p-value** is the probability of obtaining the value in the sample or a more extreme value, assuming that the null hypothesis is true.

- When calculating the **p-value** with a one-sample \( z \) test, a cumulative normal distribution is used. Mean = the population mean \( \mu \)

  Standard deviation \( \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \)

- \( \alpha \) is the **significance level**, the value (usually 0.05) at which the null hypothesis is rejected.

- If the **p-value** is less than \( \alpha \), we reject the null hypothesis and support the alternative hypothesis.

- If the **p-value** is greater than \( \alpha \), we fail to reject the null hypothesis.

**One-tailed test**

A < or > sign in the alternative hypothesis indicates that a one-tailed test is required:

- the null hypothesis \( H_0: \mu = \mu_1 \)

- the alternative hypothesis \( H_1: \mu < \mu_1 \)

  \[ p-value = Pr(\bar{X} < \bar{x}) \]

where \( \bar{x} \) is the sample mean and \( \mu_1 \) is the mean of the population.

**Two-tailed test**

A ≠ sign in the alternative hypothesis indicates a two-tailed test is required:

- the null hypothesis \( H_0: \mu = \mu_1 \)

- the alternative hypothesis \( H_1: \mu \neq \mu_1 \)

  \[ p-value = 2 \times Pr(\bar{X} > \bar{x}), \text{ for } \bar{x} > \mu_1 \quad \text{or} \]

  \[ p-value = 2 \times Pr(\bar{X} < \bar{x}), \text{ for } \bar{x} < \mu_1 \]

where \( \bar{x} \) is the sample mean and \( \mu_1 \) is the mean of the population.
Critical values for a given significance level

A null hypothesis can also be tested by comparing a test statistic \( z \) with a critical value determined from the level of significance \( \alpha \).

If the test statistic is more extreme than the critical value, the null hypothesis is rejected.

When \( \sigma \) is known, the \( z \)-value for the sample statistic is given by

\[
\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = z
\]

where

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

When \( \sigma \) is not known, use \( s_{\bar{x}} = \frac{s}{\sqrt{n}} \) in place of \( \sigma_{\bar{x}} \).

Errors in hypothesis testing

- A **type I error** occurs when we reject the null hypothesis when it is true. The probability of making this type of error is \( \alpha \) as this is the level at which the null hypothesis is rejected.

- A **type II error** occurs when we fail to reject the null hypothesis when it is false.