

麻醉中的统计学：（第一部分）

每周麻醉教程辅导 302

2014年1月20日

Sameer Hanna-Juma

英国南安普顿医院国家医疗服务基金会

联系方式: Samnanna76@hotmail.com



问题

在继续辅导之前，先尝试回答下面的问题。答案在文章的末尾可以找到。

1. 下面哪一个表述是正确的？
 - a. 由于机遇导致的差异称为 I 型错误，在研究设计中，这类错误主要是统计学误差。
 - b. 柱状图对阐述数据集分布的模式很有用
 - c. II 型错误是指存在真正的差异，但统计学上并未显示有差异
 - d. 在正态分布中，均数，中位数以及众数是相等的
 - e. 百分位数图通常用于阐述正态分布数据

2. 下面哪一个是正确的
 - a. 平均数的标准误(SEM)是分布的测量方式
 - b. 需要治疗的人数(NNT)是相对危险度减少率的倒数
 - c. 随着样本量的增加，平均数的标准误会减少
 - d. 绝对危险度减少率等于干预组的百分比减去对照组的百分比
 - e. 样本量的大小不影响可信区间

引言

统计学是收集和分析大量数字资料的科学, 主要目的是通过具有代表性的样本来推断整体(人群)的比例。

本教程将涵盖与临床实践相关的统计学基础知识包括以下的主题

- 数据的种类
- 数据的收集
- 数据的展示
- 统计推断

数据的种类

数据是通过测量一系列变量后收集起来的观察值。变量是易变的元素、特征或要素。正确处理数据的第一步是识别数据的种类。它决定了统计数据的描述, 同时也是决定使用何种检验方法的第一步。数据主要分为定性资料和定量资料, 大多数研究同时包含了这两种数据类型。定量资料易于分析并且比较可靠, 而定性资料主要用于对样本更深入描述。

定性资料

定性资料是指不包含数值的变量。它通常描述某种含义并赋予变量某个名称或称号。这种称号可能是一个数字, 比如 ASA-1, 但是它并没有真正的数值。定性资料可以是有序的或只是名义(详见下图)

名义(或分类或属性)数据	有序(或等级或分级)数据
<ul style="list-style-type: none">• 类别间不重叠• 不包含数值• 如果只有两类比如男性或女性, 称为二分类数据• 通常用于描述样本的特征• 例如血型以及手术种类	<ul style="list-style-type: none">• 有意义的排序分类• 本质的等级是无法测量的• 通常用于调查和分析非数值变量• 例如 ASA 分级以及疼痛评分的分类

注意在疼痛评分的分类中: 无疼痛=0, 轻度疼痛=1, 中度疼痛=2, 重度疼痛=3, 其中中度疼痛只表示疼痛程度高于轻度疼痛, 并不是指中度疼痛程度是轻度疼痛的两倍。

定量资料

定量资料是指包含数值的变量。定量资料的种类有很多，但是它通常不影响统计检验方式的选择。定量资料可分为离散型资料或连续型资料如下图。

离散型	连续型
<ul style="list-style-type: none">• 数据通过有限的阶梯变化• 例如数值型的疼痛量表	<ul style="list-style-type: none">• 通过连续性测量得到的数据• 例如疼痛视觉模拟评分

连续型数据可以是等比资料或非等比资料。在等比量表中，变量零表示数量或特征的缺失，同时量表中任何部分单位数量的变化都代表相同的含义。但是在非等比量表中，单位数量的变化是非线性的。

数据的收集

收集数据的方法有很多。最常用的数据抽样方式包括单纯随机抽样以及分层抽样。

单纯随机抽样

真正的随机抽样是指人群中的每一个成员都具有相等的机会进入研究。这样才最能代表整体人群，但是在医学研究中往往难以做到真正的随机抽样。比如研究血流动力学监测对急诊开腹手术结局的影响，那么就应当可以纳入我们所选择的每一个急诊开腹手术患者。

分层随机抽样

在不同的亚组(层)中进行随机抽样。它能够使得样本在人群中具有更好的代表性并且减少混杂因素的影响。比如在上述的研究中，样本组首先被分为两个亚组-吸烟或不吸烟。然后再将患者随机分配到没有血流动力学监测装置的对照组以及有血流动力学监测装置的试验组。这可以减少吸烟作为混杂因素对结果所产生的影响，同时，两组纳入患者的数量也是相等的。其他的抽样方式还包括整群抽样，多级抽样，串联抽样以及多期抽样。

