参数（Parameter）：指描述总体的数值

统计量（Statistic）：指描述样本的数值

称名量表（Nominal）顺序量表（Ordinal）

等距量表（Interval）：零点是人为设定的，不代表 “零含量”摄氏温标，海拔，经度，公元纪年，净收入，势能

等比量表（Ratio）：零点代表变量的零含量 长度（身高），面积，速度，重量，体积，反应时间，开氏温标，年龄

控制组（Control Condition）：组内个体不接受实验处理（或接受安慰剂处理），为实验组提供比较基准。

总体 N 样本 n 组距数量通常控制在 **10** 个左右

处理连续型变量：表观界限（Apparent Limits） 真实界限（Real Limits）+-0.5 (组距宽度=上限-下限+1)

多边形图（Polygon）：图形两端需“延伸至频数为 0 的位置”/常用平滑曲线代替直方图或多边形图展示分布

正偏态（Positively Skewed）：尾巴在右侧（分数集中在左侧，右侧有极端高分）众数（峰值位置）<中位数<均值（均值被极端高分拉高）

负偏态（Negatively Skewed）：尾巴在左侧（分数集中在右侧，左侧有极端低分）均值<中位数<众数（均值被极端低分拉低）

茎叶图（Stem-and-Leaf Display）

集中趋势：平均数 中位数 众数

离散程度：全距 四分位距 Q1（第一四分位数）：25% 的分数≤Q1，75% 的分数≥Q1/Q2（第二四分位数）：即中位数，50% 的分数≤Q2 方差和标准差 离均差/离均差平方和（SS） $s^2 = \frac{\sum(X-M)^2}{n-1} = \frac{SS}{df}$

Z 分数：形状不变，均值为 0，标准差恒为 1

二项分布： μ=np σ^2=np(1−p)

累积分布函数 CDF：随机变量 X 取值小于或等于 x 的概率 非递减性

连续变量的 CDF 是“连续曲线”，离散变量的 CDF 是“阶梯状曲线”

概率密度函数 PDF：曲线在任意两点 a 与 b 之间的面积，等于随机变量 X 落在 a 与 b 之间的概率，总面积为 1，单个点概率为 0

概率质量函数 PMF：随机变量 X 取 x 值的概率

指数分布：无记忆性 $f(x) = \lambda e^{-\lambda x}\ (x>0,\ \lambda>0)$

$E(X) = 1/\lambda,\ Var(X) = 1/\lambda^2$

泊松分布： $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (k为事件发生次数,

$E(X) = \lambda,\ Var(X) = \lambda$

正态分布：PDF 公式 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},\quad -\infty < x < +\infty$

在$x = \mu - \sigma$和$x = \mu + \sigma$ 出现拐点 方差小的曲线更瘦高

二项分布的正态近似：npq>5 np>10 且 nq>10 当 p 接近 0 或 1，或者 n 较小的时候，近似效果差（注意连续性矫正，+-0.5）

样本均值分布：样本均值分布的均值等于总体均值 n 足够大时近似服从于标准差为标准误的正态分布

大数定律：样本量 n 越大，样本均值 M 越可能接近总体均值 μ。样本均值的 z 分数是除以标准误。标准误常用“SE”或“SEM”（Standard Error of the Mean）表示

虚无假设（Null Hypothesis, H_0）

备择假设（Alternative Hypothesis, H_1）

α 水平（显著性水平，Alpha Level）：判断样本结果是否属于‘低概率区域’的临界标准

临界区域（又称“拒绝域，Rejection Region”）：零假设为真时极不可能出现的结果集合

P 值：当零假设（H_0）为真时，观测到当前样本所呈现的结果（如样本均值、差异值等），或比当前结果更极端的结果出现的概率（双侧要乘 2）

假设检验的结论仅为“拒绝 H_0”或“无法拒绝 H_0”

I 类错误（拒真错误）：α。零假设为真（无处理效应），但样本数据错误地显示“存在处理效应”

II 类错误（取伪错误）：β。零假设为假（有处理效应），但样本数据未显示“存在处理效应”

检验力（Power）：零假设为假（有处理效应）时，检验正确拒绝 H_0的概率

$Cohen's\ d = \frac{|\mu_{处理后} - \mu_{处理前}|}{\sigma}$ 0.2 小/0.5 中/0.8 大

$n = \frac{\sigma^2 \cdot (z_{1-\beta} + z_{1-\alpha})^2}{(\mu_0 - \mu_1)^2}$ 单 $n = \frac{\sigma^2 \cdot (z_{1-\beta} + z_{1-\alpha/2})^2}{(\mu_0 - \mu_1)^2}$ 双

双侧检验的检验力略低于单侧检验（需更大样本量）

最小方差无偏估计量（MVUE）样本均值

df 越小 t 分布尾部越厚（极端值概率更高）峰部越低。

自助法（Bootstrap）：通过对原始样本进行重抽样，推导参数（如均值、中位数、相关系数等）置信区间。不适用场景：原始样本存在严重偏差/原始样本量过小

偏度(skewness)：衡量单峰分布的对称性，反映分布尾部的偏斜方向与程度。偏度<0：左偏分布 偏度>0：右偏分布 <0.5 基本对称 <1 中等偏斜 >1：高度偏斜 样本量越大，偏度的标准误（SE）越小，越容易检测到 “轻微偏斜” $\gamma_1 = \frac{E[(X-\mu)^3]}{\sigma^3}$

峰度（Kurtosis）：$Kurt[X] = \frac{\mu_4}{\sigma^4} = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2}$ 衡量分布的尾部厚度（即极端值出现的概率），而非分布的陡峭程度。=3 正态 >3 尾部更厚数据更集中极端值更多 <3 尾部更薄数据更分散极端值更少

箱线图：箱体（Box）：上下边界分别为 Q3 和 Q1，箱内横线为中位数 Q2 须（Whiskers）：从箱体边界延伸至 “非异常值的最远数据点”，两侧须的长度可能不同 异常值：用单独的点（如 “○” 或 “×”）标注在须的外侧。