

Mathematical Formulae, Statistical Tables & Optical Character Recognition (Math Formula OCR) Reference Guide

Islam Ayoub

Independent Researcher, Information Technology

Research Area: Mathematical Notation, LaTeX & MathML Encoding, Statistical Tables, Optical Character Recognition (Math Formula OCR), Client-Side WebAssembly Computing

ABSTRACT

Mathematical notation and statistical reference tables form the foundational language of quantitative science, engineering, finance, and academic publishing. Unlike one-dimensional natural language text, mathematical expressions exhibit complex two-dimensional spatial topologies—incorporating nested fractions, definite integrals, matrix arrays, radical bounds, superscript and subscript hierarchies, and specialized Greek and relational glyphs. Digitizing printed or scanned mathematical documents into structured, machine-readable markup such as LaTeX and Mathematical Markup Language (MathML) requires specialized Optical Character Recognition architectures known as Mathematical Formula OCR (Math Formula OCR).

This reference guide serves a dual technical objective. First, it provides a standardized, comprehensive mathematical formulae and statistical tables datasheet covering core domains: algebraic series, binomial and Taylor expansions, trigonometric and hyperbolic identities, three-dimensional vector geometry (including vector equations of lines and planes), differential and integral calculus, integral transforms, vector calculus, numerical approximation methods, linear algebra, probability distributions, and standard normal, Student's t , and chi-squared critical value tables—paired directly with their canonical LaTeX source encodings.

Second, this work investigates the computational pipeline of modern Math Formula OCR systems, examining spatial structural parsing, Vision-Encoder Decoder transformer architectures, Abstract Syntax Tree (AST) normalization, and hybrid execution topologies that balance high-accuracy zero-retention cloud inference with 100% on-device WebAssembly browser execution for sensitive scientific and financial documents.

Keywords: Math Formula OCR, Mathematical Formulae, Statistical Tables, LaTeX Conversion, Optical Character Recognition, MathML, Vector Equation of a Plane, Normal Distribution Tables, WebAssembly, Privacy-Preserving Document Analysis.

1. Introduction to Two-Dimensional Mathematical Notation

Standard optical character recognition (OCR) engines are designed to process sequential, one-dimensional lines of characters separated by uniform horizontal whitespace. When applied to scientific manuscripts, examination datasheets, engineering blueprints, or quantitative financial reports, conventional text OCR

frequently fails to preserve the two-dimensional structural semantics of mathematical equations. In mathematical notation, spatial position carries explicit algebraic meaning: a character displaced vertically by half a baseline transforms a linear product (x_2) into an exponent (x^2) or an indexed tensor component (x_2) [1, 2].

Furthermore, mathematical typography relies on variable-scale delimiters, multi-line array environments, and relational operators standardized across international frameworks including ISO 80000-2, the NIST Digital Library of Mathematical Functions (DLMF), Unicode Technical Report #25, and the W3C MathML 3.0 specification [1, 3, 4, 7]. Accurate digitization therefore requires both a rigorous reference mapping between mathematical identities and their canonical LaTeX representations, and a structure-aware recognition pipeline capable of reconstructing two-dimensional syntax trees from raster images.

Central Research Question: *How can standard mathematical formulae, vector geometry identities, and statistical reference tables be systematically mapped to canonical LaTeX markup while enabling high-precision, privacy-preserving optical character recognition across cloud and browser-native execution environments?*

1.1 Implementation Context

The mathematical notation standards, LaTeX syntax mappings, and optical character recognition architectures examined in this reference guide are directly relevant to browser-based scientific document processing platforms, including Utilvo, a web application providing dual-mode (zero-retention cloud AI and 100% on-device WebAssembly browser) mathematical formula OCR and LaTeX conversion. Utilvo is referenced in this document as an implementation context for discussing hybrid formula recognition, structural LaTeX generation, and client-side data minimization. This reference does not constitute an independent experimental validation of the security or recognition properties discussed in this study.

Further information about the mathematical formula recognition implementation is available at <https://utilvo.com/pdf-tools/ocr-formula> [9].