数据的展示

在研究过程中，所有的研究都收集了大量的原始数据。这些原始数据可以通过三种方式来展示：列表，图表以及数值（描述性统计）。每一种方式都有他的应用范围、优点以及缺点。

列表

临床研究中，频数表或者累积频数表常常用于数据的展示。它可以用于展示包括名义数据、有序数据以及定量资料在内的所有数据类型。为了使用这种方式能展示出连续型数据，首先必须将数据分配到不同的组别(区间)中。例如 ASA 评分以及研究人群年龄的分组

ASA 分级	频数(患者数量) n=390	年龄组	频数(患者数量) n=390	累积频数 (患者数量)
ASA-1	60	20-40 岁	24	24
ASA-2	230	41-60 岁	210	234
ASA-3	65	61-80 岁	126	360
ASA-4	32	81-100 岁	30	390
ASA-5	3			

相对频数表是展示以上同样数据的另外一种方式。表中数字表示该组患者占总人数的百分比

ASA 分级	相对频数%(每组患者百分比)n=390
ASA-1	15.3
ASA-2	58.9
ASA-3	16.6
ASA-4	8.2
ASA-5	0.77

频数表的另一种应用是交叉表以及数据的排序。

图表

图表非常重要，特别是在展示大量数据的时候。它能够突出不同变量间重要的关系，具有强烈的视觉效果。在临床统计中，常使用以下的图表：

饼图

饼图用于展示构成总体样本的不同组别所占的比例。整个圆饼代表总体样本，而每个组所占的大小是由他们所占总体样本的比例来决定的。饼图用于有序数据和名义数据。它能够突出显示研究样本潜在的不平衡以及潜在的混杂因素。

例如图 1 所示，不同手术类型中声门上通气设备的临床应用研究。



图 1 饼图

条形图

条形图用于比较数据的不同分类。x 轴通常是无量纲的，而 y 轴通常代表每个分类的频数。每一个分类可以代表单独的一个组，或者进一步分成不同的亚组。

例如图 2 所示，上述相同的数据通过条形图来展示。注意图中每一类都展示了性别亚组。

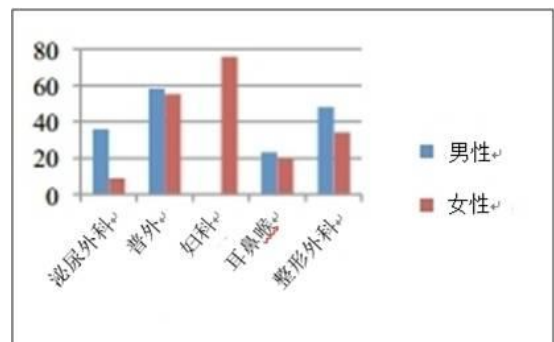


图 2 条形图

直方图

直方图是展示区间数据的特殊条形图。定量资料，特别是连续型数据，为了制作成频数表，通常会被划分成不同的区间。直方图在展示数据分布时很有用。它对统计分析的选择具有明显的影响，这将在文章的后面进一步讨论。在图 3 中我们可以明显的看到数据并非正态分布，否则

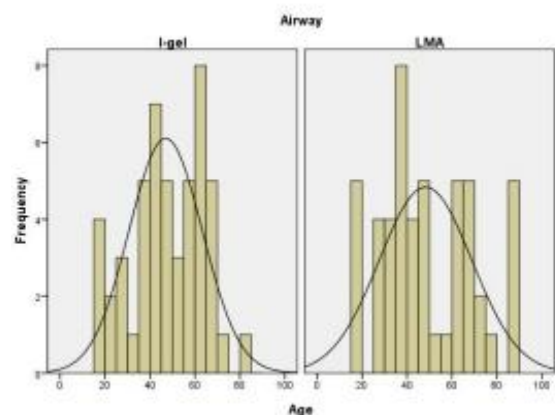


图 3 直方图

每一条形状的中点将会落在正态分布曲线上。因此统计分析方法也就与正态分布数

据的统计分析方法不同。此外，直方图在统计描述，例如均数、众数以及标准差中也很有用。

频数曲线

频数曲线与直方图很相似，但没有直条。它的应用与直方图也类似，但是频数曲线具有一个优势，就是在一个图中可进行 2 个或多个组间分布的比较。在图 4 中，我们可以比较试验组和对照组的血压的趋势。试验组中血压高于 140mmHg 的患者数量是小于对照组的。

