

第一章 绪论

第一节 医学统计学的作用

医学统计学的核心价值：在高度变异的医学世界中，区分“真实效应”与“随机波动”。

医学统计学（medical statistics）是一门服务于临床医学、基础医学、公共卫生与卫生服务研究的基础学科，其本质不是算数，而是基于数据进行科学推断（statistical inference）。

■ 医学统计学的研究对象与任务

- 研究对象：

医学中**结果具有不确定性（uncertainty）**的现象
→ 生物个体差异大、影响因素复杂、随机因素不可避免

- 核心任务：

- 收集数据（data collection）
- 分析数据（data analysis）
- 由样本推断总体（inference from sample to population）

👉 关键前提：

样本结果 ≠ 真实规律

统计学的使命是判断：

「观察到的差别，是真实效应，还是偶然波动？」

✿ 为什么医学研究离不开统计学？

以临床常见问题为例：

- 两种治疗冠心病的药物
 - 试验药：有效率 87%
 - 对照药：有效率 82%

？问题不在于“谁高”，而在于：

- 这 5% 的差别：
 - 是药物真实疗效差异？
 - 还是样本量有限下的随机误差 (random variation) ？

👉 仅凭样本结果，无法直接下结论

必须借助统计方法判断差别是否“具有统计学意义 (statistical significance)”。

💡 经典案例 1：没有统计设计 → 错误医学结论

背景

- 1962 年，JAMA 报道：
 - 👉 胃冷却（向胃内导入冷冻液）可缓解胃溃疡
- 依据：
 - 动物实验
 - 24 名患者的临床观察

致命问题

- ✗ 未合理设置对照组
- ✗ 研究设计存在系统性偏倚 (bias)

后果

- 该方法被广泛应用
- 后续**严格随机对照试验 (randomized controlled trial, RCT) **证明：
 - 仅短暂缓解疼痛
 - ! 不能治疗溃疡，甚至可能加重病情

👉 统计学启示：

没有正确设计的研究，数据越“真实”，结论可能越危险。

经典案例 2：统计推断让“可能性”变成“证据”

问题背景

- 观察性研究提示：
 孕期补充维生素可能降低新生儿神经管缺陷
- 但始终无法证实 (confounding 无法排除)

关键研究

- 1991 年，英国医学委员会开展 **大规模 RCT**

结果数据

组别	新生儿数	NTD 例数
服用叶酸组	593	6
未服用叶酸组	602	21

统计学结论

- 两组差异 **不是偶然出现**
- 可以归因于**叶酸的作用**

 **统计学在此完成的不是“计算”，而是“因果判断的可信化”。**

从两个案例中引出的统计学核心问题

医学研究中反复出现的关键问题包括：

1. **?** 为什么早期胃溃疡研究会得出错误结论?
→ **研究设计问题**
2. **?** 为了证明叶酸的作用，应如何分组?
→ **随机分组、对照原则**
3. **?** 如何估计两组 NTD 的**总体发病率**?
→ **点估计与区间估计 (estimation)**
4. **?** 如何判断两组发病率是否有差别?
→ **假设检验 (hypothesis testing)**

👉 这些问题，都不是直觉能解决的。

⌚ 一句话总括本节

医学统计学研究的不是“确定的数字”，而是“不确定世界中的可靠结论”；
它通过分析数据的偶然性，揭示医学现象背后的规律性。

📘 对医学工作者的意义

掌握医学统计学，意味着你能够：

- 设计可信的医学研究
- 选择合适的统计分析方法
- 正确解读医学文献
- 撰写经得起质疑的科研论文
- 避免“数据在，结论错”的灾难

第二节 医学统计学的基本内容

医学统计学是一条完整工作流：从“如何收集数据”到“如何对不确定性下结论”。

医学统计学的基本内容由**四个相互衔接的模块**构成：

统计设计 → 数据整理与核查 → 统计描述 → 统计推断

任何一个环节出问题，后面的分析都会失去意义。

🧠 医学统计的整体逻辑图（先有全局）

设计决定上限，数据决定下限，推断决定结论。

- **统计设计**：决定研究是否“值得做”
- **数据整理**：决定数据是否“可信”
- **统计描述**：决定结果是否“看得懂”

- **统计推断**: 决定结论是否“站得住”
-

1 统计设计 (statistical design)

统计分析不能修复设计错误。

研究类型的根本区分

- **实验性研究 (experimental study)**
 - 人为干预 (如给药、手术)
 - 因果推断能力强
- **观察性研究 (observational study)**
 - 自然发生、无干预
 - 易受混杂因素影响

设计的“两条线”

- **专业设计 (医学问题本身)**
 - 研究目的与选题
 - 研究对象
 - 处理因素 / 干预措施
 - 观察指标 (outcome)
- **统计设计 (如何让结论可信)**
 - 分组 / 抽样方法
 - **样本量估计 (sample size)**
 - 数据管理与质量控制
 - 拟采用的统计分析方法