2. Algebraic Formulae, Series & Power Expansions

Algebraic identities and infinite series expansions appear ubiquitously across pure mathematics, physics, and numerical computing. When converting scanned mathematical datasheets via Math Formula OCR, structural disambiguation between summation bounds, binomial coefficients, and fractional exponents is critical [2, 3].

2.1 Fundamental Algebraic Identities & Logarithmic Laws

Mathematical Concept	Standard Mathematical Notation	Canonical LaTeX Representation
Quadratic Equation Roots (for $ax^2 + bx + c = 0, a \neq 0$)	$x = [-b \pm \sqrt{b^2 - 4ac}] / (2a)$	$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
Laws of Logarithms (Multiplication, Division, Power)	$\log_a(xy) = \log_a x + \log_a y$ $\log_a(x/y) = \log_a x - \log_a y$ $\log_a(x^k) = k \log_a x$	$\log_a(xy) = \log_a x + \log_a y$ $\log_a\left(\frac{x}{y}\right) = \log_a x - \log_a y$ $\log_a(x^k) = k \log_a x$
Change of Base Identity	$\log_b(x) = \ln(x) / \ln(b)$	$\log_b(x) = \frac{\ln x}{\ln b}$
Binomial Coefficient & Factorial	${}^nC_r = C(n, r) = n! / [r!(n - r)!]$	$\binom{n}{r} = \frac{n!}{r!(n-r)!}$
Binomial Series Expansion (for positive integer $n \in \mathbb{N}$)	$(a + b)^n = \sum_{k=0..n} C(n, k) a^{n-k} b^k$	$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$
General Binomial Expansion (for $ x < 1, n \in \mathbb{R}$)	$(1 + x)^n = 1 + nx + [n(n-1)/2!]x^2 + [n(n-1)(n-2)/3!]x^3 + \dots$	$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \dots$

2.2 Arithmetic, Geometric & Power Series

Series Type	Closed-Form Sum Formula	Canonical LaTeX Encoding
Arithmetic Progression (AP) First term a , common difference d	$u_n = a + (n - 1)d$ $S_n = \frac{1}{2}n[2a + (n - 1)d] = \frac{1}{2}n(a + l)$	$S_n = \frac{n}{2}\left[2a + (n-1)d\right]$
Geometric Progression (GP) First term a , common ratio $r \neq 1$	$u_n = ar^{n-1}$ $S_n = a(1 - r^n) / (1 - r)$	$S_n = \frac{a(1 - r^n)}{1 - r}$
Convergent Infinite Geometric Series (Condition: $ r < 1$)	$S_\infty = a / (1 - r)$	$S_{\infty} = \frac{a}{1 - r}, \quad r < 1$
Sum of First n Natural Integers	$\sum_{r=1..n} r = \frac{1}{2}n(n + 1)$	$\sum_{r=1}^n r = \frac{1}{2}n(n+1)$
Sum of Squares of First n Integers	$\sum_{r=1..n} r^2 = \frac{1}{6}n(n + 1)(2n + 1)$	$\sum_{r=1}^n r^2 = \frac{1}{6}n(n+1)(2n+1)$
Sum of Cubes of First n Integers	$\sum_{r=1..n} r^3 = \frac{1}{4}n^2(n + 1)^2$	$\sum_{r=1}^n r^3 = \frac{1}{4}n^2(n+1)^2$

2.3 Maclaurin and Taylor Power Series Expansions

For a function $f(x)$ infinitely differentiable in a neighborhood of $x = a$, Taylor's expansion is given by:

[CANONICAL LATEX – TAYLOR & MACLAURIN SERIES EXPANSIONS]

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \dots + \frac{(x-a)^n}{n!}f^{(n)}(a) + R_n(x)$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots \quad (x \in \mathbb{R})$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{n+1}\frac{x^n}{n} + \dots \quad (-1 < x \leq 1)$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^n\frac{x^{2n+1}}{(2n+1)!} + \dots \quad (x \in \mathbb{R})$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots + (-1)^n\frac{x^{2n}}{(2n)!} + \dots \quad (x \in \mathbb{R})$$

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \dots + (-1)^n\frac{x^{2n+1}}{2n+1} + \dots \quad (-1 \leq x \leq 1)$$

3. Trigonometry, Hyperbolic Functions & 3D Vector Geometry

Trigonometric identities, complex exponential representations, and three-dimensional vector equations of lines and planes represent a core benchmark in mathematical document recognition. In particular, vector equations require distinguishing boldface vector symbols ($\mathbf{r}$, $\mathbf{a}$) from scalar parameters (λ, μ) [3, 7].