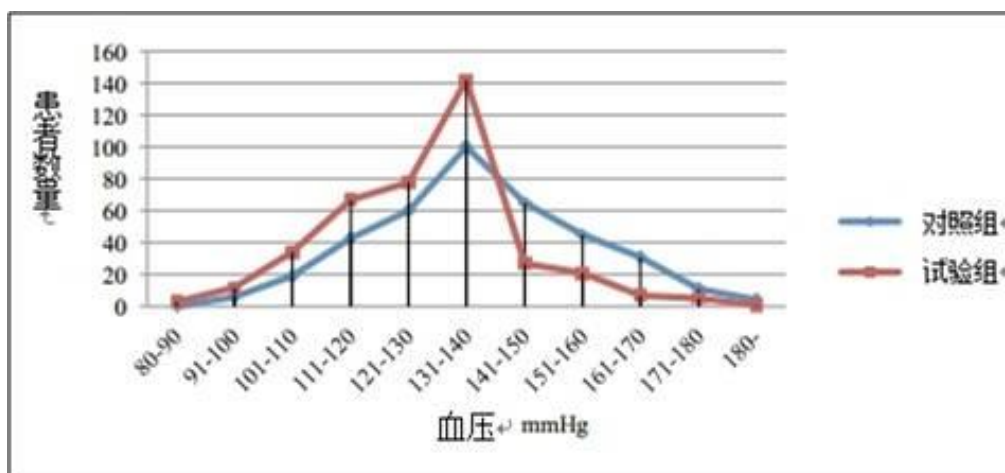


图 4 频数曲线

累积频数曲线(肩形图)

累积频数曲线是累积分布图，它在水平轴上显示数据值，在纵轴上显示累积频数、累积相对频数或者累积频数百分比。这类图主要用于识别低于或高于某种限制条件的样本成分。

散点图

散点图用来确定两个样本变量间是否存在关联。同时也可以用一个相关系数来计算出关联的强度。例如图 5 所示的研究是根据体重计算出的新型肌松药的给药剂量与为了达到某个 TOF 值而需要重复给

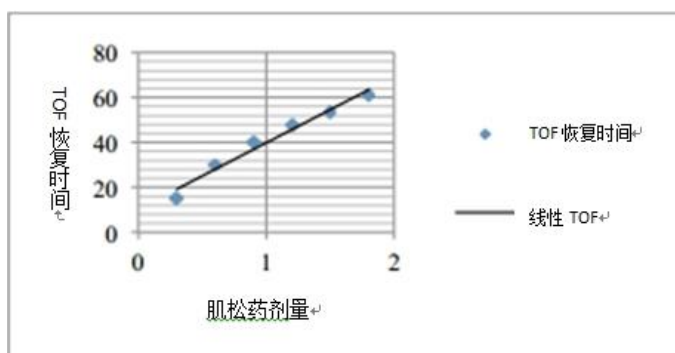


图 5 散点图

药的间隔时间之间的关系。数据显示的是增加的药量与 TOF 的恢复时间之间的直接关系。

- 直方图通常用于确定数据的分布类型，从而选择合适的统计分析方法
- 散点图是一种非常重要的视觉辅助方式，主要用于确定多变量分析中变量间的关联强度

数值(描述性统计)

虽然描述性统计是指可以通过列表、图表以及数值来展示数据，但是在医学研究中，主要还是指通过数值来表示。描述性统计的主要目的并不是要得到关于整体人群的某个结论，而是对抽样数据进行一个有意义的总结。在确定合适的统计分析方法进行推断分析之前，这是最基本的。尽管描述性统计在进行组间变量比较时能够提供强有力的总结分析，但是仍存在丢失重要细节信息的风险。描述性统计可用于单因素分析和多元分析。单因素分析的三个重要特征是分布，集中趋势和离散趋势。

