

健康教育 常用定量调查方法

中国疾病预防控制中心健康教育所
培训与评价中心
钱玲
qianlingzh@126.com

定量调查(quantitative survey)

- 根据事先设计的调查问卷/表对一定数量的研究对象通过询问、测量等方式获得量化资料的方法。
- 数据资料的统计分析---计算调查指标---说明调查目的
- 调查结果具有一定的可靠性和精确性
- 根据不同调查范围，分为**普查**和**抽样调查**。
- 在健康教育现场调查中，由于普查本身的局限性以及不同的研究目的，往往以**抽样调查**应用较多。

2008-9-11
2

普查 (overall survey)

- 全面调查 (complete survey)
- 在特定时间对特定范围人群 (即确定的所有调查总体) 全部加以调查或测量。
 - 人口普查 (census)
 - 在有限的规定范围内进行的调查
- 普查不存在抽样误差，理论上只有普查才能准确取得目标人群的状况。
- 一时性的---一般用于了解调查总体在某一特定“时点”的情况；
- 能够全面描述行为及有关因素的分布特征；
- 能起到普及医学科学知识的作用。
- 工作量大，不易做得细致，常难免漏查；
- 耗费人力物力大，成本较高。

2008-9-11
3

普查应注意

- 明确划定普查的范围；
- 统一调查时点或时期；
- 尽可能于短期内完成；
- 统一项目的规定和标准；
- 统一组织领导，培训调查人员，统一普查方法、步骤、进度等；
- 尽量减小漏查：一般要求普查应答率在**85%**以上。

2008-9-11
4

抽样调查 (sampling survey)

- 是一种非全面调查
- 通过在特定的调查总体中随机抽取一定数量的样本对象进行调查或测量
- 由样本来获得对总体的推断

2008-9-11
5

抽样调查

- 调查对象较普查少，节省时间、人力、物力；
- 调查范围小，调查工作容易做得细致。
- 设计、实施和资料分析比较复杂；
- 重复和遗漏不容易发现；
- 不适用于变异较大的资料调查；
- 如果某种行为发生率较低时，样本量大小不能提供足够的信息资料，样本量扩大到近于总体的75%时，直接进行普查更有意义。

2008-9-11
6

抽样调查应注意

- 选择**合适的抽样方法**;
- 估计**样本调查对象的大小**;
- 抽样误差的适当控制:
 - 不可避免,可以测量;
 - 通过样本大小和抽样设计来控制。
- 必须防止非抽样误差的影响:
 - 非抽样误差产生偏倚,是一种错误。
 - 主要由于抽样方法的设计不随机,设计者主观选择调查对象;
 - 在抽样过程中未按照预先设计进行随机抽样,任意变换抽样方法。
- 统一调查的时限;统一并固定调查标准和方法;
- 尽量提高调查的应答率(一般要求在85%以上)。

2008-9-11
7

抽样方法 *sampling*

- **概率抽样**
 - 单纯随机抽样
 - 系统抽样/机械抽样/等距抽样
 - 分层抽样/分群抽样
 - 整群抽样
 - 多阶段抽样
- **非概率抽样**
 - 方便抽样
 - 目的抽样
 - 滚雪球抽样
 - 定额抽样
 - ...

2008-9-11
8

非概率抽样/非随机抽样

- 方便抽样:根据方便程度选择容易接近的对象构成样本。
 - 适用于预调查/预试验
- 目的抽样:根据研究目的需要和研究者主观判断选择样本。
 - 取决于研究者对调查对象代表总体的程度的确认
- 滚雪球抽样:首先确定符合研究目的对象进行调查,根据他们的推荐扩大调查,不断滚动,至达到样本量为止。
- 定额抽样:类似分层抽样,只是每层的样本量由研究者根据研究目的确定。