3.1 Compound-Angle, Double-Angle & Factor Formulae

Identity Category	Mathematical Formula	LaTeX Syntax
Pythagorean Identities	$\sin^2 \theta + \cos^2 \theta = 1$ $1 + \tan^2 \theta = \sec^2 \theta$ $1 + \cot^2 \theta = \csc^2 \theta$	<code>\sin^2\theta + \cos^2\theta = 1</code> <code>1 + \tan^2\theta = \sec^2\theta</code> <code>1 + \cot^2\theta = \csc^2\theta</code>
Compound-Angle Identities	$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$ $\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$ $\tan(A \pm B) = (\tan A \pm \tan B) / (1 \mp \tan A \tan B)$	<code>\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B</code> <code>\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B</code> <code>\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A \tan B}</code>
Double-Angle & Half-Angle Identities (where $t = \tan(\frac{1}{2}A)$)	$\sin(2A) = 2 \sin A \cos A = 2t / (1 + t^2)$ $\cos(2A) = \cos^2 A - \sin^2 A = (1 - t^2) / (1 + t^2)$	<code>\sin A = \frac{2t}{1+t^2}, \quad \cos A = \frac{1-t^2}{1+t^2}</code>
Sum-to-Product (Factor) Formulae	$\sin P + \sin Q = 2 \sin(\frac{1}{2}(P+Q)) \cos(\frac{1}{2}(P-Q))$ $\cos P + \cos Q = 2 \cos(\frac{1}{2}(P+Q)) \cos(\frac{1}{2}(P-Q))$	<code>\sin P + \sin Q = 2\sin\frac{P+Q}{2}\cos\frac{P-Q}{2}</code>
Hyperbolic Definitions & Identities	$\sinh x = \frac{1}{2}(e^x - e^{-x}), \quad \cosh x = \frac{1}{2}(e^x + e^{-x})$ $\cosh^2 x - \sinh^2 x = 1$ $\operatorname{arsinh} x = \ln[x + \sqrt{x^2 + 1}]$	<code>\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}</code> <code>\cosh^2 x - \sinh^2 x = 1</code> <code>\operatorname{arsinh} x = \ln\left(x + \sqrt{x^2+1}\right)</code>
Euler's Relation & De Moivre's Theorem	$e^{i\theta} = \cos \theta + i \sin \theta$ $[r(\cos \theta + i \sin \theta)]^n = r^n(\cos n\theta + i \sin n\theta)$	<code>e^{i\theta} = \cos\theta + i\sin\theta</code> <code>\left[r(\cos\theta + i\sin\theta)\right]^n = r^n(\cos n\theta + i\sin n\theta)</code>

3.2 Three-Dimensional Vector Geometry: Equations of Lines and Planes

In three-dimensional Euclidean space $\mathbb{R}^3$, straight lines, planes, and vector products are defined as follows:

- Vector Equation of a Straight Line:** The straight line passing through a point A with position vector $\mathbf{a}$ and parallel to a direction vector $\mathbf{b}$ has vector equation: `\mathbf{r} = \mathbf{a} + \lambda \mathbf{b}`.
- Vector Equation of a Plane through Three Non-Collinear Points:** The plane through non-collinear points A, B , and C with position vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ has vector equation:

$$r = a + \lambda(b - a) + \mu(c - a) = (1 - \lambda - \mu)a + \lambda b + \mu c$$

- **Normal Form of the Equation of a Plane:** The plane through the point with position vector $\mathbf{a}$ and perpendicular to the normal vector $\mathbf{n}$ satisfies $(r - a) \cdot n = 0$, or $r \cdot n = d$ where $d = a \cdot n$. The perpendicular distance from (x_0, y_0, z_0) to the plane $n_1x + n_2y + n_3z + d = 0$ is $D = \frac{|n_1 x_0 + n_2 y_0 + n_3 z_0 + d|}{\sqrt{n_1^2 + n_2^2 + n_3^2}}$.
- **Vector (Cross) Product & Scalar Triple Product:** For vectors $a = a_1i + a_2j + a_3k$ and $b = b_1i + b_2j + b_3k$, the vector product is $\mathbf{a} \times \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \sin \theta \hat{\mathbf{n}}$, and the volume of the parallelepiped spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is $V = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$.