分布

在卫生统计分析里，正态（高斯）分布十分重要，健康人群的生物数据(身高，体重，血压)是最常见的正态分布数据。虽然视觉辅助工具，例如直方图，可以用来确定数据的分布类型，但是依然有正式的统计检验方法来确定数据是否是正态分布，例如 Shapiro-Wilkes 检验以及 D'Agostino-Pearson 综合检验。正态分布(参数的)的特征是含有一个峰(单峰)并且两侧变量分布对称。在正态分布中，所有关于集中趋势的测量结果(均数，众数，中位数)应当是一致的，并且是最大频数对应的这个点。数据在峰值两侧的分布是相同的，可以用标准差来表示。两个参数(均数和标准差)能够完全描述出图 6 所示的曲线的形态。

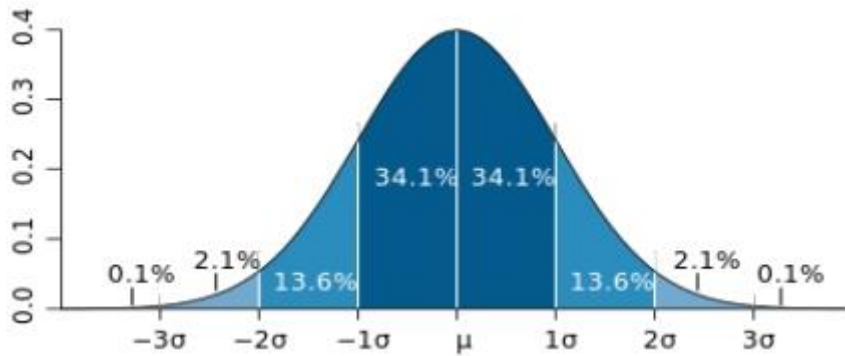


图 6 正态分布曲线。68%的人群处于均数 \pm 一个标准差的范围内。95%的人群处于均数 \pm 两个标准差的范围内。99.7%的人群处于均数 \pm 三个标准差的范围内。大多数关于医学概率的研究采用的是均数 \pm 两个标准差这个范围(95%人群)。

对于非参数数据，变量不是均匀地分布在集中趋势的顶点两侧的。可能在一侧分布较多，另一侧分布较少，我们称之为偏态(图 7)。数据也可能含有不止一个峰(多峰)。峰度是描述曲线顶峰的另一个名词，正态分布的峰度为零。曲线的顶峰更尖锐，尾部更长称为正峰度，而负峰度则指更宽更平坦的分布。在这种情况下，应当使用非参数检验进行统计分析。在分布模式不清楚的情况下，使用非参数检验更为妥当。

