Math Notes

Chang Si Yuan

November 15, 2017

(ver 1.0)

This compilation of notes are to be used as a reference for the GCE "A"-level Mathematics paper, both as a refresher in theories as well as for general descriptions of presentation form. These notes are meant for free, public use, but at the reader's own risk.
Good luck with your exams.

1 Assumed Knowledge

1.1 Algebra

1.1.1 Completing the Square

$$x^2 + bx + c = (x + \frac{b}{2})^2 + c - (\frac{b}{2})^2$$

1.1.2 Polynomial Expansions

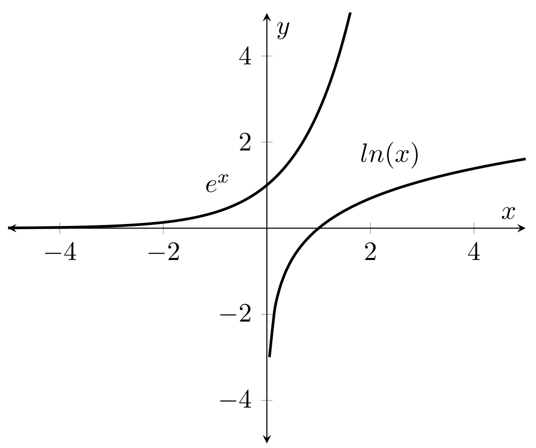
$$(a \pm b)^2 = a^2 \pm 2ab + b^2$$
$$a^2 - b^2 = (a+b)(a-b)$$
$$a^3 \pm b^3 = (a \pm b)(a^2 \mp ab + b^2)$$

1.1.3 Partal Fractions

$$\frac{f(x)}{(ax+b)(cx+d)}$$
$$= g(x) + \frac{A}{ax+b} + \frac{B}{cx+d}$$
$$\frac{f(x)}{(ax+b)(cx+d)^2}$$
$$= g(x) + \frac{A}{ax+b} + \frac{B}{cx+d} + \frac{C}{(cx+d)^2}$$
$$\frac{f(x)}{(ax+b)(x^2+c)}$$
$$= g(x) + \frac{A}{ax+b} + \frac{Bx+C}{x^2+c}$$

1.1.4 Exponent and Logarithm

$$e^n = \underbrace{e \times e \times e \times ... \times e}_{\text{n times}}$$
$$e^{\frac{1}{2}} = \sqrt{e}$$
$$\log_e(x) = \ln(x)$$
$= $ how many times e is multiplied by itself to get x
$$\log_{10}(x) = \lg(x)$$
$$x = e^{ln(x)}$$
$$\log_x(y) = \frac{\log_{base}(y)}{\log_{base}(x)}$$



1.2 Trigonometry

1.2.1 Sine and Cosine Rule

For any triangle with length of sides a, b and c and with opposite angles A B and C:

$$\frac{a}{\sin(A)} = \frac{b}{\sin(B)} = \frac{c}{\sin(C)}$$
$$a^2 = b^2 + c^2 - 2bc\cos(A)$$
$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}$$

1.2.2 Sum of Angles

$$\sin(A \pm B) = \sin(A)\cos(B) \pm \cos(A)\sin(B)$$
$$\sin(2A) = 2\sin(A)\cos(A)$$
$$\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$$
$$\begin{aligned}\cos(2A) &= \cos^2(A) - \sin^2(A)\\ &= 2\cos^2(A) - 1\\ &= 1 - 2\sin^2(A)\end{aligned}$$
$$\tan(A \pm B) = \frac{\tan(A) \pm \tan(B)}{1 \mp \tan(A)\tan(B)}$$
$$\tan(2A) = \frac{2\tan(A)}{1 - \tan^2(A)}$$
Area of Triangle $= \frac{1}{2}absin(C)$

1.2.3 Factor and Reverse Factor Formula

$$\sin(A) + \sin(B) = 2\sin(\tfrac{A+B}{2})\cos(\tfrac{A-B}{2})$$
$$\sin(A) - \sin(B) = 2\cos(\tfrac{A+B}{2})\sin(\tfrac{A-B}{2})$$
$$\cos(A) + \cos(B) = 2\cos(\tfrac{A+B}{2})\cos(\tfrac{A-B}{2})$$
$$\cos(A) - \cos(B) = -2\sin(\tfrac{A+B}{2})\sin(\tfrac{A-B}{2})$$
$$\sin(A)\cos(B) = \tfrac{1}{2}[\sin(A+B) + \sin(A-B)]$$
$$\cos(A)\sin(B) = \tfrac{1}{2}[\sin(A+B) - \sin(A-B)]$$
$$\cos(A)\cos(B) = \tfrac{1}{2}[\cos(A+B) + \cos(A-B)]$$
$$\sin(A)\sin(B) = -\tfrac{1}{2}[\cos(A+B) - \cos(A-B)]$$

Factor formulae are given in MF10. Reverse factor formula can be derived using factor formula.

$$\sin(A) + \sin(B) = 2\sin(\tfrac{A+B}{2})\cos(\tfrac{A-B}{2})$$

$$\tfrac{1}{2}[\sin(X+Y) + \sin(X-Y)] = \sin(X)\cos(Y)$$

2 Inequalities

2.1 Inequalities

2.1.1 Properties of Inequalities

$$a > b, c > 0 \implies ac > bc$$
$$a > b, c < 0 \implies ac < bc$$
$$\frac{a}{b} > 0 \implies ab > 0$$

$$\frac{a}{b} < 0 \implies ab < 0$$

Positive sides of inequalities suggest that both terms share similar positive or negative signs, negative sides of inequalities suggest that both terms have opposite positive or negative signs.

2.1.2 Quardratic Inequalities

Find where $f(x) = 0$ by completing square or quadratic formula and sketch graph.

2.1.3 Inequality Reduction

For any inequality $\frac{f(x)}{g(x)} > or < 0$ where $f(x)$ or $g(x)$ is strictly positive or negative, reduce inequality to non-strictly positive/negative function and change sign accordingly.
Careful for elements of the form $(x+a)^2$, though these can be assumed to be strictly positive, the case where $(x+a)^2 = 0$ needs to be accounted for.

2.1.4 Modulus Inequalities

$$|x| < a \iff -a < x < a$$
$$|x| > a \iff x < -a \quad or \quad a < x$$
$$|x-a| < b \iff a-b < x < a+b$$
$$|x-a| > b \iff x < a-b \quad or \quad a+b < x$$

To solve inequalities, sketch and find intercept, then deduce suitable range of x.

