序

医学研究，犹如执灯探路于迷雾之中，而样本含量恰似灯光的强度——不足，则前路模糊，结论难立；过度，则眩目耗神，徒费资源。如何把握“恰到好处”的尺度，照亮科学探索之路，是每一位医学研究者必须直面之问，亦是统计学者始终钻研之艺。

近日，钟晓妮教授携其新作《医学研究样本含量估计》邀我作序。欣然提笔之际，不由想起我们因参与《卫生统计学》（第 6 版）编撰而结缘的往事。当时钟教授作为参编团队的重要成员，她治学严谨、思路清晰、为人谦和，令我印象深刻。如今，钟教授及谢彪副研究员等围绕“样本含量估计”这一专题深耕成书，融入软件操作与鲜活案例，化繁为简，足见其团队学术功底之深厚与传播科学之热忱。

当今医学研究设计日益复杂，数据形态愈发多元，从随机对照试验到真实世界研究，从诊断试验到高维组学分析，样本含量的估计方法亦随之演进、不断丰富。不少研究者，尤其是青年学者，常在此处举步维艰：公式繁多、假设抽象、软件生疏，仿佛面前横着一堵无形之墙。

《医学研究样本含量估计》正是为“推墙”而来。作者团队立足实践，以研究设计为体，软件操作为翼，将 PASS、R 和 nQuery 等软件融入具体场景，追求“即学即用、即用即会”。全书贯通理论、案例与实践，参数阐释清晰，语言简明易懂，既延续了钟教授一贯的教学智慧，亦体现了其“为用而著”的务实风格。

这是一本“授人以渔”的优秀工具书，可以帮助研究者在设计之初便树立“样本意识”，从而提升医学研究的质量与效率；既传递严谨的科研方法，亦示范系统的科学思维与求实学风。

寥寥数语，难尽全书之妙。唯愿读者朋友能明其“算法”，通其“道理”，使研究设计胸有成竹，软件操作得心应手，在科学之路上行稳致远。

方积乾

于广州，2025 年冬

前　言

样本含量估计是医学研究设计中的重要环节，也是医学研究者开展科学研究最关注的问题之一。然而，许多研究者在实际研究过程中，面临着样本含量估计的挑战。一方面，样本含量估计方法基于复杂的统计学公式和假设，非统计学专业的研究者难以理解和应用；另一方面，随着医学研究设计日益复杂，样本含量估计方法也在不断演进，这对研究者提出了更高的要求。因此，本书从应用视角，介绍样本含量估计的方法和常用软件的操作，为研究者提供借鉴。

本书共分七章，第一章为绪论，介绍了样本含量估计的基本概念、操作步骤和常用软件；第二章和第三章基于研究设计类型，分别介绍了实验研究的样本含量估计和观察研究的样本含量估计；第四章诊断试验的样本含量估计、第五章临床试验的样本含量估计、第六章贝叶斯样本含量估计、第七章其他情形的样本含量估计，均基于医学研究的应用场景进行了详细介绍。同时，第七章引入了高维组学数据的样本含量估计、所需参数不足时的样本含量估计、经验法的样本含量估计等内容。

本书基于研究设计，融合了样本含量估计方法与 PASS、nQuery 软件演示和 R 代码实例操作，减少了烦琐的计算，便于使用者快速掌握并应用，是医学科研人员、临床医生、医学及相关专业研究生，以及从事医学统计、流行病学方法研究的科研人员的参考书籍。本书具有以下特点：

1. 全面性　本书内容不仅包括实验研究，还包括观察研究的样本含量估计；不仅涉及基础研究，还涉及临床研究应用场景，如诊断试验的样本含量估计、临床试验的样本含量估计；不仅涵盖经典的样本含量估计内容，还及时跟进热点研究方法涉及的样本含量估计有关内容，如贝叶斯样本含量估计、高维组学数据的样本含量估计等；不仅有相关理论的铺垫，还有案例应用的操作实践。

2. 实用性　本书从实用角度出发，针对医学专业特点设置相关内容，形式上不将软件操作单独作为一章，而将相关理论、案例、PASS、nQuery 与 R 软件操作融为一体，紧随案例分析之后提供软件操作，避免理论知识与应用的割裂，提高查阅本书的便捷性和可操作性。