4. Differential & Integral Calculus Reference Datasheet

Calculus expressions introduce tall vertical symbols—including Leibniz differential fractions, indefinite and definite integral signs with upper and lower limits, and partial derivative operators—that require vertical bounding-box segmentation during Math Formula OCR [2, 3].

Function $f(x)$	Derivative $f'(x) = df/dx$ (LaTeX)	Indefinite Integral $\int f(x) dx$ (LaTeX)
x^n ($n \neq -1$)	$\frac{d}{dx}(x^n) = nx^{n-1}$	$\int x^n dx = \frac{x^{n+1}}{n+1} + C$
$e^{kx}, 1/x$	$\frac{d}{dx}(e^{kx}) = ke^{kx}$, $\quad \frac{d}{dx}(\ln x) = \frac{1}{x}$	$\int e^{kx} dx = \frac{1}{k}e^{kx} + C$, $\quad \int \frac{1}{x} dx = \ln x + C$
$\sin(kx), \cos(kx)$	$\frac{d}{dx}(\sin kx) = k \cos kx$, $\quad \frac{d}{dx}(\cos kx) = -k \sin kx$	$\int \sin kx dx = -\frac{1}{k} \cos kx + C$
$\tan(x), \sec(x)$	$\frac{d}{dx}(\tan x) = \sec^2 x$, $\quad \frac{d}{dx}(\sec x) = \sec x \tan x$	$\int \tan x dx = \ln \sec x + C$
$\arcsin(x/a)$	$\frac{d}{dx} \arcsin\left(\frac{x}{a}\right) = \frac{1}{\sqrt{a^2 - x^2}}$	$\int \frac{1}{\sqrt{a^2 - x^2}} dx = \arcsin\left(\frac{x}{a}\right) + C$
$\arctan(x/a)$	$\frac{d}{dx} \arctan\left(\frac{x}{a}\right) = \frac{a}{a^2 + x^2}$	$\int \frac{a}{a^2 + x^2} dx = \arctan\left(\frac{x}{a}\right) + C$
$\operatorname{arsinh}(x/a), \operatorname{arcosh}(x/a)$	$\frac{d}{dx} \operatorname{arsinh}\left(\frac{x}{a}\right) = \frac{1}{\sqrt{x^2 + a^2}}$	$\int \frac{1}{\sqrt{x^2 - a^2}} dx = \operatorname{arcosh}\left(\frac{x}{a}\right) + C$
Integration by Parts & Quotient Rule	$\frac{d}{dx} \left(\frac{u}{v} \right) = \frac{v \frac{du}{dx} - u \frac{dv}{dx}}{v^2}$	$\int u \frac{dv}{dx} dx = uv - \int v \frac{du}{dx} dx$

4.1 Integral Transforms & Vector Calculus Identities

In differential equations, signal processing, and continuum mechanics, Laplace and Fourier transforms and vector differential operators (`\nabla`) follow standardized LaTeX encodings [3, 7]:

[LAPLACE TRANSFORM, FOURIER TRANSFORM & VECTOR CALCULUS IDENTITIES]

% Unilateral Laplace Transform & Continuous Fourier Transform

$\mathcal{L}\{f(t)\}(s) = F(s) = \int_0^{\infty} e^{-st}f(t)\,dt, \quad \mathcal{F}\{f(t)\}(\omega) = \hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt$

% Gradient, Divergence, Curl & Laplacian Operators

$\nabla \phi = \frac{\partial \phi}{\partial x}\mathbf{i} + \frac{\partial \phi}{\partial y}\mathbf{j} + \frac{\partial \phi}{\partial z}\mathbf{k}, \quad \nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}, \quad \nabla^2 \phi = \nabla \cdot (\nabla \phi)$