3 Vectors

3.1 Representation

3.1.1 Point Representation

O is always defined as the origin
Written: $\underset{\sim}{r}$ or $\overrightarrow{OR}$
Column: $\mathbf{r} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$
Unit Vector: $\mathbf{r} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ where $\mathbf{i}$,$\mathbf{j}$ and $\mathbf{k}$ are unit vectors in the x, y and z dimensions

3.1.2 Line Representation

Vector Equation:

$$l : \mathbf{r} = \mathbf{a} + \lambda\mathbf{b} \quad \lambda \in \mathbb{R}$$

Parametric Equation:

$$l : \begin{cases} x = a_x + \lambda b_x \\ y = a_y + \lambda b_y \\ z = a_z + \lambda b_z \end{cases} \quad \lambda \in \mathbb{R}$$

Cartesian Equation:

$$\frac{x-a_x}{b_x} = \frac{y-a_y}{b_y} = \frac{z-a_z}{b_z}$$

3.1.3 Plane Representation

Vector Equation:

$$\pi : \mathbf{r} = \mathbf{a} + \lambda\mathbf{b} + \mu\mathbf{c}$$

Scalar Product Equation:

$$\pi : \mathbf{r} \cdot \mathbf{n} = d$$

Cartesian Equation:

$$\pi : xn_x + yn_y + zn_z = d$$

3.2 Manipulation

3.2.1 Vector Algebra

Vector Addition: remove same inside or outside terms

$$\overrightarrow{OA} + \overrightarrow{AB} = \overrightarrow{OB}$$

Negative Vectors: reverse the points

$$\overrightarrow{AB} = -\overrightarrow{BA}$$

Vector Subtraction: reverse points, then add

$$\overrightarrow{OB} - \overrightarrow{OA} = \overrightarrow{OB} + \overrightarrow{AO} = \overrightarrow{AB}$$

3.2.2 Vector Properties

Modulus / Magnitude : The total distance of a vector

$$\left| \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right| = \sqrt{x^2+y^2+z^2}$$

Unit Vector / Cap : A vector which defines a direction and has modulus of 1

$$\hat{\mathbf{a}} = \frac{1}{|\mathbf{a}|}\mathbf{a}$$

3.2.3 Ratio Theorem

Points between two direction vectors $\mathbf{a}$ and $\mathbf{b}$ are in the form:

$$\mathbf{r} = \frac{\mu\mathbf{a} + \lambda\mathbf{b}}{\mu + \lambda}$$

3.2.4 Scalar Product

Scalar/Dot product produces a scalar which is a representation of how inline two vectors are with each other.

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = x_1x_2 + y_1y_2 + z_1z_2$$

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}|cos(\theta)$$
$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$$
$$\mathbf{a} \cdot (\perp \mathbf{a}) = 0$$
$$\mathbf{a} \cdot (\mathbf{b} \pm \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$$
$$(\lambda\mathbf{a}) \cdot \mathbf{b} = \lambda(\mathbf{a} \cdot \mathbf{b})$$

3.2.5 Vector Product

Vector/Cross product produces a vector which has direction perpendicular to its input vectors and has magnitude similar to area subtended by its input vectors.

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \times \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} y_1 z_2 - z_1 y_2 \\ z_1 x_2 - x_1 z_2 \\ x_1 y_2 - y_1 x_2 \end{pmatrix}$$

$$\mathbf{a} \times \mathbf{b} = \hat{\mathbf{n}}|\mathbf{a}||\mathbf{b}|sin(\theta)$$

$$\mathbf{a} \times \mathbf{a} = 0$$

$$\mathbf{a} \times (\perp \mathbf{a}) = \hat{\mathbf{n}}|\mathbf{a}||\mathbf{b}|$$

$$\mathbf{a} \times (\mathbf{b} \pm \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$$

$$(\lambda \mathbf{a}) \times \mathbf{b} = \lambda(\mathbf{a} \times \mathbf{b})$$

3.3 Angles Between Vectors

General formula:

$$cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

Lesser used formula:

$$sin(\theta) = \frac{|\mathbf{a} \times \mathbf{b}|}{|\mathbf{a}||\mathbf{b}|}$$

3.3.1 Point-Point Angle

Let $\mathbf{a}$ and $\mathbf{b}$ be point vectors.

3.3.2 Line-Line Angle

Let $\mathbf{a}$ and $\mathbf{b}$ be direction vectors of lines.

3.3.3 Line-Plane Angle

Let $\mathbf{a}$ be direction vector of line and $\mathbf{b}$ be normal vector of plane
Angle between line and plane will be $90° - \theta$.

3.3.4 Plane-Plane Angle

Let $\mathbf{a}$ and $\mathbf{b}$ be normal vectors of planes.

3.4 Intersection

3.4.1 Point-Line Intersection

Solve series of parametric equations or find λ which lets point equal to point on line.

3.4.2 Line-Line Intersection

If direction vectors are scalar multiples of each other, lines are parallel.
Find a point which satisfies both lines, solving parametric equations of both line equations.

$$\mathbf{a_1} + \lambda \mathbf{b_1} = \mathbf{a_2} + \mu \mathbf{b_2} \quad \lambda, \mu \in \mathbb{R}$$

If lines are both non-parallel and non-intersecting, lines are skew.

3.4.3 Line-Plane Intersection

If dot product of direction vector of line and normal of plane equals to 0, line is parallel to plane.
Substitute line equation into plane equation and expand to solve for λ.

$$\begin{aligned} p &= (\mathbf{a} + \lambda \mathbf{b}) \cdot \mathbf{n} \\ &= \mathbf{a} \cdot \mathbf{n} + \lambda(\mathbf{b} \cdot \mathbf{n}) \end{aligned}$$

3.4.4 Plane-Plane Intersection

If normal of planes are scalar multiples of each other, planes are parallel.
Equating two planes results in a line.
Cross product of normal vectors of two planes produces the direction vector of line.
Position vector of line can be observed from equations, find a vector which satisfies both plane equations.

3.5 Projections

3.5.1 Point-Point Projection

To find distance d of projection of point vector $\mathbf{a}$ on point vector $\mathbf{b}$:

$$d = \mathbf{a} \cdot \hat{\mathbf{b}}$$

3.5.2 Point-Line Projection

To find distance d of projection of point vector $\mathbf{a}$ on direction vector of line $\mathbf{b}$, similar to Point-Point Projection.

3.5.3 Point-Line Perpendicular

To find perpendicular of point vector $\mathbf{a}$ on line $l : \mathbf{r} = \mathbf{b} + \lambda \mathbf{c}$, let $\mathbf{P}$ be a point on the line such that $\overrightarrow{AP}$ is perpendicular to $\mathbf{c}$ and solve for λ :

$$\begin{aligned} 0 &= \overrightarrow{AP} \cdot \mathbf{c} \\ &= (\mathbf{b} + \lambda \mathbf{c} - \mathbf{a}) \cdot \mathbf{c} \\ &= \lambda \mathbf{c} \cdot \mathbf{c} + (\mathbf{b} - \mathbf{a}) \cdot \mathbf{c} \end{aligned} \qquad \lambda |\mathbf{c}|^2 = (\mathbf{a} - \mathbf{b}) \cdot \mathbf{c}$$