3. 可读性　本书以通俗易懂的语言，将基本理论简介、样本含量估计公式、医学案例、公式计算、PASS 软件操作、nQuery 软件操作、R 软件操作等一体化呈现，在阐述样本含量计算中所涉及的基本理论基础上，对涉及的关键参数予以简明扼要的解释，提高了可读性。

本书的出版融入了众多人员的心血，谨此向每一位编者致以敬意。同时，感谢方积乾教授、易东教授、彭斌教授给予的指导和帮助，感谢汪倩、蒋佳明、曾海娇、刘馨璟、刘恬、刘颖橙、李文龙、谢雨鑫、尚愿、覃芹、陈红、程晨、阎巧坪、秦莉、黄雅婷等研究生在内容校对、案例计算结果的复核等方面付出的辛勤劳动。

在大家的努力下，我们完成了书稿。现在我们鼓起勇气将本书奉献给大家。由于编者水平有限，书中难免存在疏漏与不足之处，恳请各位读者批评指正，以便再版时进一步完善。

钟晓妮　谢　彪

2025 年 12 月，重庆

目　录

第一章 绪　　论

样本含量估计（sample size estimation）是医学科学研究设计的关键环节，需综合考虑检验效能、研究的可行性、结论的时效性、医学伦理及非随机误差等多重因素，研究对象的数量并非越多越好，而是需要在满足医学科研统计学要求、确保一定推断精度的前提下，选择最为合适的样本含量，在科学可靠性、伦理合规性和资源可行性之间取得平衡。样本含量的估计通常涉及多个因素，包括研究设计类型、预期效果、统计功效、显著性水平和允许误差等。本章重点介绍样本含量估计所涉及的基本概念、操作步骤，以及实践中常用的软件工具。

第一节　样本含量估计的基本概念和操作步骤

在医学研究领域，无论是观察性研究，还是实验性探索，抽样研究方法均扮演着核心角色。其主要目的是通过分析实际观测到的样本信息，推断出关于未知总体的特征性结论，该过程被称为统计推断。在设计抽样研究方案时，一个关键因素需要被考虑，即合理确定样本中应包含的研究对象数量（包括人、动物、生物学材料等），以确保研究既满足严格的统计学要求、支持有效的统计推断，又兼顾实际可行性和伦理道德等多维度因素。这种平衡的目的是控制研究成本，降低潜在风险，并提高研究的效率和效果。

一、基 本 概 念

样本含量估计涉及一系列基本概念，融合了统计学原理、实验设计的核心要素及实际操作等因素。

1. 总体（population）**和样本**（sample）　总体是根据研究目的确定的、所有同质研究对象某一（组）指标值的集合；样本是从总体中随机抽取的、数量足够的、能代表总体特征的部分研究对象某一（组）指标值的集合。总体是我们想知道的，样本是我们能知道的，因此医学科学研究常采用抽样研究，从总体中随机抽取一部分作为样本，通过样本特征推断总体特征。若样本含量不足，根据样本进行推断所下的结论则不可靠。

2. 抽样误差（sampling error）　是由于抽取样本的随机性所造成的样本值与总体值之间的差异，亦称为代表性误差。抽样误差的存在是不可避免的，一般而言，随着样本含量的增加，抽样误差会相应减小。

3. 效应量（effect size）　是反映研究中关注变量间差异或关联程度的关键指标，其大

小直接影响样本含量的需求。按效应量的统计意义将其分为三类：差异类（difference-type）、相关类（correlation-type）、组重叠（group-overlap）。

4. 统计功效（power）　是指在原假设为假的情况下，接受备择假设的概率。通俗而言，$P<0.05$ 时，结果显著（接受备择假设）；在此结论下有多大的把握发现这种差别，此时需要用到统计功效来表示这种“把握”。通常，研究者期望的统计功效至少要达到 80%。

5. 显著性水平（significance level）　是在原假设为真时，错误拒绝原假设的概率，即Ⅰ类错误（type Ⅰ error）的最大允许概率。

6. Ⅱ类错误（type Ⅱ error）　是指在真实效应存在的情况下，研究未能正确检测到该效应，其概率为 β。统计功效与Ⅱ类错误概率之和为 1，即统计功效等于 $1-\beta$。样本含量与统计功效密切相关，即增加样本含量可以有效提升研究的统计功效，从而降低Ⅱ类错误的风险。

7. 正态分布（normal distribution）　又称为常态分布，是统计学中最重要的分布之一。其描述了许多自然现象和社会现象的数据分布特征，如身高、体重、考试成绩等。正态分布的特点是数据集中在均数附近，呈现钟形曲线。众多统计检验都基于数据遵循正态分布的假设，在进行样本含量估计时，需要评估数据是否满足这一假设条件。