👉 核心原则:

设计阶段的错误，在分析阶段**不可逆**。

统计学不是“算后补救”，而是“事前保障”。

2 数据整理与核查 (data management & checking)

干净的数据，比复杂的方法更重要。

数据整理的基本操作

- 分类编码 (categorical coding)
- 数值编码 (numeric coding)
- 建立数据结构 (变量定义)

数据核查的三类“必查点”

① 逻辑检查 (logic check)

- 如：男性的“绝经年龄”
→ 明显逻辑错误

② 有效性检查 (validity check)

- 变量是否超出定义范围
- 如性别变量出现非法数值

③ 异常值处理 (outlier handling)

- 面对离群值，先问：
 - 是录入错误？
 - 还是真实但罕见的现象？

👉 重要原则：

没有充分理由，不应随意剔除“看起来不正常”的数据。

过度清洗 ≠ 高质量数据

有时，新发现正藏在异常值里（如基因突变）。

3 统计描述 (descriptive statistics)

统计描述回答的是：数据“长什么样”。

目的

- 概括大量数据的主要特征
- 指出后续分析的方向

三种核心表达方式

① 统计指标 (statistical measures)

- 用少量数字概括数据特征
- 如：
 - 平均水平 (mean, median)
 - 变异程度 (SD, IQR)

② 统计表 (tables)

- 替代冗长文字
- 便于比较与分析
- 是论文与报告的基础结构

③ 统计图 (figures)

- 直观、形象
- 快速传递信息
- 正确使用对论文质量影响极大

👉 关键提醒：

描述 ≠ 推断
统计描述不能回答 “是否有差异” ，
只能回答 “差异长什么样” 。

4 统计推断 (statistical inference)

统计推断，是从“样本”走向“总体”的桥梁。

统计推断包括 **参数估计** 和 **假设检验** 两大类。

◆ 参数估计 (parameter estimation)

① 点估计 (point estimation)

- 给出一个“最可能的值”

② 区间估计 (interval estimation)

- 给出一个可信范围
- 同时量化“不确定性”

👉 医学意义

区间估计不仅告诉你“差多少”，
还告诉你“你有多不确定”。

例

- 舒张压差值的 95% 置信区间：
10.41 ~ 13.05 mmHg
- 含义：
 - 真值落在区间外的概率 $\approx 5\%$

◆ 假设检验 (hypothesis testing)

核心问题

观察到的差异，是否可能只是随机误差？

基本思想

- 构建“无差别”的原假设 (null hypothesis)
- 计算观察结果在该假设下出现的概率

P 值的角色

- $P < 0.05$:
 - 差异由随机波动造成可能性较小

- 支持“存在真实差异”

❖ 重要认知

- P 值 \neq 效果大小
 - P 值 \neq 临床意义
 - 它只回答一个问题：
“这个差异像不像随机误差？”
-

▣ 常见统计方法（按问题选工具）

- t 检验
- 方差分析 (ANOVA)
- χ^2 检验
- 非参数检验
- 相关分析
- 回归分析

👉 没有“万能方法”

方法的选择取决于：

- 数据类型
 - 研究设计
 - 研究目的
-

⌚ 一句话总括本节

医学统计学不是一堆检验方法，而是一条从研究设计到科学结论的完整逻辑链。

第三节 医学统计学中的基本概念

医学统计学处理的核心矛盾：生命体天然“变异”，但我们仍想从样本推出总体的规律。

一、同质与变异 (homogeneity & variation)

同质是“可比的前提”，变异是“统计学存在的理由”。

█ 同质 (homogeneity) 是什么？

- 指观察单位/个体在**主要研究指标的关键影响因素**上相同或相近
- 例：研究“某药治疗高血压”
 - 若药物针对**原发性高血压** → 原发性高血压患者才算“同质”
 - 肾病引起的继发性高血压 → **不同质**（机制不同、反应可能不同）
 - 若疗效受病情轻重影响 → 只能把**病情相同**的人视为同质观察单位

◆ 观察单位 (observation unit)

- 研究对象的基本单元：一位患者 / 一只动物 / 一份样品……

✿ 变异 (variation) 是什么？

- 同一测量在总体中不同个体间的差别
- 例：同年龄同种族同性别健康人，脉搏/体温/呼吸也可能差异很大
- 例：同药同病情，疗效仍可能不同

👉 统计学的角色

- 变异往往来自未控制/无法控制/未知因素造成的随机波动
- 所以不能“看 1-2 例有效就下结论”，必须依靠统计从大量偶然中找必然

二、变量与数据类型 (variable & data types)

数据类型决定你能用什么统计方法。

█ 变量 (variable) 与数据 (data)