To find distance d between point vector $\mathbf{a}$ and its perpendicular on line with direction vector $\mathbf{b}$, find magnitude of the cross product of $\mathbf{a}$ and unit vector of $\mathbf{b}$:

$$d = |\mathbf{a} \times \hat{\mathbf{b}}|$$

3.5.4 Point-Plane Perpendicular

To find perpendicular of point vector $\mathbf{a}$ on plane $\pi : \mathbf{r} \cdot \mathbf{n} = d$, consider a line containing $\mathbf{a}$ and with direction vector $\mathbf{n}$, equate the two equations and then solve for λ:

$$l = \mathbf{a} + \lambda \mathbf{n}$$

$$\begin{aligned} d &= (\mathbf{a} + \lambda \mathbf{n}) \cdot \mathbf{n} \\ &= \mathbf{a} \cdot \mathbf{n} + \lambda \mathbf{n} \cdot \mathbf{n} \end{aligned} \qquad \lambda |\mathbf{n}|^2 = d - \mathbf{a} \cdot \mathbf{n}$$

To find distance d between point vector $\mathbf{a}$ and plane with normal vector $\mathbf{n}$, find projection of $\mathbf{a}$ on unit vector of $\mathbf{n}$:

$$d = |\mathbf{a} \cdot \hat{\mathbf{n}}|$$

4 Complex Numbers

4.1 Imaginary Numbers

$$i = \sqrt{-1}$$
$$i^2 = -1$$
$$i^4 = i^2 \times i^2 = 1$$

4.2 Cartesian Representation

4.2.1 Complex Numbers

Complex numbers are written in the form:

$$z = a + ib \quad a, b \in \mathbb{R}$$
$$\text{Where} \quad Re(z) = a \quad Im(z) = b$$

And populate the set $\mathbb{C}$

4.2.2 Conjugates

$$\text{For} \quad w = a + ib \quad z = c + id$$
$$w^* = a - ib$$
$$ww^* = a^2 + b^2 = |w|^2$$
$$(w + z)^* = w^* + z^*$$
$$(wz)^* = w^* z^*$$

4.2.3 Algebraic Manipulation

$$\text{For} \quad w = a + ib \quad z = c + id$$
$$w = z \implies a = c, b = d \quad \textbf{IMPT}$$
$$w + z = (a + c) + i(b + d)$$
$$w - z = (a - c) + i(b - d)$$
$$w * z = (ac - bd) + i(ad + bc)$$
$$|w| = \sqrt{a^2 + b^2}$$
$$\sqrt{w} \quad \text{occurs with a } \pm \text{ sign}$$

For division, remove i from denominator by multiplying numerator and denominator by conjugate and then solve:

$$\frac{w}{z} = \frac{w}{z} \times \frac{z^*}{z^*}$$
$$= \frac{wz^*}{zz^*}$$
$$= \frac{wz^*}{c^2 + d^2}$$

4.3 Complex Polynomial Roots

4.3.1 Fundamental Theorem of Algebra

A polynomial of degree n has n real or complex roots.

4.3.2 Finding Complex Roots

If a polynomial has all real coefficients, complex roots occur in conjugate pairs. Write as such in exams:

$$f(x) \quad \text{has real coefficients}$$
$$a + ib \quad \text{is a root} \implies a - ib \quad \text{is a root}$$

For a polynomial with complex coefficients, use quadratic general formula. Note that a $\pm$ will still be present somewhere.

4.4 Polar Representation

4.4.1 Polar Representation

$$z = re^{i\theta} \quad r \in \mathbb{R}_0^+ \quad -\pi < \theta \leq \pi$$
$$|z| = r \quad arg(z) = \theta = \tan^{-1}(\frac{a}{b})$$
$$Re(z) = r\cos(\theta)$$
$$Im(z) = r\sin(\theta)$$
$$re^{i\pi} = -1 \quad re^{i0} = 1$$
$$re^{i\frac{\pi}{2}} = i \quad re^{i\frac{-\pi}{2}} = -i$$

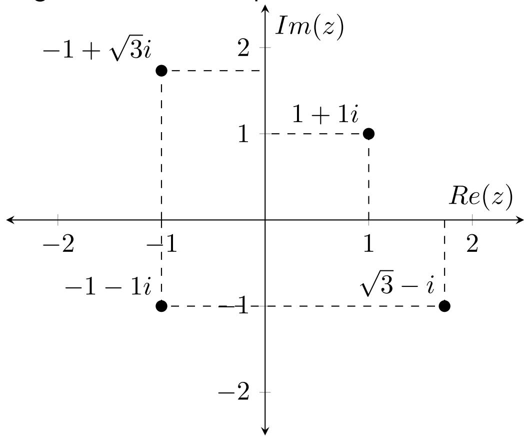
4.4.2 Algebraic Manipulation

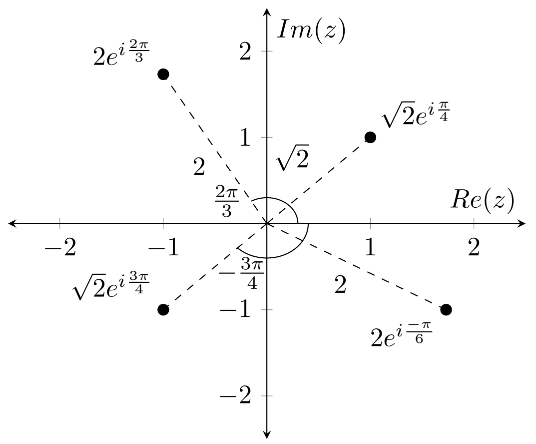
$$\text{For} \quad z_1 = r_1 e^{i\theta_1} \quad z_2 = r_2 e^{i\theta_2}$$
$$z_1 z_2 = r_1 r_2 e^{i(\theta_1 + \theta_2)}$$
$$\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$$
$$z_1^n = r_1^n e^{in\theta_1}$$
$$\sqrt{z_1} = \sqrt{r_1} e^{i(\frac{\theta_1}{2})}, \sqrt{r_1} e^{i(\frac{\theta_1}{2} + \pi)}$$

4.5 Geometric Representation

4.5.1 Argand Diagrams

Cartesian form describes the x and y coordinate on the real and imaginary axis.
Polar form describes the distance between the point and the origin as well as the angle a line from the point to the origin makes with the positive real axis.





4.5.2 Geometric Manipulation

Multiplying by $re^{i\theta}$ scales the number by a factor of r and rotates anticlockwise by an angle of θ about the origin.
Multiplying by i rotates a complex number by $\frac{\pi}{2}$, or $90°$ anticlockwise.
The conjugate of a complex number is a reflection of the complex number on the x axis.

5 Functions

5.1 Properties of a Function

5.1.1 Individual Properties

A function f is a relation which maps input of a set D_f to outputs of a set R_f using a certain rule.
Multiple elements in the input set can have the same output, but one single element in the input set can only have one output.