8. 方差（variance）**和标准差**（standard deviation）　方差是各个变量值与其均数离差平方的平均数，反映了样本中各个观测值到其均数的平均离散程度。标准差是方差的平方根。样本的方差和标准差是衡量样本变异性的重要指标。在样本含量估计过程中，需要充分考虑样本的变异性。

二、操 作 步 骤

在考虑资源限制及符合伦理要求的前提下，样本含量估计的过程应遵循以下一系列严谨、有序的步骤。

第一步　明确研究设计。

（1）明确研究目标：确立清晰的研究问题和相应的假设。

（2）界定研究类型：确定研究的性质，如描述性、相关性或因果性研究。

第二步　确定参数。

（1）选择显著性水平（α）：一般设为 0.05，表示可接受的假阳性概率为 5%。

（2）确定统计功效（$1-\beta$）：期望的统计功效通常应达到 80%或以上，以减少假阴性的风险。

（3）定义效应量大小：界定希望检测到的最小差异或关联的程度，该标准可基于既有研究、理论预测或临床实践意义。故实际操作中，多以反映样本效应值的统计量替代，如用样本均数替代总体均数。

（4）评估变异性：对变量的变异性进行估计或测量，多以反映样本变异性的统计量替代，如用样本标准差替代总体标准差。

第三步　选择公式或软件进行计算。通过应用特定的样本含量计算公式或借助统计软件，估计所需样本含量。

第四步　考虑脱落与失访。考虑可能的样本损失情况（如参与者退出研究），并据此调整样本含量。

此外，还需检查不同参数（如效应量大小、变异性）的变动对样本含量需求的影响。详细记录样本含量估计的依据和过程，以便于在研究设计和方法部分进行报告。综合以上步骤和考虑因素，最终确定研究所需的样本含量。

第二节　样本含量估计的常用软件

样本含量估计的常用软件较多，它们能够帮助研究者精确地计算出所需样本含量，以确保研究的准确性和可靠性。样本含量估计的常用软件包括简单的在线计算工具和专业的统计分析软件，以满足研究者的各类需求。

一、简单的在线计算工具

1. Power And Sample Size

（1）访问途径：http://www.powerandsamplesize.com/。

（2）主要功能：该平台提供多达 30 种不同数据类型下的样本含量计算服务，涵盖单样本均数、两样本均数比较及多个样本均数比较等多样化需求。

（3）显著特点：除了基本的样本含量计算功能外，该工具还为用户提供了详尽的样本含量计算公式及对应的 R 代码示例，适用于学术论文及标书撰写的实际应用场景。

2. 常笑医学网在线样本含量计算工具

（1）访问途径：http://www.cxmed.cn/statistool.html。

（2）主要功能：该工具全面覆盖了单组设计、两组平行设计、2×2 交叉设计及多组平行设计等多种研究设计类型的样本含量计算需求，适用于比较率的差异、是否优于/非劣于某一界值等多种研究场景。

（3）显著特点：用户可根据自身研究设计的具体类型，直接选择对应的在线计算工具，从而快速、准确地获取样本含量计算结果。

3. MedSci 样本含量计算软件（MedSci sample size tool，MSST）

（1）访问途径：MSST 通常是通过梅斯医学（MedSci）的官方网站或 App 进行访问，并没有直接的外部网址链接到样本含量计算软件本身。用户可以通过下载梅斯医学 App，在 App 内部找到并使用 MSST。

（2）主要功能：MSST 涵盖了十多种常见的样本含量计算方法，包括随机对照试验、诊断性研究、病例对照研究等，能够满足临床上 90%以上的样本含量计算需求。

（3）显著特点：MSST 支持各种研究设计和结局指标的样本含量计算，帮助用户在不同研究场景下估计样本含量。同时，MSST 提供全中文的用户界面，方便国内用户使用和理解。此外，该软件还提供了详细的引导和指示，帮助用户快速上手并准确使用各项功能。

4. Epitools

（1）访问途径：Epitools 并未设置特定的网址链接，而是作为由澳大利亚生物安全合作研究中心资助的 Ausvet 动物健康服务机构所创建的动物流行病学在线计算平台。其并非独立的网页或应用，用户通常需借助搜索引擎或相关网站指引，才可找到并访问 Epitools 的相关页面或功能。