% Gauss's Divergence Theorem & Stokes' Theorem

$\iiint_V (\nabla \cdot \mathbf{F})\,dV = \oiint_{\partial V} (\mathbf{F} \cdot \mathbf{n})\,dS, \quad \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{r} = \oint_{\partial S} \mathbf{F} \cdot d\mathbf{r}$

5. Numerical Analysis & Iterative Approximation Methods

When closed-form analytical solutions do not exist, numerical root-finding, quadrature, and differential equation solvers provide bounded approximations:

[NUMERICAL ANALYSIS FORMULAE – LATEX ENCODING]

% 1. Newton-Raphson Iterative Root Approximation

$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad f'(x_n) \neq 0$

% 2. Composite Trapezium Rule (with step $h = (b - a) / n$)

$\int_a^b y\,dx \approx \frac{1}{2}h[(y_0 + y_n) + 2(y_1 + y_2 + \dots + y_{n-1})]$

% 3. Composite Simpson's Rule (for even n , step $h = (b - a) / n$)

$\int_a^b y\,dx \approx \frac{1}{3}h[(y_0 + y_n) + 4(y_1 + y_3 + \dots + y_{n-1}) + 2(y_2 + y_4 + \dots + y_{n-2})]$

% 4. Classical Fourth-Order Runge-Kutta (RK4) Method for $dy/dx = f(x, y)$

$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \quad k_1 = hf(x_n, y_n), \quad k_2 = hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2})$

6. Linear Algebra, Matrix Decompositions & Array Grammar

In LaTeX typesetting and Math Formula OCR, two-dimensional matrices are delimited by `\begin{pmatrix}` (parentheses), `\begin{bmatrix}` (square brackets), or `\begin{vmatrix}` (determinant vertical bars), with column entries separated by `&` and rows terminated by `\\` [2].

[2x2 & 3x3 DETERMINANTS, CRAMER'S RULE & SVD IN LATEX]

$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad \det(\mathbf{A}) = ad - bc \neq 0$

$\det \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$

$\det(\mathbf{A} - \lambda \mathbf{I}) = 0, \quad \mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

7. Probability Distributions & Statistical Estimators

Quantitative research and statistical examination datasheets rely on standard discrete and continuous probability distributions alongside sample covariance, rank correlation, and linear regression estimators [3]:

Distribution	Probability Mass / Density Function (LaTeX)	Mean $E(X)$	Variance $\text{Var}(X)$
Binomial $B(n, p)$	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson $Po(\lambda)$	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$	λ	λ
Geometric $\text{Geo}(p)$	$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \dots$	$1/p$	$(1-p)/p^2$
Continuous Uniform $U(a, b)$	$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$	$\frac{1}{2}(a+b)$	$(b-a)^2/12$
Exponential $\text{Exp}(\lambda)$	$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$	$1/\lambda$	$1/\lambda^2$
Normal (Gaussian) $N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2

Product Moment Correlation, Spearman's Rank Correlation & Least-Squares Regression:

Given bivariate observations $(x_1, y_1), \dots, (x_n, y_n)$, the summary statistics S_{xx} , S_{yy} , and S_{xy} , Pearson correlation r , and Spearman's rank correlation r_s are given by:

[BIVARIATE CORRELATION, SPEARMAN'S RANK & CHI-SQUARED ESTIMATORS IN LATEX]

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}, \quad r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad \chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

8. Standard Statistical Reference Tables

If Z has a standard normal distribution with mean 0 and variance 1, $Z \sim N(0, 1)$, then the cumulative distribution function $\Phi(z) = P(Z \leq z)$ and the upper-tail percentage points z where $P(Z > z) = p$ are tabulated below [3]:

8.1 Cumulative Normal Distribution Function $\Phi(z) = P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

8.2 Percentage Points of the Standard Normal Distribution

If Z has a normal distribution with mean 0 and variance 1, then for each probability p , the table gives the value of z such that $P(Z > z) = p$ (one-tailed) and $P(|Z| > z) = 2p$ (two-tailed):