- 变量：研究对象特征（随机变量，random variable）的简称
- 数据：变量的观测值

- 例：体温/身高/体重/性别/血型/疗效都是变量；测出来的结果就是数据

1) 定量数据 (quantitative data, 计量资料)

- 观测值是**数值**，描述数量特征，通常有单位
- 连续型**：理论上可取无限多值（身高、体重、血压、温度）
- 离散型**：多为整数计数（家庭成员数、脉搏、白细胞计数）
- 医学中常不强行区分，统称定量数据

2) 定性数据 (qualitative data, 计数资料)

- 观测值是**类别/属性**，类别互斥
- 例：性别（男/女），血型（A/B/O/AB）
- 可数字编码，但**数字不代表“量”**

3) 有序数据 (ordinal data, 等级/半定量资料)

- 也是定性，但类别之间有**顺序/程度**
- 例：尿糖 “-、+、++、+++”；疗效 “显效、有效、好转、无效”
- 可编码 1,2,3,...,k
- 能比大小，但差多少不确定**（1 和 2 的差不等于 3 和 4 的差）

▣ 数据类型可以转换 (按分析目的)

- 血红蛋白（定量）
 - 分“正常/异常” → 二分类定性
 - 分“正常/轻/中/重度贫血” → 有序数据
- 👉 但转换会丢信息：把连续变成分组通常降低统计效率 (power)。

三、总体与样本 (population & sample)

统计推断的本质：用样本去近似总体。

▣ 总体 (population)

- 研究对象的全体：由**同质观察单位**组成
- 关注的是总体的一个或多个特征（可视为一维/多维变量的集合）

样本 (sample)

- 从总体中抽取的**有代表性的一部分** (通常要求随机抽取)
- 例：社区随机抽 200 名成年男性，测 5 个变量 (身高体重血压脉搏心率)
 - 总体：该社区全部成年男性的这些生命体征
 - 样本：200 人的测量结果

有限总体 vs 无限总体

- 有限总体**：对象有明确范围 (如两个社区的冠心病患者)
- 无限总体**：无时间空间限制的同质人群 (如“所有心梗且接受尿激酶原治疗的患者”)

参数 (parameter) vs 统计量 (statistic)

- 参数：描述总体特征的指标 (总体均值、总体率...)
- 统计量：由样本算出来的指标 (样本均值、样本率...)

👉 关键要求：先严格定义总体范围，再随机化抽样，才能保证代表性与推断可信。

四、误差 (error)

误差 = 观测值偏离真实值；统计量偏离参数。

1) 系统误差 (systematic error)

- 来自固定因素，**有方向性**，大小恒定或按规律变化
- 例：仪器未归零、试剂校准不准、读数偏向；入组不当、疗效标准把握不准
- ✅ 可通过：研究设计更严密、测量标准化等控制/消除

2) 随机测量误差 (random measurement error)

- 偶然因素导致同一对象重复测量不一致
- 多由生物自然变异与不可预知因素造成
- 无固定方向**，但有统计规律 (如近似正态)
- ✅ 不可避免，但可用重复测量更准确估计真值

3) 抽样误差 (sampling error)

- 最重要的随机误差：样本统计量 \neq 总体参数
 - 根源：个体变异
 - 规律：**样本量越大，抽样误差越小**，统计量越接近参数
- ✓ 可用统计方法量化（标准误、置信区间、检验等）
-

五、概率与概率分布 (probability & distribution)

概率是“不确定性”的语言，分布是“概率如何分配到各种结果”的地图。

■ 概率 (probability)

- 描述随机事件发生可能性： $0 \leq P(A) \leq 1$
- 越接近 1 越可能；0 不可能；1 必然

两种常见定义

- **古典概率**： $P(A) = m/n$, 要求基本结果等可能
- **统计概率 (频率解释)**：当样本量很大时， $P(A) \approx n_A/n$
 - 例：调查 20000 人，高血压 4000 \rightarrow 患病率 20%
 - 随机抽一名成年人，估计其患病概率约 20%

👉 医学研究结果不确定，所以经常用概率表达；同时样本统计量的“随机性”也用概率刻画，这是推断基础。

■ 概率分布 (probability distribution)

- 随机变量所有可能取值与其概率的对应关系
- 例：红细胞计数在不同区间的概率；四种血型各自的概率

❖ 小概率事件与 P 值直觉

- 习惯上 $P \leq 0.05$ 称“小概率事件”
- 在假设检验里：
 - 即便两药总体效果相同，样本均数也会因抽样误差不同
 - 若计算得到 $P \leq 0.05$ ，意味着当前差异由随机误差导致的可能性很小
 - 因而支持“两药疗效有差别”的推断