（2）主要功能：Epitools 集成了多元化的样本含量计算方法，包括但不限于：单一比例的估算、单一均数的估算、两比例的比较、两均数（无论样本量大小或方差是否相等）的比较、真实患病率（动物或群体层面）的估算、队列研究的样本含量估计及病例对照研究的样本含量估计。此外，该平台还涵盖了流行病学数据分析、贝叶斯估计等功能。

（3）显著特点：主要体现在其全面性和用户友好性。该平台提供了多种样本含量计算方法和数据分析工具，以满足流行病学研究和动物健康监测的多样化需求。其界面设计简洁直观，用户只需按照提示输入相关参数，即可迅速获得准确的计算结果。值得一提的是，Epitools 的大部分功能均免费提供，极大地方便了科研和教学工作。

5. Zstats 风暴统计平台

（1）访问途径：该平台是浙江中医药大学基于 R 语言推出的在线统计分析平台，具体网址：https://zstats.medsta.cn/ samplesize/。

（2）主要功能：可计算单样本均数、两样本均数比较、k 个样本均数比较；支持单个率、两个率比较，配对率比较，两样本率比较，k 个样本率比较，时间-事件数据（生存数据）比较，OR 值比较等功能。

（3）显著特点：适用于医学研究、临床试验、公共卫生研究等多个领域，能够满足不同研究设计的样本含量计算需求。计算报告会直接附带参考文献，方便用户进行文献回溯和引用。界面简洁明了，操作简便，用户友好度高，即使是零基础的用户也能轻松使用。

二、专业统计分析软件

1. PASS（power analysis and sample size）　是国际公认的统计效能分析与样本含量计算金标准工具，专注于功效分析和样本含量估计。该软件具备超过 1100 种统计检验和置信区间场景的样本含量计算工具，其功能的丰富性远超同类软件。历经 25 余年的不断优化与完善，PASS 已成为临床试验、制药及广泛医学研究领域的首选样本含量估计软件。该软件操作简便，提供 30 天的免费试用版本，用户可通过下载试用版来全面体验其功能。

2. SAS（statistical analysis system）　作为一款备受认可的统计分析软件，其在样本含量估计领域展现了卓越的性能。该软件内置了如 POWER 和 GLMPOWER 等多个过程步，这些过程步能够支持不同统计检验下的样本含量估计需求。SAS 软件通过提供详尽的实例和教程，旨在帮助用户深入理解和掌握样本含量估计的方法。尽管 SAS 在统计分析领域的功能强大且全面，但用户普遍反映其学习曲线较为陡峭，可能需要一定的时间和努力来掌握。

3. R 语言　作为一款功能卓越的统计编程语言，提供了广泛的包和函数支持，极大地助力研究者进行样本含量的精确估计。R 语言具备极高的灵活性，研究者可以根据其特定

的研究设计和统计需求，灵活地定制样本含量计算方法。特别是，R 语言内包含多个专门针对样本含量计算的包，其中，“pwr 包”涵盖了多种统计检验的样本含量估计功能，如 t 检验、χ^2 检验（chi-square test）；而“pmsampsize 包”则侧重于临床数据回归分析领域的样本含量估计。这些“包”通常设计有用户友好的接口，使得即便是非统计学背景的研究者也能轻松进行样本含量的估计。

值得一提的是，R 语言及其多数包均采取开源和免费的策略，为研究者提供了无成本使用这些强大工具进行样本含量估计的便利。此外，R 语言拥有一个活跃的开发者与用户社区，该地区提供了大量的教程、指南和示例代码，极大地帮助了新用户快速掌握 R 语言的使用。

R 语言的样本含量计算功能广泛适用于多种研究设计，包括但不限于横断面研究、病例对照研究、队列研究和随机对照试验等。不仅如此，R 语言还能够实现数据清洗、探索性数据分析、统计建模和结果可视化等一系列数据处理流程，为研究者提供从数据准备到分析报告生成的一站式解决方案。

4. G*Power 作为一款在医学研究中广泛应用的统计软件，其核心功能在于精确计算统计检验的样本含量和效应量，并高效执行功效分析。G*Power 全面支持多种统计检验，包括但不限于 t 检验、方差分析、相关性分析及回归分析等。在效应量计算方面，G*Power 提供了包括 Cohen’s d、Pearson’s r 和 f 效应量在内的多种计算方法。用户可以根据具体的研究设计，灵活输入包括效应量、显著性水平（α）及统计功效（$1-\beta$）在内的各项参数。