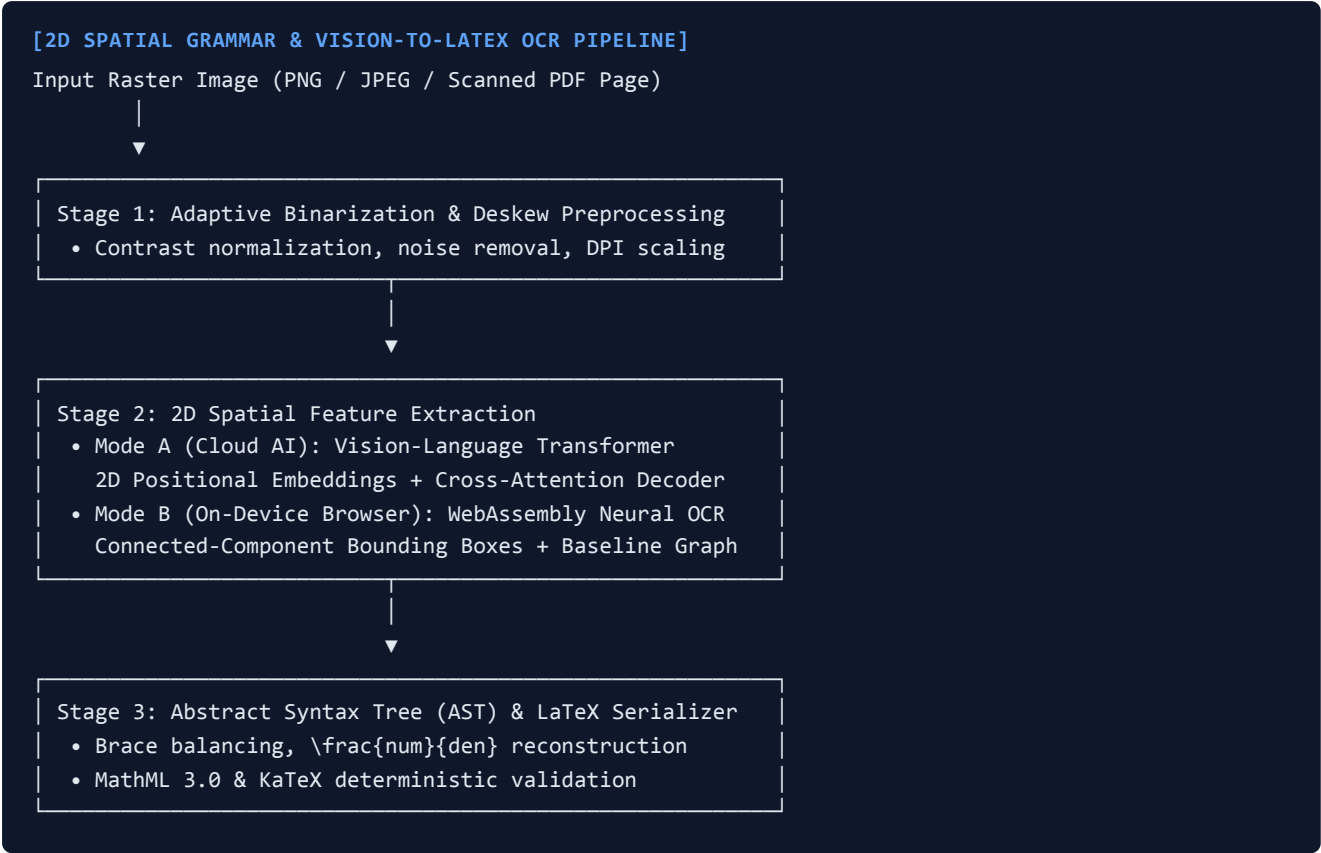
Probability p (Upper Tail)	0.5000	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
Critical Value z	0.0000	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