5.1.2 Function Presentation

When questions ask for functions in a similar form, be sure to maintain presentation.

$$f(x) = x^2 \quad x \in (-\infty, \infty)$$

$$f : x \mapsto x^2 \quad x \in (-\infty, \infty)$$

$$g(x) = \begin{cases} x^2 & x \in \mathbb{R}, x > 0 \\ -x & x \in \mathbb{R}, x < 0 \end{cases}$$

$$D_f = (-\infty, \infty) \quad R_f = [0, \infty)$$

Note: infinity is always written as non-inclusive

5.2 Inverse Functions

Inverse functions map the output of a function to the input of a function. Inverse functions only exist when each element of the output set of the original function is mapped to one and only one element in the input set, i.e. inverse functions only exist if a function is one-one.
Inverse functions are written as f^{-1} and

$$D_{f^{-1}} = R_f \quad R_{f^{-1}} = D_f$$

5.2.1 Proving existence and inexistence

$f(x)$ cuts each line $y = k$, $k \in R_f$ at one and only one point, f is one-one, hence f^{-1} exists
Replace R_f with the actual set
The line $y = k$ cuts $f(x)$ at more than one point, f is not one-one, hence f^{-1} does not exist .
Replace k with the actual edge case.

5.2.2 Finding Inverse Functions

$$f(x) = x^2 + 1 \quad x \in [0, \infty)$$
$$\text{Let} \quad y = f(x) = x^2 + 1$$
$$x^2 = y - 1$$
$$x = \sqrt{y-1}$$
$$f^{-1}(y) = \sqrt{y-1}$$
$$f^{-1}(x) = \sqrt{y-1}$$
$$D_{f^{-1}} = R_f = [1, \infty)$$
$$R_{f^{-1}} = D_f = [0, \infty)$$

5.2.3 Graphical Relationships Between Functions and Inverse Functions

Inverse functions are, in essence, a function reflected along the line $y = x$.
Intersections of functions with their inverse also satisfy the condition $f(x) = x$.

5.3 Composite Functions

Considering two functions f and g, the composite function fg is obtained when inputs of g are mapped to their outputs of g, which are then used as inputs to f and mapped to outputs of f:

$$fg(x) = f(g(x))$$

5.3.1 Deriving Composite Functions

For the composite function fg to exist, $R_g \subseteq D_f$.
The domain of function fg follows the domain of function g, i.e. $D_{fg} = D_g$.
The rule of fg is obtained by substituting the rule of g into the rule of f.
The range of fg is a subset of R_f and may be limited due to the fact that R_g may be smaller than D_f, hence R_{fg} must be reevaluated after creating its rule.

6 APGP

6.1 Arithmetic Progression

6.1.1 Arithmetic Sequence

An arithmetic progression is a sequence of numbers which have the same difference between consecutive elements. Sequences are defined by their initial term a and their constant difference d

$$u_1 = a = a + (1-1)d$$
$$u_2 = u_1 + d = a + d = a + (2-1)d$$
$$u_3 = u_2 + d = a + d + d = a + (3-1)d$$
$$...$$
$$u_n = a + (n-1)d$$

6.1.2 Arithmetic Series

An Arithmetic Series is defined as the sum of a certain number of consecutive elements in an arithmetic sequence.

$$S_1 = u_1$$
$$S_2 = S_1 + u_2 = u_1 + u_2$$
$$S_3 = S_2 + u_3 = u_1 + u_2 + u_3$$
$$...$$
$$S_n = u_1 + u_2 + ... + u_{n-1} + u_n$$

For an arithmetic sequence of known initial term and constant difference, the term S_n can be derived from the equation

$$S_n = \frac{n}{2}(2a + (n-1)d) = \frac{n}{2}(u_1 + u_n)$$

6.1.3 Proving an AP

To prove an AP, show that $u_n - u_{n-1} = d$ for all $n \geq 2$

6.2 Geometric Progression

6.2.1 Geometric Sequence

A geometric progression is a sequence of numbers which have the same constant ratio between consecutive elements. Sequences are defined by their initial term a and their constant ratio r

$$u_1 = a = ar^{1-1}$$
$$u_2 = u_1 r = ar^{2-1}$$
$$u_3 = u_2 r = ar^{3-1}$$
$$...$$
$$u_n = ar^{n-1}$$

6.2.2 Geometric Series

A Geometric Series is defined as the sum of a certain number of consecutive elements in a geometric sequence.
For a geometric sequence of known initial term and constant difference, the term S_n can be derived from the equation

$$S_n = \frac{a(r^n - 1)}{r - 1} = \frac{a(1 - r^n)}{1 - r}$$

The first equation is typically used for $r > 1$ while the second is used for $r < 1$.

6.2.3 Convergence

For a $r < 1$, is can be proven that as n tends to infinity, the value of u_n tends to zero and the value of S_n converges to a certain value and the series of this geometric sequence is said to be convergent. The value to which a series converges to is given by:

$$S_\infty = \frac{a}{1 - r}$$

The equation can be derived from the general formula of a geometric series as the numerator term $1 - r^n$ can be reduced to 1 as r^n tends to zero.

6.2.4 Proving a GP

To prove a GP, show that $\frac{u_n}{u_{n-1}} = r$ for all $n \geq 2$

7 Calculus I

7.1 Differentiation

7.1.1 Standard Differentiation Forms

$$\frac{d}{dx}f(x) = f'(x)$$
$$\frac{d}{dx}f(x)^n = f'(x) \quad n \quad f(x)^{n-1}$$
$$\frac{d}{dx}ln(f(x)) = \frac{f(x)}{f'(x)}$$
$$\frac{d}{dx}e^{f(x)} = f'(x) \quad e^{f(x)}$$
$$\frac{d}{dx}k^{f(x)} = \frac{d}{dx}e^{f(x)ln(k)}$$
$$= k^{f(x)}ln(k)f'(x)$$

7.1.2 Standard Differentiation Methods

Chain Rule: $\frac{dy}{dx} = \frac{du}{dx}\frac{dy}{du}$
Product Rule: $\frac{d}{dx}f(x)g(x) = f(x)g'(x) + g(x)f'(x)$
Quotient Rule: $\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}$

7.1.3 Implicit Differentiation

$$f(x) = g(y)$$
$$\frac{d}{dx}f(x) = \frac{d}{dx}g(y)$$
$$= \frac{d}{dx}\frac{dy}{dy}g(y)$$
$$= \frac{dy}{dx}\frac{d}{dy}g(y)$$
$$\frac{dy}{dx} = \frac{d}{dx}f(x) / \frac{d}{dy}g(y)$$
$$= \frac{f'(x)}{g'(y)}$$

Take differential of both sides and differentiate $g(y)$ in terms of y and $\frac{dy}{dx}$, then isolate $\frac{dy}{dx}$ to obtain a suitable expression.