G*Power 具备广泛的适用性，不仅适用于实验设计，同时也支持观察性研究等多种研究场景。其操作界面设计直观简洁，即便是初学者也能迅速掌握并应用。作为一款免费且开源的软件，G*Power 为研究者提供了零成本的样本含量和功效分析解决方案。

此外，G*Power 还具备良好的跨平台性，可在 Windows、macOS 和 Linux 等多种操作系统上稳定运行。在学术界，G*Power 凭借其卓越的性能和稳定性，获得了广泛的认可，众多研究论文和教材均推荐使用其进行样本含量估计和功效分析。

5. Stata 作为一款功能强大的统计分析软件，具备卓越的样本含量估计能力。其内置了多样化的工具和命令，如 power 命令家族（包括 power onesample，power twosample 等），旨在协助研究者进行精确的样本含量估计。研究者可以根据具体的研究设计和统计检验需求，灵活设置效应量、显著性水平、统计功效等关键参数。Stata 的直观界面允许用户直接在软件环境中处理和分析数据，从而便捷地将样本含量估计与数据预分析过程相结合。

除样本含量估计外，Stata 还集成了统计功效分析工具，以辅助用户评估在特定样本含量下检测效应的能力。此外，Stata 社区积极贡献了大量由用户编写的第三方命令和程序，这些工具极大地扩展了软件的样本含量估计功能。

在学术界，Stata 凭借其卓越的性能获得了广泛的应用，诸多文献和教材均提供了详尽的 Stata 样本含量估计操作指南。该软件支持多平台运行，包括 Windows、macOS、Linux 等操作系统，确保了广泛的适用性。Stata 通常提供试用版本，使用户能够在购买前充分评估软件的功能和性能。

6. nQuery 作为一款专注于临床研究设计与样本含量估计的专业统计软件，其核心功能在于为研究者提供样本含量计算、统计功效分析及试验设计支持。nQuery 广泛应用于药

物临床试验、生物统计学研究及医学研究方案设计中，能够满足固定设计和适应性设计等多类研究场景，还提供了贝叶斯样本含量估计模块。研究者可依据具体研究目的与设计类型，输入效应量、显著性水平、统计功效、组间分配比例、事件发生率、失访率等参数，完成相应的样本含量估计。

nQuery 支持多种类型终点和分析方法的样本含量估计，包括连续型结局、二分类结局、生存结局，以及均值比较、率的比较、生存分析等常见统计情形。在临床试验设计方面，nQuery 提供了针对平行设计、交叉设计、成组序贯设计及部分自适应设计情境下的设计支持，可辅助研究者在方案制订阶段进行更为系统的统计规划。

nQuery 长期被制药企业、合同研究组织（CRO）及医学研究机构用于试验设计与样本含量论证，在临床研究方法学领域具有较高的认可度。其功能模块覆盖较为丰富，并提供图形化操作界面，便于用户根据研究设计逐步完成参数输入和结果解释。此外，nQuery 属于商业软件，分为 Base、Plus、Pro、Expert 四个版本，首次注册可申请试用版本，但须购买授权后方可长期使用。

三、软件计算的样本含量与手动公式结果存在细微差异的原因

本书软件计算的样本含量与手动公式结果存在细微差异，在此进行统一说明。这种差异是正常的，主要原因：①算法维度不同，PASS 采用迭代模拟技术处理效应量分布和协变量交互，而手动计算依赖简化假设；②连续性校正方法不同，PASS 默认启用边缘概率校正，而公式常忽略 Yates 校正；③分布假设差异，如二项数据中 PASS 使用精确 Clopper-Pearson 区间而非正态近似；④多重检验控制策略不同，PASS 通过闭合检验程序优化，比手动 Bonferroni 校正更高效。此外，PASS 还能处理缺失数据补偿和复杂设计，如非劣效性检验中的非线性特征模拟，导致差异幅度可达 15%～20%，尤其在基线率低或小样本时更为显著。当用户面临选择专业软件还是采用公式手动计算样本含量时，建议优先使用 PASS 处理复杂设计，如非劣效性检验、生存分析、小样本/极端比例数据或需多重校正的研究，因其采用迭代模拟和精确算法更可靠；而手动公式适合教学演示或资源有限时的快速估算。无论选择哪种方法，都应在报告中明确标注工具版本及关键参数。若研究涉及法规申报等高风险决策，必须使用专业软件以确保准确性。