⌚ 一句话总括本节

同质让比较成立，变异让统计必要；变量类型决定方法；总体/样本决定推断框架；误差决定可信度；概率与分布提供推断语言。

第四节 概率的基本运算法则

概率运算的本质：把“多重不确定性”组合成一个可判断的风险。

这一节不是为了让你记公式，而是为了让你在后面理解：

- P 值从哪来
 - 联合风险怎么估
 - 为什么筛查阳性 ≠ 真有病
-

1 概率的乘法法则 (multiplication rule)

当事件彼此独立时，“同时发生” = “分别发生的概率相乘”。

■ 独立事件 (independent events)

- 事件 A 是否发生 不影响 事件 B 发生的概率

△ 公式

$$P(AB) = P(A) P(B)$$

UrlParser 医学例子 (独立假设下)

- 40 岁以上人群
 - 2 型糖尿病概率 $P(A) = 22.5\%$
 - 甲状腺结节概率 $P(B) = 20.2\%$

若两病互不影响：

$$P(AB) = 0.225 \times 0.202 \approx 0.045$$

👉 含义：约 4.5% 的人同时患两种病

📌 关键提醒

- 乘法法则成立的前提是**独立**
 - 在医学中，“真正独立”的情况并不多，**必须先判断合理性**
-

2 概率的加法法则 (addition rule)

“至少发生一个”的概率，要小心是否存在**重叠**。

情形一：互不相容 (mutually exclusive)

▣ 互不相容

- A 发生就不可能发生 B (如血型 A 与 B)

△ 公式

$$P(A \cup B) = P(A) + P(B)$$

💡 医学例子

- A 型血：29.7%
- B 型血：34.2%

$$P(A \cup B) = 29.7\% + 34.2\% = 63.9\%$$

情形二：不互斥（可同时发生）

这是医学里更常见的情况。

△ 公式

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

💡 医学例子

- 糖尿病: 22.5%
- 甲状腺结节: 20.2%
- 假设互不影响 (独立)

$$P(A \cup B) = 0.225 + 0.202 - 0.225 \times 0.202 \approx 0.382$$

👉 含义: 约 38.2% 的人至少患有其中一种病

📌 直觉解释

- 直接相加会把“同时患两病的人”算两次
 - 所以要减去交集
-

3 条件概率 (conditional probability)

条件概率回答的是: 在“已知一件事发生”的前提下, 另一件事发生的可能性。

☰ 定义

- $P(B | A)$: 在 A 已发生的条件下, B 发生的概率
- $P(A | B)$: 在 B 已发生的条件下, A 发生的概率

▷ 联合概率的一般形式

$$P(AB) = P(A) P(B | A)$$

或

$$P(AB) = P(B) P(A | B)$$

💡 医学例子 (不独立的疾病)

- 糖尿病: $P(A) = 12.5\%$
- 高血压: $P(B) = 20.2\%$
- 糖尿病患者中 65.2% 合并高血压

$$P(AB) = 0.125 \times 0.652 = 0.0815$$

👉 约 8.15% 的人同时患两病

进一步：

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{0.0815}{0.202} \approx 0.404$$

👉 高血压患者中，约 40% 合并糖尿病

⭐ 重要直觉

- $P(A | B) \neq P(B | A)$
 - 医学中最常见的误判之一就是把二者混为一谈
-

4 Bayes 公式 (Bayes' theorem)

Bayes 的本质：用“检测结果”反推“真实状态”的概率。

█ Bayes 公式

$$P(A | B) = \frac{P(B | A) P(A)}{P(B | A) P(A) + P(B | \bar{A}) P(\bar{A})}$$

- $P(A)$: 先验概率 (prior)
 - $P(A | B)$: 后验概率 (posterior)
-

💡 经典医学筛查陷阱：肝癌例子

已知

- 肝癌患病率： $P(C) = 0.4\% = 0.004$
- 患癌者联合检测阳性： $P(A | C) = 85\%$
- 健康者检测阴性： $90\% \rightarrow P(A | \bar{C}) = 10\%$

△ 计算

$$P(C | A) = \frac{0.85 \times 0.004}{0.85 \times 0.004 + 0.10 \times 0.996} \approx 0.033$$

👉 检测阳性的人中，真正患肝癌的概率只有 3.3%

🔴 极其重要的医学直觉

- 灵敏度高 \neq 阳性就是真的
- 特异度高 \neq 筛查就可靠
- 基线患病率（先验概率）极低时，假阳性会淹没真阳性

👉 这是医学诊断、筛查、过度医疗争议的**核心数学原因**

🧠 本节在医学统计中的真正位置

这些公式**以后很少直接计算**，
但它们构成了以下概念的底层逻辑：

- P 值
 - 置信区间
 - 假设检验
 - 筛查试验评价（灵敏度 / 特异度 / 阳性预测值）
-

⌚ 一句话总括本节

概率运算法则不是算概率，而是教你如何在不确定世界中避免“直觉型错误判断”。