7.1.4 Trigonometric Derivatives

$$\begin{aligned}
\frac{d}{dx}\sin(x) &= \cos(x)\\
\frac{d}{dx}\cos(x) &= -\sin(x)\\
\frac{d}{dx}\tan(x) &= \sec^2(x)\\
\frac{d}{dx}\sec(x) &= \tan(x)\sec(x)\\
\frac{d}{dx}\csc(x) &= -\csc(x)\cot(x)\\
\frac{d}{dx}\sin^{-1}(x) &= \frac{1}{\sqrt{1-x^2}}\\
\frac{d}{dx}\cos^{-1}(x) &= -\frac{1}{\sqrt{1-x^2}}\\
\frac{d}{dx}\tan^{-1}(x) &= \frac{1}{1+x^2}
\end{aligned}$$

7.1.5 Parametric Differentiation

$$\begin{aligned}
x = f(t) &\quad y = g(t)\\
\frac{dx}{dt} = f'(t) &\quad \frac{dy}{dt} = g'(t)\\
\frac{dy}{dx} = \frac{dy}{dt}\Big/\frac{dx}{dt} &= g'(t)/f'(t)
\end{aligned}$$

7.2 Applications of Differentiation

7.2.1 Minima and Maxima

Minimum points occur when a graph changes from increasing to decreasing gradient. Maximum points occur when a graph changes from increasing to decreasing gradient.

To find minima and maxima, find where $\frac{dy}{dx} = 0$. To test whether a point is a minimum or a maximum, use the first derivative test or second derivative test.

7.2.2 First Order Derivative Test

7.2.3 Second Order Derivative Test

Find x where $\frac{dy}{dx} = 0$. Evaluate $\frac{d^2y}{dx^2}$ at this value of x. A positive value indicates a minimum point while a negative value indicates a maximum point.

7.2.4 Concavity and Points of Inflection

A range $[u, v]$ of a graph $f(x)$ is concave down if all points within this range is equal to or above a line passing through $f(u)$ and $f(v)$. A range is concave up if all points within this range is equal to or above this line. Strictly concave graphs do not have the property of "or equal to".

The concavity of a graph can also be related to gradient. A range $[u, v]$ of a graph $f(x)$ is concave down if its gradient is non-increasing across this interval. A range is concave if its gradient is non-decreasing across this interval.

Points of inflection occur when there is a change from concavity. Therefore, to find points of inflection, set the derivative of the gradient to zero, hence set the second derivative to zero and solve: $\frac{d^2y}{dx^2} = 0$.

7.3 Graphing Techniques

7.3.1 Graphs with Asymptotes

Consider graphs of the form $y = \frac{f(x)}{g(x)}$.

Observe that if there are points where $g(k) \to 0$, there will be a vertical asymptote at the line $x = k$.

Also notice that if $f(x)$ and $g(x)$ are polynomials, partial fractions can be used to simplify the graph if the order of $f(x)$ is larger than the order of $g(x)$.

Consider $f(x) = ax^2 + bx + c$ and $g(x) = dx + e$. y can then be simplified using partial fractions to obtain y in the form $y = px + r + \frac{s}{dx+e}$. Such a graph will have a vertical asymptote at $dx + e = 0$ and have oblique asymptotes tending towards $y = px + r$.

Note that if $a = 0$ and hence $f(x)$ is an order 1 polynomial, the oblique asymptote will become a horizontal asymptote. Also note that $\lim_{y\to\pm\infty} y = px + r$ does not mean that $y \neq px + r$ at any x. There can be interceptions of the curve with the asymptote.

To find range of values of y, either locate turning points of the graph using differentiation or equate y to some constant and solve where discriminant is more than or equal to zero.

7.3.2 Conic Sections

Conic Sections are a special category of graphs because they can be derived through the intersection of a plane with a three-dimensional biconic function.

Circles have equations of the form $\frac{(x-a)^2}{r^2} + \frac{(y-b)^2}{r^2} = 1$, where coordinates (a, b) indicate the center of the circle and r is the radius of said circle.

Ellipses have equations of the form $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$, where coordinates (h, k) indicate the center of the ellipse, a is the largest distance from the center to any point in the x axis and b is the largest distance from the center to any point in the y axis.

Hyperbola have equations of the form $\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1$ or $\frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1$, where coordinates (h, k) indicate the center of the curves, and its asymptotes are given by the equation $y - k = \frac{b}{a}(x - h)$.

7.4 Mclaurin Series

Any equation of a n-degree polynomial can be recovered from the values of its first n derivatives at $x = 0$.

$$f(x) = a + bx + cx^2 + dx^3 \ldots \Rightarrow f(0) = a$$
$$f'(x) = b + 2cx + 3dx^2 \ldots \Rightarrow f'(0) = b$$
$$f''(x) = 2c + 6dx \ldots \Rightarrow f''(0) = 2c$$
$$f'''(x) = 6d \ldots \Rightarrow f'''(0) = 6d$$
$$\ldots$$
$$f(x) = f(0) + f'(x) + \frac{f''(x)}{2!} + \frac{f'''(x)}{3!} \ldots$$

By assuming that all equations can be approximated by a polynomial, approximations in a polynomial form can be given to non-polynomial functions. Values of these polynomials can be evaluated and used if the polynomial is convergent. I.e. for an arbitary function $f(x)$:

$$f(x) \approx f(0) + \frac{f'(0)}{1!} + \frac{f''(0)}{2!} + \frac{f'''(0)}{3!} \ldots$$

7.5 Graph Transformatons

7.5.1 Linear Transformations

Four linear transformations of graphs need to be known:

- Scale along x axis:
  Scaling by factor a: $f(x) \Rightarrow f(\frac{x}{a})$
- Scale along y axis:
  Scaling by factor a: $f(x) \Rightarrow f(x)a$
- Translate along x axis:
  Shifting left a units: $f(x) \Rightarrow f(x+a)$
  Shifting right a units: $f(x) \Rightarrow f(x-a)$
- Translate along y axis:
  Shifting up a units: $f(x) \Rightarrow f(x) + a$
  Shifting down a units: $f(x) \Rightarrow f(x) - a$

7.5.2 Inverse Graphs

Transformation of a graph $f(x)$ to a graph $\frac{1}{f(x)}$ holds the properties:

- X intercepts $\Rightarrow$ vertical asymptotes
- Vertical asymptotes $\Rightarrow$ X intercepts with exclusion circle
- Horizontal asymptotes remain
- Exact coordinates for all marked points that are not on the x-axis

7.5.3 Derivative Graphs

Transformation of a graph $f(x)$ to a graph $f'(x)$ holds the properties:

- Maximum, minimum and points of inflection $\Rightarrow$ x-intercepts
- Vertical asymptotes remain in location
- Oblique asymptotes $\Rightarrow$ Horizontal asymptotes

8 Calculus II

8.1 Integration

8.1.1 Standard Integral Forms

$$\int f'(x)dx = f(x) + c$$
$$\int f'(x)(f(x))^n dx = \frac{1}{n+1} \quad f(x)^{n+1} + c$$
$$\int \frac{f'(x)}{f(x)}dx = \ln(|f(x)|) + c$$
$$\int f'(x)e^{f(x)}dx = e^{f(x)} + c$$

8.1.2 Special and Trigonometric Integrals

$$\int \frac{1}{x^2+a^2}dx = \frac{1}{a}\tan^-1\left(\frac{x}{a}\right) + c$$
$$\int \frac{1}{\sqrt{a^2-x^2}}dx = \sin^-1\left(\frac{x}{a}\right) + c$$
$$\int \frac{1}{x^2-a^2}dx = \frac{1}{2a}\ln\left(\frac{x-a}{x+a}\right) + c$$
$$\int \frac{1}{a^2-x^2}dx = \frac{1}{2a}\ln\left(\frac{a+x}{a-x}\right) + c$$
$$\int \tan(x)dx = \ln(\sec(x))$$
$$\int \cot(x)dx = \ln(\sin(x))$$
$$\int \csc(x)dx = -\ln(\csc(x) + \cot(x))$$
$$\int \sec(x)dx = \ln(\sec(x) + \tan(x))$$

8.1.3 Integration by Substitution

Given a substitution $u(x)$ where $f(x) = g(u)$

$$\int f(x)dx = \int g(u)\frac{dx}{du}du$$

For definite integrals, substitute u back to receive the integral as a function of x. For definite integrals with limits $[a,b]$, change limits to $[u(a), u(b)]$ to evaluate the integral.