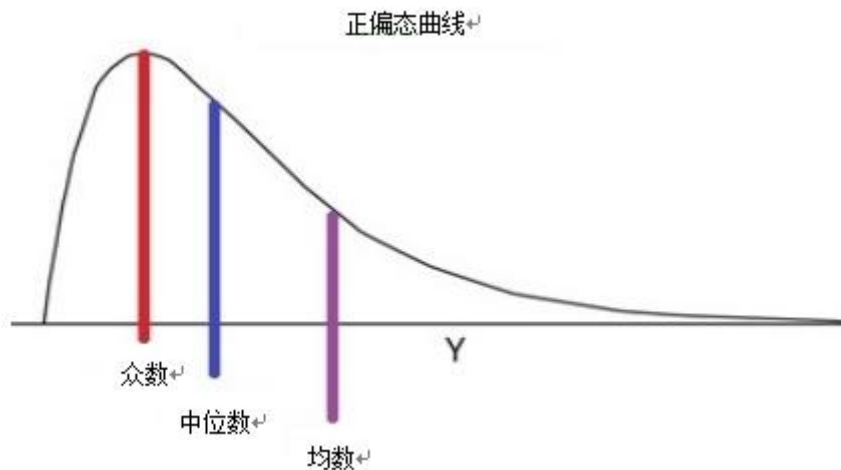


图 7 正偏态数据。注意众数，中位数和均数位于不同的方位。

集中趋势

集中趋势是对变量分布中心值的估计。集中趋势的三个指标分别是：均数，中位数和众数。

- 均数是指变量的总值除以变量的个数(算数均数)。它可以用于参数数据

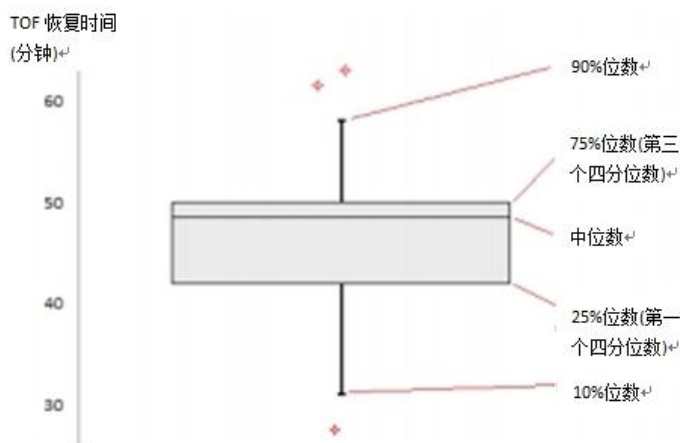
中集中趋势的描述，但不能用于有序数据和名义数据。

- 中位数是指将数据按观察值的大小进行排序后位置居中的观察值。这说明在这类数据中有 50% 的数值大于中位数，另有 50% 的数值小于中位数。在描述非参数数据的集中趋势时，我们更愿意选用中位数指标，因为相比较于均数而言，它受到异常值的影响更小。
- 众数是指在频数分布表中频数分布最多的那一组的观察值。它并不是描述集中趋势的一个很好的指标，但是在名义数据和有序数据中却是描述集中趋势唯一的指标。

离散趋势

极差是离散趋势中最简单的指标，但在实践中的应用却非常有限。极差是指数据中最大值和最小值的差值。在参数数据的离散趋势描述中，方差和标准差是最常用的指标。在计算这两个指标时需要用到总样本量数，因此这两个指标更精确。方差是每一个测量值与均数的差值的平方和除以样本总数得到的值。在计算方差时是经过平方计算的，这是由于在正态分布的数据中，测量值与均数的差值表现为正负差值会相互抵消。标准差是方差的平方根。百分位数(四分位数)是非参数数据中描述离散趋势的常用指标。四分位数顾名思义：有 25% 的数据小于第一个四分位数，第二个四分位数与中位数相等同时有 50% 的数据小于它，有 75% 的数据小于第三个四分位数。

百分位数通常在百分位数图即箱形图中表示(图 8)



在图例中，方块代表 50% 的样本(在使用了肌松药之后，50% 的患者的恢复时间在 42 到 50 分钟之间)。两条直线代表 10% 和 90% 位数，星号代表异常值(31 分钟后，有 10% 的患者肌松恢复，而 58 分钟后有 90% 的患者肌松恢复)。注意中位数并不在方块的中间，这说明该数据是非正态分布(在 42 到 48 分钟间以及 48 到 50 分钟间，肌松恢复的患者数量相等)。

图 8 百分位数图

统计推断

推断是从已知的或者假设的事实中通过逻辑分析得出结论的过程，并且假定样本人群能够代表整体人群。重要的是，我们需要知道这种假设是有程度之分的，这就引出了概率这个概念。统计推断可以是描述性的(样本均数代表总体均数)或者分析性的(研究样本人群中不同的变量间的关系并推广到总体人群中)。因此，统计推断中两个基本组成部分是样本和概率。

样本

纳入研究的这部分称为样本。例如，在研究围术期血流动力学监测对急诊开腹手术患者结局的影响中，在研究阶段纳入研究的患者就是样本，而所有需要行急诊开腹手术的患者就是总体。在开始任何一个试验前，计算所需要的样本量大小是非常重要的。这和检验效能有一定的关系。

概率

任何一个事件都有许多可能的结局。任何一个结局的可能性都是从 0(不可能发生)到 1(总是发生)。在临床研究中，进行任何一类数据推断的时候，概率(或者叫 P 值)都是必不可少的部分。这让读者明确，结局是由研究变量的效应产生的，而不是单纯由于机遇而产生的。因此，当某个研究显示需要输血在对照组和接受氨甲环酸的试验组之间存在差异，也就是 P 值等于 0.01 时，这就意味着单纯由于机遇所产生而不是由于使用了氨甲环酸所产生的差异的概率是很小的。在医学统计学中，描述事件发生概率的另两个常用术语是比值比和危险比。

比值比

机率是指发生事件与未发生事件的比值。比如在 160 位产妇当中，有 110 位自然分娩而有 50 位行剖宫产术，因此剖宫产的机率是 50: 110，也就是 0.45。在临床研究中，一个更为重要和常用的概念就是比值比。它是用于测量某种干预因素对发生事件概率的影响。在自然分娩产妇中，以硬膜外镇痛与对照组产妇的比值比为例。在硬膜外组中，有 90 例自然分娩产妇，60 例剖宫产产妇。则剖宫