（石　磊　钟晓妮）

第二章　实验研究的样本含量估计

实验研究是指研究者根据研究目的主动地对实验对象设置干预措施，按照对照、重复、随机化的基本原则控制非干预措施的影响，通过对实验结果的分析，评价干预措施的效果。与观察研究相比，实验研究能较好地控制非干预因素的影响，减少偏倚。样本含量估计是实验研究设计的重要内容之一，其目的是确定研究对象中具有代表性的样本数量，对于研究结果的准确性和可靠性具有重要影响。根据实验设计的方法，可分为单组设计、配对设计、平行组设计、交叉设计、重复测量设计和其他设计。本章节将根据此分类分别进行样本含量估计的介绍。

第一节　单组设计的样本含量估计

单组设计又称单组目标值法，没有为试验组设立传统意义上的同期对照组，而是采用目标值作为理论对照值，将试验结果与理论值进行比较。单组设计根据评价指标的不同可分为定性（二分类）指标的单组设计和定量指标的单组设计。

一、定量指标的样本含量估计

在研究设计中，双侧检验的样本含量估计见公式（2-1-1）。

$$n=\frac{\left(Z_{\alpha/2}+Z_{\beta}\right)^2\sigma^2}{\left(\mu_1-\mu_0\right)^2} \qquad (2\text{-}1\text{-}1)$$

若为单侧检验，将上式 $Z_{\alpha/2}$ 替换为 Z_{α}。

式中：

（1）n 为样本含量；

（2）α 为检验水准；

（3）β 为Ⅱ类错误概率；

（4）μ_1 为总体均数；

（5）μ_0 为目标值；

（6）σ 为总体标准差。

例 2-1-1　本研究拟检测降压药物氨氯地平是否能显著降低患者的收缩压。根据既往研究，已知在目标人群中，收缩压下降值的总体标准差约为 10mmHg。研究者期望氨氯地平能够使平均收缩压比用药前至少降低 5mmHg，设目标值为 0，要求 α 为 0.05，把握度（$1-\beta$）

为 80%，需要多少样本含量来验证该药物的有效性？

根据公式（2-1-1）计算样本含量（本例为单侧检验）。

$$n=\frac{\left(Z_{\alpha}+Z_{\beta}\right)^{2}\sigma^{2}}{\left(\mu_{1}-\mu_{0}\right)^{2}}=\frac{\left(1.645+0.842\right)^{2}\times10^{2}}{5^{2}}\approx25$$

在样本量计算中，Z_{α} 为Ⅰ类错误小于等于 α 时的 Z 值，反映了原假设正确时错误拒绝原假设这一错误情况对应的标准正态分布下的临界值；Z_{β} 是Ⅱ类错误 β 所对应的 Z 值，与统计功效相关，体现了原假设错误时接受原假设这一错误情况在标准正态分布下的对应值。