8.1.4 Integration by Parts

Recall the chain rule:

$$\frac{d}{dx}f(x)g(x) = f(x)g'(x) + f'(x)g(x)$$
$$f(x)g(x) = \int f(x)g'(x) + f'(x)g(x)dx$$
$$\int f(x)g'(x) = f(x)g(x) - \int f'(x)g(x)dx$$

To find the indefinite integral of a function involving multiplication, use integration by parts to eliminate one part of the integral.

8.2 Definite Integrals

8.2.1 Area Under Curve

Area bounded by x axis, $x = a$, $x = b$ and $f(x)$:

$$\int_a^b |f(x)|dx$$

Area bounded by y axis, $y = a$, $y = b$ and $f(y)$:

$$\int_a^b |f(y)|dy$$

8.2.2 Area of Parametric

Area bounded by x axis, $x = a$, $x = b$ and parametric curve $(x(t), y(t))$:

$$\int_a^b ydx = \int_c^d y(t)\frac{dx}{dt}dt$$

Where $x(c) = a, x(d) = b$

8.2.3 Volume of Rotation

Volume enclosed by curve rotated about x axis:

$$\pi \int_a^b (f(x))^2 dx$$

9 Differential Equations

9.1 Variable Separable Differential Equations

A differential equation is a equation which expresses derivatives of a function in terms of the function itself. Variable Separable Differential Equations are differential equations whose parts in terms of the original function can be removed from the parts which are in terms of other variables. To solve these, separate the equations and then integrate both sides respectively.

$$\begin{aligned}\frac{dy}{dx} &= xy\\ \frac{1}{y}\frac{dy}{dx} &= x\\ \int \frac{1}{y}dy &= \int xdx\\ \ln(y) + c_1 &= \frac{1}{2}x^2 + c_2\end{aligned}$$

Be sure to correctly manipulate constants. Lowercase constants typically denote added or subtracted constants while uppercase constants typically denote constants as coefficients.

$$\begin{aligned}\ln(y) + c_1 &= \frac{1}{2}x^2 + c_2\\ \ln(y) &= \frac{1}{2}x^2 + c_3\\ y &= e^{\frac{1}{2}x^2 + c_3}\\ &= C_1 e^{\frac{1}{2}}\end{aligned}$$

9.2 Initial Conditions

Differentiation is a destructive process. Information is lost when a function is differentiated, such as when the y-intercept of a $y = mx + c$ graph is no longer known once differentiated. When questions provide initial conditions, known values of x, y and their derivatives can be substituted to find the actual value of these constants.

10 Combinatorics and Probability

10.1 Permutations and Combinations

$${}^nP_r = \frac{n!}{(n-r)!}$$

${}^nP_r$ gives the number of ways to order r objects out of a possible n objects.

If a group of objects to be ordered contains $k_1, k_2 \ldots$ similar objects, divide the number of permutations by $k_1!k_2! \ldots$.

If n objects are placed in a ring, find the number of different ways to organize the n objects in a line and then divide by n as each permutation is overcounted for each n possible rotations.

$${}^nC_r = \frac{n!}{(n-r)!r!}$$

${}^nC_r$ gives the number of ways to choose r objects out of a possible n objects.

10.2 P & C Heuristics

10.2.1 'Group' Method

When arranging n objects, of which k must placed next to each other, consider the k objects as one single unit. Arrange the $n-k+1$ objects, then multiply by the number of ways the k objects can be arranged among themselves.

10.2.2 'Slot' Method

When arranging n objects, of which k cannot be placed next to each other, arrange the remaining $n - k$ objects and then 'slot' each of the k objects in the gaps between the $n - k$ objects.

10.3 Probability and Search Space

For a number of events which have equal chance to occur, the function P represents the fraction of cases where this condition is true.

$$\text{P(condition)} = \frac{\text{Cases where condition is true}}{\text{Count of all possible cases}}$$

As a result, $\mathrm{P}(\text{Guaranteed true condition}) = 1$ and $\mathrm{P}(\text{Guaranteed false condition}) = 0$.

10.4 Inverse, Unions and Intersections

The probability of multiple conditions can be compounded using their Inverse, Unions and Intersections.

The inverse A' of an event is the set of events which do not appear in the event A.

The union A ∪ B of two events is the set of events which appear at least once in A or B.

The intersection A ∩ B of two events is the set of events which appear in both A and B.

Two events are mutually exclusive if the intersection between the two events is empty.

$$\mathrm{P}(\text{A}) = 1 - \mathrm{P}(\text{A'})$$
$$\mathrm{P}(\text{A} \cup \text{B}) = \mathrm{P}(\text{A}) + \mathrm{P}(\text{B}) - \mathrm{P}(\text{A} \cap \text{B})$$
$$\mathrm{P}(\text{A}) = \mathrm{P}(\text{A}|\text{B}) + \mathrm{P}(\text{A}|\text{B'})$$

10.5 Conditional Probability

Conditional probability considers the probability of an event given that another event has already occurred.

$$\begin{aligned}&\mathrm{P}(\text{B}|\text{A}) \\ &= \text{Probability that B occurs given that A has occurred} \\ &= \frac{\mathrm{P}(\text{B} \cap \text{A})}{\mathrm{P}\,A}\end{aligned}$$

10.6 Independence

Two events are independent if one event occurring does not affect the probability of another from occurring.

$$\begin{aligned}&\text{A and B are Independent} \\ \iff\ &\mathrm{P}(\text{B}|\text{A}) = \mathrm{P}(\text{B}) \\ \iff\ &\mathrm{P}(\text{A}|\text{B}) = \mathrm{P}(\text{A}) \\ \iff\ &\mathrm{P}(\text{A} \cap \text{B}) = \mathrm{P}(\text{A})\,\mathrm{P}(\text{B})\end{aligned}$$

11 Discrete Random Variable

11.1 Probability Distribution

A discrete random variable X (or denoted by other capital letters) is a variable which can take a finite and distinct set of values. Its probability distribution is a description of the probability that it takes some value.

x	1	2	...	5
$\mathrm{P}(X = x)$	0.2	0.3		0.1

11.2 Expectation and Variance

The expectation value of a discrete random variable is the sum of the products of its possible values and the probability for it to be some value.

$$\mathrm{E}(X) = \sum_x \mathrm{P}(X = x) \times x$$

The variance of a discrete random variable is a quantification of how spread out are the possible values of X from its mean.

$$\begin{aligned}\mathrm{Var}(X) &= \sum_x \mathrm{P}(X = x) \times (x - \mathrm{E}(X))^2 \\ &= \mathrm{E}(X^2) - \mathrm{E}(X)^2\end{aligned}$$

The standard deviation σ of a discrete random variable is another representation of spread from its mean, but more accurate to the scale of the original random variable.