产的机率为 60: 90，也就是 0.66。因此，硬膜外镇痛与对照组的剖宫产比值比是 0.66: 0.45，也就是 1.46。当比值比为 1 时意味着两组间没有差异，而在上述这种情况下，比值比则意味着使用硬膜外镇痛时剖宫产的机率会增加。

危险比

危险比是流行病学研究中的常用术语，并且与比值比相似。在计算危险比时最重要的是要知道这里的分母是总人数。例如以上的数据中，剖宫产的风险是 50: 160，也就是 0.312，而机率是 0.45。在大规模的流行病学研究中，危险比、绝对危险度减少率和相对危险度减少率都是用于确定某种干预因素的有效性以及推广某种治疗方案可能带来的经济影响的重要统计学指标(图 9)。危险比在病例对照研究中是无法计算出来的，但是可以通过比值比来近似估计危险比。

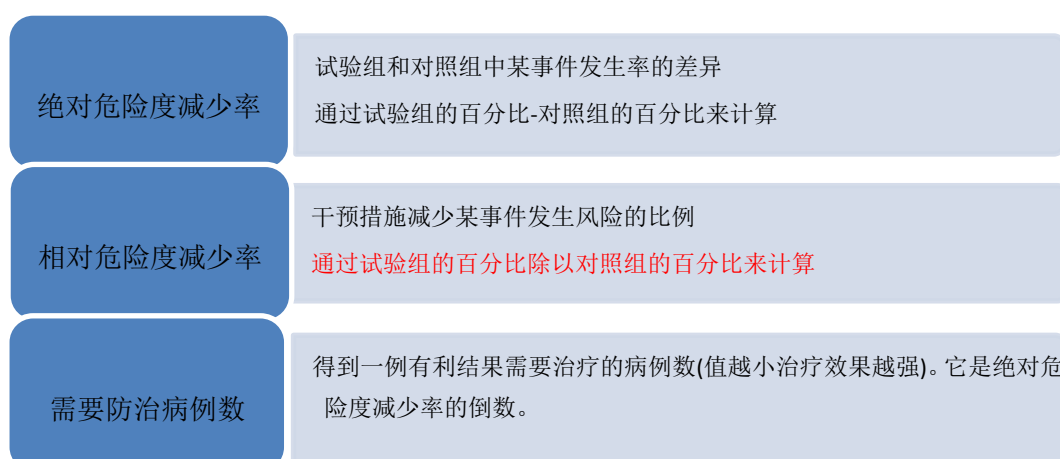


图 9 绝对危险度减少率、相对危险度减少率以及需要防治病例数之间的区别

以上这些结果的重要性需要在研究的正文中予以评价。在某个研究中，如果只计算相对危险度的减少率，那么一个大样本的研究就有可能提供误导信息。

例如：在防治术后恶心呕吐的风险中，对比新的止吐药与金标准。

- A 组(对照组): 1000 个患者中 10 个患者出现了呕吐。
- B 组(新药物): 1000 个患者中 5 个患者出现了呕吐。

绝对危险度减少率为 $99.5\% - 99\% = 0.5\%$ ，而相对危险度减少率为 $0.005:0.01 = 0.5$ 或 50%

有些作者会使用相对危险度减少率来突出术后恶心呕吐风险降低了 50%。但是，如果你计算需要防治的病例数(NNT)=100:0.5=200，你会发现与金标准相比，需要治疗 200 例患者才能减少 1 例术后恶心呕吐的患者。如果你知道这种新药价格昂贵并且有明显的副作用例如心律失常，那么这些结果可能会改变你使用这种新药的想法。

无效假设和备择假设

统计分析就是通过试验观察值来定量的分析证据的强度，从而接受或者拒绝假设。在医学统计学中，通常是假设不同组间的某个变量是没有差异的，除非有证据证明他们之间存在差异。这称之为无效假设。备择假设则是相反的，并且研究者通常对于这个问题更感兴趣。假设差异来自于某种治疗或干预措施。那么统计检验的主要目的则是拒绝无效假设同时证实备择假设。