据此，该研究需要 25 例研究对象来进行药物有效性的研究。

PASS 软件操作：

打开 PASS 软件，操作过程如下：

→单击左侧菜单【Means】

→选择【One Mean】

→选择【T-Test（Inequality）】

→单击右侧界面【Tests for One Mean（Simulation）】

→弹出【Tests for One Mean（Simulation）】对话框进入单组设计定量指标的样本含量估计界面，如图 2-1-1。

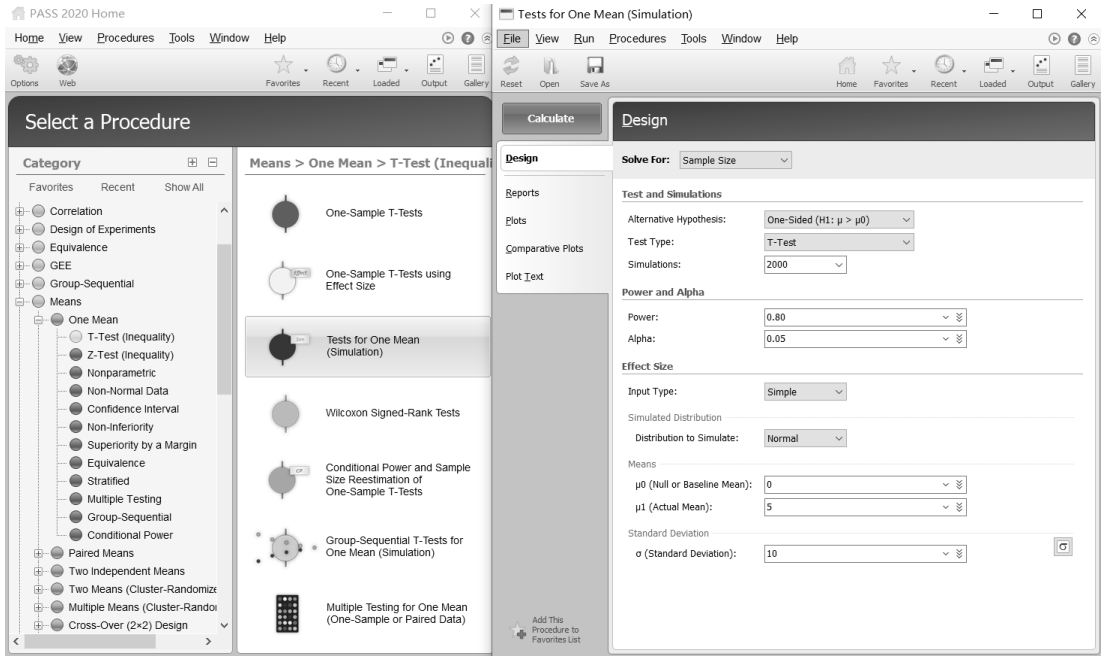


图 2-1-1　PASS 软件单组设计定量指标的样本含量估计操作示意图

弹出的对话框中：

（1）Solve For 选择 Sample Size，说明本次所求的结果为样本含量；

（2）Power（$1-\beta$）一般选择 0.80 及以上，本例 Power 为 0.80；

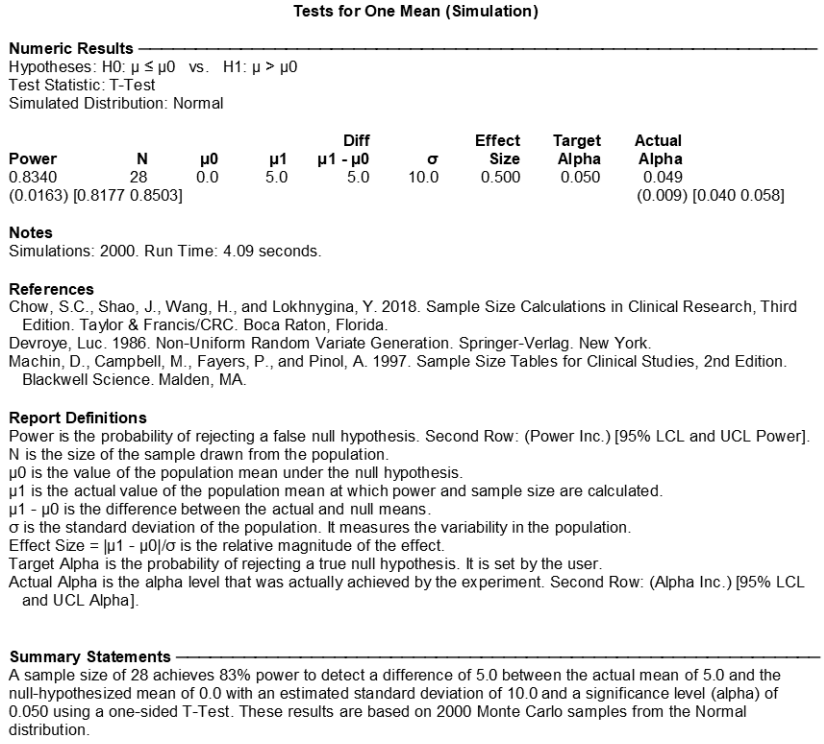
（3）Alpha 一般选择 0.05，本例 α 为 0.05；

（4）$\mu 0$（Null or Baseline Mean）：指定公认的标准值或总体均数，本例总体均数为 0；

（5）$\mu 1$（Actual Mean）：指定预期的标准值或总体均数，本例中总体均数为 5；

（6）σ（Standard Deviation）：设定标准差，本例标准差估计值为 10；

（7）单击【Calculate】按钮，完成样本含量估计，详见图 2-1-2。



Tests for One Mean (Simulation)

Numeric Results

Hypotheses: H0: μ ≤ μ0 vs. H1: μ > μ0
Test Statistic: T-Test
Simulated Distribution: Normal

Power	N	μ0	μ1	Diff μ1 - μ0	σ	Effect Size	Target Alpha	Actual Alpha
0.8340	28	0.0	5.0	5.0	10.0	0.500	0.050	0.049
(0.0163) [0.8177 0.8503]								(0.009) [0.040 0.058]

Notes

Simulations: 2000. Run Time: 4.09 seconds.