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

11.3 Binomial Distribution

A binomial distribution $X \sim B(n, p)$ is the probability distribution describing the number of successful trials over n total trials, each with a probability of success p.

$$\mathrm{P}(X = x) = ^n C_x \times p^x \times (1 - p)^{n-x}$$
$$\mathrm{E}(X) = np$$
$$\mathrm{Var}(X) = np(1 - p)$$

12 Continuous Random Variable

12.1 Continuous Random Variable

Unlike discrete random variables, continuous random variables are described by a probability distribution function (p.d.f.). The actual probability of the random variable having a value between a range is its direct integral of the p.d.f. within the range.

A p.d.f., when integrated across the range $(-\infty, \infty)$ should have a total probability of 1. Such a function is said to be normalized.

12.2 Expectation and Variance

The expectation value of a continuous random variable is the integral of the product of its instantaneous value and the probability distribution function at that value.

$$\mathrm{E}(X) = \int \mathrm{P}(X = x)x\ dx$$

The variance of a continuous random variable is obtained in a similar fashion to discrete random variables.

$$\begin{aligned}\mathrm{Var}(X) &= \int \mathrm{P}(X = x) \times (x - \mathrm{E}(X))^2\ dx \\ &= \mathrm{E}(X^2) - \mathrm{E}(X)^2\end{aligned}$$

12.3 Normal Distribution

A normal distribution is a graph of the form e^{-x^2}, with added terms to ensure that its mean and standard deviation is equal to a specified value. Normal distributions are defined as $X \sim N(\mu, \sigma^2)$ by assigning them a mean and a standard deviation.

12.4 Combination of Random Variables

$$X \sim N(\mu_X, \sigma_X^2) \qquad Y \sim N(\mu_Y, \sigma_Y^2)$$
$$X + c \sim N(\mu_X + c, \sigma_X^2)$$
$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$
$$X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$$
$$k_1 X + k_2 Y \sim N(k_1\mu_X + k_2\mu_Y, k_1^2\sigma_X^2 + k_2^2\sigma_Y^2)$$

- Adding a constant to a normal distribution only adjusts its mean
- Adding two normal distributions adds their means together and adds their variances
- Subtracting a normal distribution from another requires subtracting their mean but adding their variances
- Multiplying a random variable by a constant factor k requires adding its mean with a factor k and adding its variance with an additional factor of k^2

12.5 Standard Normal Distribution

The standard normal distribution Z is a specially defined normal distribution as $Z \sim N(0, 1)$. Other normal random variables can be adjusted in order to equal the standard normal distribution, by offsetting $\bar{X}$ to be centered at 0 and then dividing by its standard deviation σ.

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

From this, the calculation of probabilities can be done through the use of data tables or the inverse normal distribution function.

12.6 Modeling as Normal Distribution

To model a random variable as a normal distribution:

- Random variable should have bell curve-like distribution
- Random variable should 'make sense' with values from $(-\infty, \infty)$, or at least minimize the probability of a nonsensical value
- Multiple observations of the random variable should be independent of each other
- Observations of multiple different random variables should be independent of each other

13 Sampling

13.1 Populations and Samples

A population is the entire collection of data to be studied. When populations are too large, too fluid/changing to measure at once or the amount of data to be collected is limited, a sample may be taken to represent the distribution of the population instead.

Given a population described by a random variable X, multiple independent observations of that random variable can be used to obtain a sample. From this sample, an approximate of the population's parameters of mean and variance can be obtained.

13.2 Sample Parameters

When a sample of size n is taken from a population X with unknown population mean μ and unknown population variance σ^2, sample mean and sample variance can be obtained.

From the sample parameters, an unbiased estimate of the population mean $\bar{x}$ and an unbiased estimate of the variance s^2 can be obtained.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x}{n}$$
$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$$

13.3 Sample Mean

Different samples of a population may get a different unbiased estimate of mean $\bar{x}$. As such, the distribution of such sample means can be said to be a random variable $\bar{X}$, defined as

$$\bar{X} = \frac{1}{n}\sum X = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

with $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

13.4 Central Limit Theorem

The central limit theorem states that the sample mean from of large sample size $n \geq 50$, regardless of the distribution of the population, will *approximately* follow a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

14 Hypothesis Testing

14.1 Hypotheses

A null hypothesis $H_0 : \mu = k$ is a proposed hypothesis of what property a random variable should hold.

An alternative hypothesis $H_1 : \mu \neq k$ is a proposed hypothesis which contradicts a null hypothesis, and describes how a null hypothesis will be tested.

The Level of Significance α of a test is the probability of incorrectly rejecting the null hypothesis, i.e. the the chance a false positive is identified. Uses of α can range from 0.000001 in physical sciences to 0.05 in general sciences to 0.3 in social sciences.

14.2 Z-Test

A Z-test is a statistical test for the validity of a null hypothesis which follows a normal distribution.

A null hypothesis is first established with its mean μ and variance $\sigma^2 \text{OR} s^2$, either from known data or using an unbiased estimate of a sample.

A sample mean with n samples and mean $\bar{x}$ is taken. This sample mean is then compared to the distribution of the null hypothesis with distribution $X \sim N(\mu, \frac{\sigma^2}{n})$.

The sample mean is then verified if it lies at the extremities of the null hypothesis' probability distribution, by finding its relative probability (p-value) or finding the critical value at which results then are beyond the level of significance (z-value).

14.2.1 Syntax for Z-Test

1. Define random variables if necessary.

   'Let X be the random variable denoting ___.'

2. Hypotheses H_0 and H_1 with their definitions.

   '$H_0 : \mu =$ ___ vs $H_1 : \mu \neq$ ___'; OR
   '$H_0 : \mu =$ ___ vs $H_1 : \mu >$ ___'; OR
   '$H_0 : \mu =$ ___ vs $H_1 : \mu <$ ___'

3. Type of test, level of significance.

   'Conduct a 2-tailed test at a $\alpha =$ ___% level of significance'; OR
   'Conduct a 1-tailed test at a $\alpha =$ ___% level of significance'

4. Describe random variable in terms of known and unknown quantities, stating the values of known quantities; state the sample used and its derived quantities.