误差的类型

不适当的拒绝了无效假设的时候可能会发生 I 型错误，也称之为 α 误差。它意味着在没有真实差异存在的时候，却发现了显著的统计学差异。显著性水平 P 值代表了拒绝无效假设 α 的最小值，在医学统计学中，P 值通常设为 0.05(1/20 的发生率)。

在研究报告中，应当计算并提及 P 值，但并不是简单的说大于或者小于 0.05。例如，同一干预措施的两个不同的研究提到的 P 值分别是 0.051 和 0.049，那么事实上，他们的结果非常相似。但是如果只报告 P 值 < 0.05 或者 P 值 > 0.05 的话，就可能产生误解。

不适当的接受了备择假设的时候可能会发生 II 型错误，也称之为 β 误差。它意味着当真正存在差异的时候却没有发现显著的统计学差异。这种误差最常见的原因是不恰当的样本量大小。因此在开始任何一个试验前，计算检验出差异所需要的样本量大小是一项非常重要的工作，从而可以避免 II 型错误的出现。这称之为检验效能。

检验效能是指当差异存在时，能够检验出显著统计学差异的概率。他的计算方式为 $(1-\beta)$ 。大多数医学研究接受的 β 值为 0.2，同时没有效能计算的研究是不被推荐的。

统计分析的类型

对于不同类型数据选择合适的统计分析方法是非常重要的。以下是如何对你的数据建立合适的统计分析方法的系统性步骤：

1. 分辨出数据是定性的还是定量的
2. 如果是定量资料，确定数据的分布类型
3. 确定需要比较的组数
4. 确定数据是否需要配对

统计分析的具体方法在本文中不作讨论，但流程如图 10 所示。

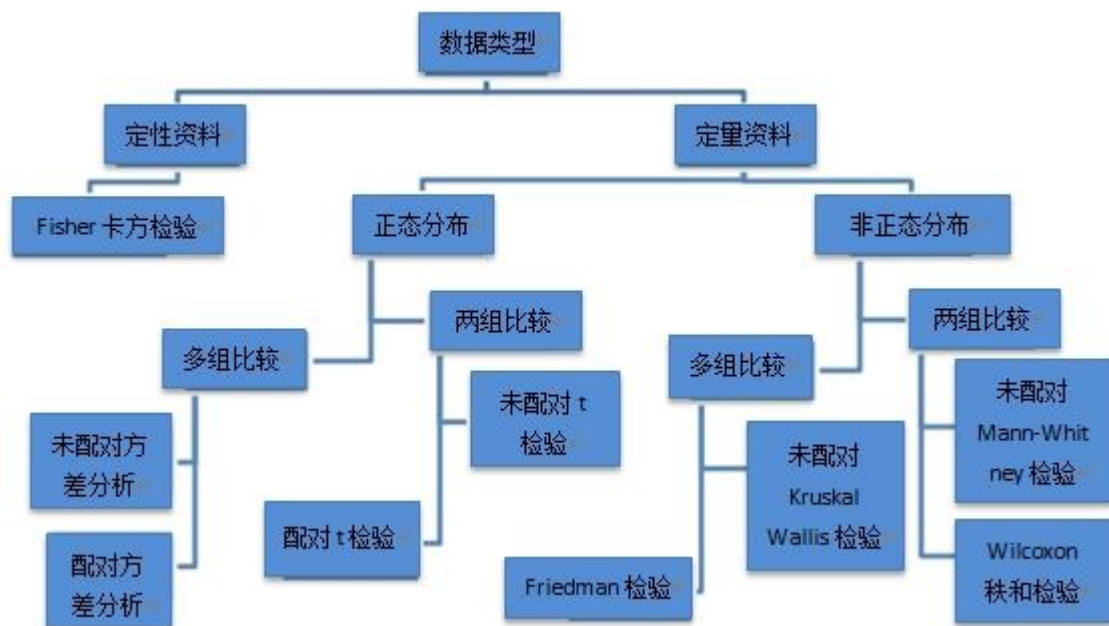


图 10 选择恰当统计分析的方法

样本和总体

为了了解代表总体人群的样本的精确性，可能会使用到其他的一些统计学概

念。包括均数的标准误(SEM)以及可信区间(CI)。

均数的标准误不是描述集中趋势或离散趋势的指标，它反映的是你的样本均数与总体均数的接近度。它可以通过以下公式来计算($SEM=SD/\sqrt{n}$)。从这个公式中我们可以看到，样本量越大，均数的标准误就越小，样本的代表性就越好。这仅仅是因为样本量越大，其中包含关于总体的信息就越多，因此结果的估计值就越精确。可信区间是未知总体参数的估计范围，这个估计范围是通过已知的样本数据所计算出的。可信区间告诉我们样本效应的大小代表总体的可能性。影响可信区间的因素包括样本量大小和标准差。