References

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.

Report Definitions

Power is the probability of rejecting a false null hypothesis. Second Row: (Power Inc.) [95% LCL and UCL Power].
N is the size of the sample drawn from the population.
μ0 is the value of the population mean under the null hypothesis.
μ1 is the actual value of the population mean at which power and sample size are calculated.
μ1 - μ0 is the difference between the actual and null means.
σ is the standard deviation of the population. It measures the variability in the population.
Effect Size = |μ1 - μ0|/σ is the relative magnitude of the effect.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment. Second Row: (Alpha Inc.) [95% LCL and UCL Alpha].

Summary Statements

A sample size of 28 achieves 83% power to detect a difference of 5.0 between the actual mean of 5.0 and the null-hypothesized mean of 0.0 with an estimated standard deviation of 10.0 and a significance level (alpha) of 0.050 using a one-sided T-Test. These results are based on 2000 Monte Carlo samples from the Normal distribution.

图 2-1-2　PASS 软件单组设计定量指标的样本含量估计结果示意图

PASS 软件给出的主要结果：样本含量估计的结果、相关参考文献、样本含量估计报告中各名词的定义、对计算结果的总结描述等。主要关注【N】结果即可。该研究最少需要 28 例研究对象，与使用公式（2-1-1）计算的结果（n=25）基本一致。

R 软件操作：

使用 R 软件对例 2-1-1 中的样本含量进行估计。打开 RStudio，在代码输入框中输入以下代码（图 2-1-3）。

```
#安装和加载"pwr"包
install.packages("pwr")
library(pwr)
#计算样本含量
pwr.t.test(n = NULL,
           sig.level = 0.05,
           type = "one.sample",
           alternative = "greater",
           power = 0.80,
           d = 5/10)
```

图 2-1-3　R 软件单组设计定量指标的样本含量估计操作示意图

图 2-1-3 中：

（1）代码“sig.level = 0.05”指定检验水准，本例 α 为 0.05。

（2）代码“power = 0.80”指定检验效能，本例 β 为 0.2。

（3）代码“type = "one.sample"”表示单样本。

（4）代码“alternative = "greater"”表示单侧检验。

（5）代码“$d=\frac{\mu 1-\mu 0}{\sigma}=5/10$”表示效应值。

（6）最后估计样本含量（n），并输出结果，见图 2-1-4。

```
One-sample t test power calculation

              n = 26.13753
              d = 0.5
      sig.level = 0.05
          power = 0.80
    alternative = greater
```

图 2-1-4　R 软件单组设计定量指标的样本含量估计结果示意图

由图 2-1-4 可知，本研究需要 27（26.13753 取整）例研究对象来进行药物有效性的研究，与公式 2-1-1 计算得到的样本含量估计结果（n=25）基本一致。

二、定性指标的样本含量估计

研究设计中，双侧检验的样本含量估计见公式（2-1-2）。

$$n=\frac{\left[Z_{\alpha/2}\sqrt{\pi_0(1-\pi_0)}+Z_{\beta}\sqrt{\pi_1(1-\pi_1)}\right]^2}{\delta^2} \qquad (2\text{-}1\text{-}2)$$

若为单侧检验，将上式 $Z_{\alpha/2}$ 替换为 Z_{α}。

式中：

（1）n 为样本含量；

（2）α 为检验水准；

（3）β 为Ⅱ类错误概率；

（4）π_0 为已知总体率；

（5）π_1 为预期实验结果的总体率；

（6）$\delta=\pi_1-\pi_0$。

例 2-1-2　某临床研究中心拟研究缬沙坦对原发性高血压患者的疗效，根据既往临床数据，同类型的血管紧张素Ⅱ受体阻滞剂药物血压控制有效率约为 80%。研究者假设缬沙坦可能具有更高的有效率，并期望其有效率达到 90%。设定 α 为 0.05，β 为 0.2，需要多少例研究对象才能发现缬沙坦的有效率有 10%的提高？

根据公式（2-1-2）计算样本含量（本例为单侧检验）。