8.3 Critical Values of Student's *t*-Distribution & Chi-Squared (χ^2) Distribution

Degrees of Freedom (<i>v</i>)	<i>t</i> _{0.90} (1-Tail 0.10)	<i>t</i> _{0.95} (1-Tail 0.05)	<i>t</i> _{0.975} (1-Tail 0.025)	<i>t</i> _{0.99} (1-Tail 0.01)	$\chi^2_{0.95}$ (Upper 0.05)	$\chi^2_{0.99}$ (Upper 0.01)
<i>v</i> = 1	3.078	6.314	12.706	31.821	3.841	6.635
<i>v</i> = 2	1.886	2.920	4.303	6.965	5.991	9.210
<i>v</i> = 5	1.476	2.015	2.571	3.365	11.070	15.086
<i>v</i> = 10	1.372	1.812	2.228	2.764	18.307	23.209
<i>v</i> = 20	1.325	1.725	2.086	2.528	31.410	37.566
<i>v</i> = 30	1.310	1.697	2.042	2.457	43.773	50.892
<i>v</i> = ∞ (Normal)	1.282	1.645	1.960	2.326	–	–

9. Computational Architecture of Mathematical Formula OCR

Transforming a raster image of a mathematical equation into valid LaTeX or MathML requires solving three distinct computational sub-problems: (1) visual glyph recognition across multi-scale fonts; (2) two-dimensional spatial relationship parsing (above, below, superscript, subscript, inside-radical, matrix-cell); and (3) syntactic serialization into balanced LaTeX tokens [1, 2].



9.1 Structural Baseline and Bounding-Box Heuristics

In local geometric layout analysis, each detected connected component C_i is assigned a bounding box $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$ and normalized centroid (x_c, y_c) . Spatial relations between an anchor symbol A and an adjacent symbol B are classified by vertical overlap ratios:

- **Superscript Relation:** $y_c(B) < y_{\min}(A) + 0.35 \times H(A)$ (in top-left origin image coordinates), serialized as $A^{\{B\}}$.
- **Subscript Relation:** $y_c(B) > y_{\min}(A) + 0.65 \times H(A)$, serialized as $A_{\{B\}}$.
- **Horizontal Fraction Bar:** A high-aspect-ratio horizontal stroke ($W/H > 6.0$) partitions components strictly above its vertical span into the numerator subgraph and components strictly below into the denominator subgraph, serialized as $\frac{\text{Numerator}}{\text{Denominator}}$.

10. Hybrid Privacy-Preserving Execution Topology

Scientific manuscripts, proprietary engineering derivations, and financial valuation models frequently contain confidential intellectual property. At the same time, complex multi-line matrix derivations and degraded camera scans benefit from large-parameter vision-language models. A transparent, user-controlled architecture resolves this tension by offering two explicit execution modes [5, 6, 8]:

Architectural Dimension	Mode A: Ephemeral AI Cloud OCR	Mode B: 100% On-Device Browser OCR
Execution Environment	Stateless GPU inference endpoint over TLS 1.3 encrypted transport.	Client-side Web Worker + WebAssembly (WASM) inside browser RAM sandbox.
Data Retention Policy	Zero-Retention / In-memory ephemeral inference; no file or database storage.	Zero network transmission; image bytes never leave the local device.
Optimal Workload	Complex multi-line calculus, matrices, nested fractions, and low-DPI scans.	Confidential equations, offline environments, and linear algebraic expressions.
Output Formats	Canonical LaTeX, MathML 3.0, Markdown, and live rendered equation preview.	Local text/symbol extraction with client-side LaTeX structural heuristic formatting.

When operating in the on-device browser mode, image buffers are transferred to a dedicated background Web Worker using zero-copy `ArrayBuffer` transfers so the main UI thread remains responsive [5, 6]:

```
// Client-Side Web Worker Execution for On-Device Formula OCR

async function runLocalBrowserFormulaOcr(imageFile) {
  const bitmap = await createImageBitmap(imageFile);
  const canvas = new OffscreenCanvas(bitmap.width, bitmap.height);
  const ctx = canvas.getContext("2d");
  ctx.drawImage(bitmap, 0, 0);

  // Execute neural OCR entirely within local browser memory (Zero Network Egress)
  const imageData = ctx.getImageData(0, 0, bitmap.width, bitmap.height);
  const rawRecognition = await recognizeLocallyInWebWorker(imageData);
  return normalizeToCanonicalLatex(rawRecognition);
}
```

11. LaTeX AST Normalization & MathML Accessibility

Raw sequence outputs from optical character recognition models often contain syntactically equivalent but non-canonical token sequences (for example, $\{x\}_{1}^{2}$ vs. x_1^2 , or unmatched `\left(` delimiters). A deterministic post-processing stage parses the token stream into an Abstract Syntax Tree (AST) to enforce balanced braces, canonicalize operator names (`\sin`, `\cos`, `\lim`, `\ln`), and compile accessible W3C MathML 3.0 markup for screen readers and academic publishing workflows [1, 2, 4].