样本量越大，样本代表总体的代表性就越好，可信区间就越窄。重要的是，如果样本的可信区间很宽，那么它可能不是代表的一个真实的总体。在比较新的治疗方案与金标准的随机对照研究中，要特别注意可信区间。例如，一个研究报告指出在感染性休克患者中，使用血管加压素比去甲肾上腺素平均增加平均动脉压 20mmHg，它的可信区间是-5 到 45mmHg，那么这个研究的结论就应该是在该人群中，治疗效应无显著性差异。这是由于基于可信区间的该人群的平均动脉压预测改变值可能是 0 或者 -5mmHg。图 11 所示的研究为锻炼对静息心率影响的可信区间。

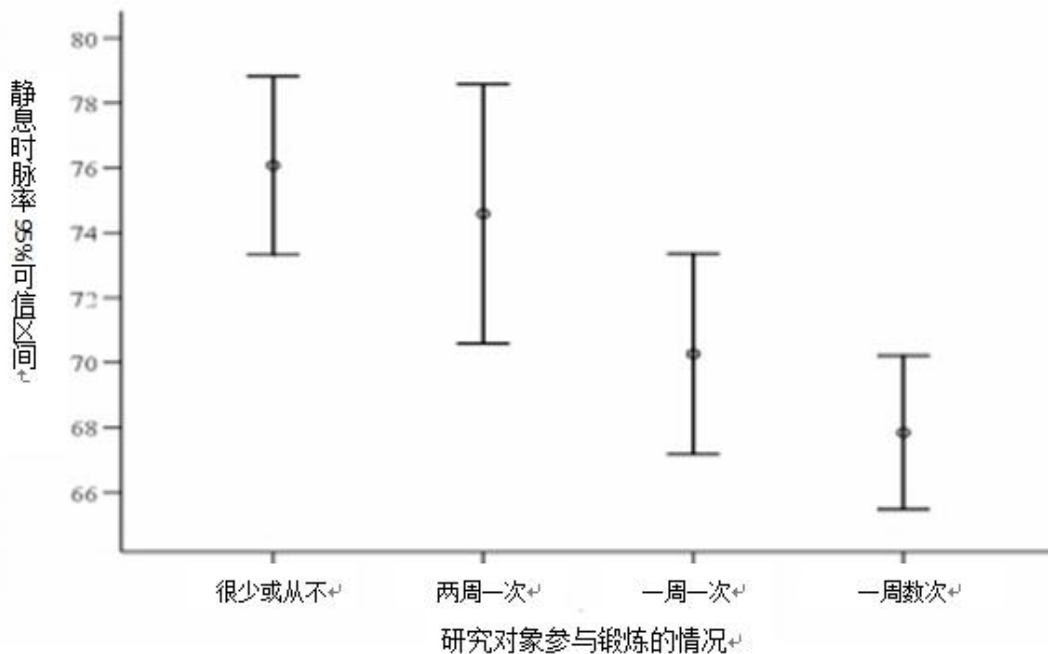


图 11 例举不同研究中，锻炼对静息心率影响的可信区间

问题的答案

1. TTTTF 2.FFTTF

网站链接

<http://www.cyclismo.org/tutorial/R/index.html>

<http://www.stats.gla.ac.uk/steps/glossary/index.html>

<http://en.wikipedia.org/wiki/Statistics>

参考文献和推荐阅读

1. McCluskey A, Lalkhen A. Statistics A publication series from Continuous Education in Anaesthesia, Critical Care and Pain Management, Oxford University Press in 2007
2. Harris M, Taylor G. Medical Statistics Made Easy, 2nd Edition, Scion Publishing Ltd, 2008.
3. Bowers D, Medical Statistics from Scratch an Introduction for Health Professionals, 2nd Edition, John Wiley & Sons Ltd, UK 2008.
4. Spoons C, Kiff K. Oxford Specialty Training: Training in Anaesthesia, 1st Edition, Oxford University Press in 2010

四川大学华西医院

翻译审校 朱涛