   If variance is given:

   'Under H_0, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ where $\mu =$ ___ and $\sigma^2 =$ ___.
   From sample: $\bar{x} =$ ___, $n =$ ___'

   If variance is estimated from sample:

   'Under H_0, $\bar{X} \sim N\left(\mu, \frac{s^2}{n}\right)$ where $\mu =$ ___.
   From sample: $\bar{x} =$ ___, $n =$ ___, $s^2 =$ ___,'

5. Calculate p-value or z-value and compare to level of significance.

   p-value approach: calculate the area of the probability distribution between the extremities and $\bar{x}$, then compare to level of significance.

   1-tailed: 'p-value $= \mathrm{P}(\bar{X} < \bar{x}) =$ ___'
   2-tailed: 'p-value $= \mathrm{P}(\bar{X} < \bar{x}) \times 2 =$ ___'
   In 2-tailed tests, p-value is two times the probability calculated as two 'tails' of probability are compared with α, do NOT halve α instead.
   End with comparing to α, writing less or greater than.

   z-value approach: find the 'critical value' corresponding to α and the standard normal distribution Z, then transform to the given distribution $\bar{X}$ and compare with $\bar{x}$.
   1-tailed: 'z-value $= \mathrm{P}(Z < \frac{c-\mu}{\sigma}) = \alpha, c =$ ___'
   2-tailed: 'z-value $= \mathrm{P}(Z < \frac{c-\mu}{\sigma}) = \frac{\alpha}{2}, c =$ ___'
   In 2-tailed tests, z-value is calculated with α divided by 2 because the probability is distributed across two tails.
   End with comparing to $\bar{x}$, writing less or greater than.

6. Conclude test with context.

   'Reject H_0 because there is sufficient evidence at a ___% level of significance to conclude that ____ '; OR

   'Reject H_0 because there is insufficient evidence at a ___% level of significance to conclude that ____ '

15 Linear Regression

15.1 2-Variable Data

Previous statistics regarding probability distributions and hypothesis testing typically involve data points with one dimension of quantity, but often there is a use to determine a relation between two different quantities of collected data. The approach of correlation and regression attempts to identify these relationships in a consistent and objective manner while handling statistical interference.

Data collected with two quantities at one point are called bivariate data. Sometimes one quantity may be controlled to be measured at predetermined values. In this scenario, the controlled variable is the 'independent' variable and the other is the 'dependent' variable.

15.2 Scatter Plots

A scatter plot displays points on a graph where data has been measured, with each set of values varying the x and y coordinates respectively. When drawing scatter plots:

- Draw and label axis with their variables x, y, arrow pointing towards increasing.
- Mark the minimum and maximum values of each axis and label their values.
- Attempt to maintain the same shape of the trend
- Preserve relative heights of plots, preserve even or uneven spacing and grouping.
- Otherwise, no need to be too precise.

A scatter plot is good for how tightly-spaced a group of data are to its best fit line and also the type of relation (linear, quadratic, exponential etc.) it displays.

15.3 Product Moment Correlation Coefficient

The Product Moment Correlation Coefficient r is a quantity describing whether a set of bivariate data follows a increasing ($r \approx 1$) or decreasing ($r \approx -1$) trend. The formula for calculation can be found in MF26, or alternatively calculated through your GC using the "2-Var Stats" function.

$r = 1$ indicates a perfect positive linear correlation with all points on the line, $r = -1$ indicates a perfect negative linear correlation with all points on the line. $r = 0$ indicates no linear correlation, meaning variables could be not related OR have a nonlinear relationship.

The value of r^2 removes the negative sign on r and its value used to determine if a relation present.

15.4 Linear Regression

After identifying a linear correlation between a set of bivariate data, the data can then be given a best fit line. The method of Linear Least-Squares Regression is an objective and consistent method to obtain an equation of the form $y = a+bx$ and calculating the necessary constants a and b.

The least squares regression of y on x assumes that the obtained data has NO error in x and that all deviation of data points from the best fit are due to errors in y. A linear graph is then fit to the data to minimize the sum of square of all residuals. At a data point (x_1, y_1) and point on the best fit (x_1, y_1'), the residual is $e_1 = y_1' - y_1$ and the sum of squares of residuals is $\sum e_r^2$. As a result of this approach, x is the independent variable and y is the dependent variable.

The process of obtaining this line of best fit is not through trial and error, but rather through selecting a central point and calculating the most suitable gradient to minimize the sum of squares of residuals. For a set of data with points (x_i, y_i), the point $(\bar{x}, \bar{y})$ is first selected and the gradient $b$is calculated with:

$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

The linear regression of x against y treats x as the dependent variable and y as the independent variable. This alternative approach is useful if data given goes against conventions of labeling x and y, but also in scenarios where neither set of data are reliably controlled. This regression assumes NO error in obtained values of y, and that all deviation of data points from the best fit are due to to errors in x.

15.4.1 Linearization

Even if two variables are not related by a linear relation of the form $y = mx + c$, some types of relationships can be manipulated into that form and to find a linear relationship between them.

$$y = ax^2 + b \xrightarrow{X=x^2} y = aX + b$$

$$\sqrt{y} = \frac{a}{x} + b \xrightarrow[Y=\sqrt{y}]{X=\frac{1}{x}} Y = aX + b$$

$$\begin{aligned} y = ab^x &\iff \ln(y) = \ln(a) + x\ln(b) \\ &\xrightarrow[Y=\ln(y)]{} Y = \ln(a) + x\ln(b) \end{aligned}$$

$$\begin{aligned} y = ax^b &\iff \ln(y) = \ln(a) + b\ln(x) \\ &\xrightarrow[Y=\ln(y)]{X=\ln(x)} Y = \ln(a) + bX \end{aligned}$$

Given a set of data and multiple ways to linearize it, the linearization which obtains a product moment correlation coefficient closes to ± 1 or the linearization which has a similar shape and trend in gradient is the more suitable one.

15.5 Applications of Linear Regression

15.5.1 Determining a Correlation

Whether a correlation is present in data can be determined by inspection of its scatter plot and its product moment correlation coefficient.

Scatter plots are used to identify the shape of a relationship and verify if a relation is linear or otherwise. Scatter plots are also effective in identifying outliers in data which may arise due to errors in data collection, after which specific data points can be ignored.

The product moment correlation coefficient is a quantitative measure of how well the data points fit in a linear relation. Calculating values of r for different methods of linearizing a set of data can help to find the most suitable linearization.

15.5.2 Predicting Values

After obtaining the equation of the linear regression between two values, one can predict and estimate the value of one variable when given the other.

If a set of data is given with a distinct independent variable x (whether stated in the problem, or if it takes 'nice'/whole values in the data given), use the equation obtained from the linear regression of y on x to predict future points.

If a set of data is given with no distinct independent variable, the linear regression equation to use depends on what information is given in the prediction. If finding y for a given x, use the regression of y against x, otherwise use the regression of x against y. This corresponds to the approach of calculating the linear regression: when predicting y for a given x, there is zero error in given value of x and hence the best regression to use would be when the data is assumed to have zero error in x, and vice versa.