12. Benchmark Evaluation Metrics for Math Formula OCR

Unlike plain-text OCR, which is evaluated primarily by Character Error Rate (CER) or Word Error Rate (WER), mathematical formula recognition must be evaluated on both token-level accuracy and two-dimensional structural fidelity:

- **Expression Recognition Rate (ExpRate):** The percentage of mathematical expressions whose compiled AST matches the ground-truth mathematical expression with zero structural or symbol errors.
- **Tree-Edit-Distance-Based Similarity (TEDS):** Measures structural similarity between the predicted and ground-truth operator trees (particularly critical for matrices and tabular statistical datasheets): $\text{TEDS}(T_a, T_b) = 1 - \frac{\text{EditDist}(T_a, T_b)}{\max(|T_a|, |T_b|)}$.
- **Token-Level BLEU-4 Score:** Evaluates n -gram overlap of normalized LaTeX command tokens against canonical AMS-LaTeX ground truth.

Architectural Principle: A mathematical OCR system must transparently disclose whether an equation image is processed locally inside the browser's WebAssembly sandbox or via an encrypted zero-retention cloud inference pipeline.

13. Conclusion

This reference guide has presented a unified datasheet of fundamental mathematical formulae, vector geometry equations, calculus identities, probability distributions, and standard normal, Student's t , and

chi-squared statistical tables paired with their canonical LaTeX encodings. By combining standardized mathematical notation with a transparent, dual-mode Optical Character Recognition (Math Formula OCR) architecture, researchers, educators, and engineers can accurately convert printed and scanned quantitative documents into structured LaTeX and MathML while maintaining strict control over data privacy and computational provenance [1, 3, 5, 9].

References

- [1] World Wide Web Consortium (W3C). *Mathematical Markup Language (MathML) Version 3.0 (2nd Edition)*. W3C Math Working Group, W3C Recommendation. Available: <https://www.w3.org/TR/MathML3/>
- [2] American Mathematical Society (AMS). *User's Guide for the amsmath Package (Version 2.1)*. American Mathematical Society TeX Resources. Available: <https://www.ams.org/tex/amslatex>
- [3] National Institute of Standards and Technology (NIST). *NIST Digital Library of Mathematical Functions (DLMF)*. U.S. Department of Commerce. Available: <https://dlmf.nist.gov/>
- [4] Unicode Consortium. *Unicode Technical Report #25: Unicode Support for Mathematics*. Unicode Standard Technical Report. Available: <https://www.unicode.org/reports/tr25/>
- [5] WebAssembly Community Group. *WebAssembly Core Specification — Version 2.0*. WebAssembly.org, W3C Recommendation. Available: <https://webassembly.github.io/spec/core/>
- [6] World Wide Web Consortium (W3C). *Web Workers Specification*. WHATWG / W3C Web Platform Working Group. Available: <https://html.spec.whatwg.org/multipage/workers.html>
- [7] International Organization for Standardization (ISO). *ISO 80000-2: Quantities and units — Part 2: Mathematical signs and symbols to be used in the natural sciences and technology*. Available: <https://www.iso.org/standard/64973.html>
- [8] World Wide Web Consortium (W3C). *Web Cryptography API*. W3C Web Cryptography Working Group, W3C Recommendation. Available: <https://www.w3.org/TR/WebCryptoAPI/>
- [9] Utilvo. (2026). *Utilvo Math Formula OCR: Client-side and zero-retention optical character recognition for mathematical equations and LaTeX conversion*. Available: <https://utilvo.com/pdf-tools/ocr-formula>