2008-9-11
9

随机抽样方法 举例

- 若采取随机抽样方法,从全市10000名大学生中抽查1000名进行调查,了解该市大学生艾滋病防治相关知识、态度和行为的情况。

2008-9-11
10

单纯随机抽样

- 从随机数字表中取出10000个随机数字记在学生卡片上;
- 按随机数字大小将卡片排序;
- 以开头的1000张或末尾的1000张卡片上的学生为样本;
- 或者每10张卡片抽取一张组成样本量为1000的样本。

2008-9-11
11

单纯随机抽样

- 以学生的学号作为编号或采取其他编号方法;
- 从随机数字表中的任一位置开始(如从第4行第3列开始),按一定的顺序(如向右的方向)抄录1000个四位数字(如随机数字表中数字为五位数,只取后面的四位数);
- 凡学生编号属于这1000个随机数字的即作为抽取的调查样本

2008-9-11
12

系统抽样

- 今总体例数 $N=10000$,样本例数 $n=1000$,抽样比 $=1000/10000=1/10$,也就是每隔10名学生要抽取一名学生。
- 现将10000名学生进行编号(1, 2, 3, ..., 10000)或使用学生学号按从小到大顺序排列;
- 按照顺序将学生分成1000个部分,第1部分编号为1, 2, 3, ..., 10;第2部分编号为11, 12, 13, ..., 20;以此类推;
- 从第1部分中随机抽取某号学生,如4号学生作为样本;则按照“10”的抽样间隔陆续抽取第二部分中14号,第三部分中的34号,.....
- 于是,以编号为4, 14, 24, 34, 44, 54, 64, 74, 84, ..., 9974, 9984, 9994的学生组成样本。

2008-9-11
13

整群抽样

- 将全校所有班级按照学生数多少进行编号(1, 2, 3, ...,) ;
- 采取简单随机抽样方法抽取班级;
 - 如果各班级学生数量均在30人左右,则随机抽取3个班级。
 - 如果各班级学生数量不同,则抽取班级数依据样本量大小而定。
- 对抽取班级的全部学生进行调查。

2008-9-11
14

分层抽样

- 按照性别分层;
- 根据男女生在该学校学生中的比例,确定具体应抽取多少名男生/女生;
- 可按照简单随机抽样、系统抽样或整群抽样的方法抽取男生或女生。
- 按照专业—年级—班级等特征进行分层。

2008-9-11
15

???

- 优缺点
- 区别

2008-9-11
16

调查研究中的样本量估计

□ 有公式支持的经验估算方法

- > 以率作为调查指标,基础样本量为100例
- > 以均数作为调查指标,基础样本量为200例
- > 根据影响调查指标的可能因素,进行扩展。

- 以艾滋病防治知识知晓率为限制指标,考虑性别(2层)、年级(4层)、文理专业(2层)等分层因素,并保证最低一层样本量为100,确定参与定量调查的大学生样本量为 $2 \times 4 \times 2 \times 100 = 1600$ 人,考虑到10%的失访率,实际调查样本量为1800人。

$$\hat{n} = \frac{n}{90\%}$$

2008-9-11
17

调查研究中的样本量估计

$$n = \left(\frac{u_{\alpha}^2 \pi (1 - \pi)}{\delta^2} \right) \quad n = \left(\frac{u_{\alpha} \sigma}{\delta} \right)^2$$

$$\delta = p - \pi \quad \delta = \bar{X} - \mu$$

$$\alpha = 0.05, u_{0.05} = 1.96 \approx 2;$$

$$\delta = 0.1\pi, n = \left(\frac{u_{\alpha}^2 \pi (1 - \pi)}{\delta^2} \right) = 400 \times \frac{1 - \pi}{\pi}$$

$$\delta = 0.15\pi, n = \left(\frac{u_{\alpha}^2 \pi (1 - \pi)}{\delta^2} \right) = 178 \times \frac{1 - \pi}{\pi}$$

$$\delta = 0.2\pi, n = \left(\frac{u_{\alpha}^2 \pi (1 - \pi)}{\delta^2} \right) = 100 \times \frac{1 - \pi}{\pi}$$

2008-9-11
18

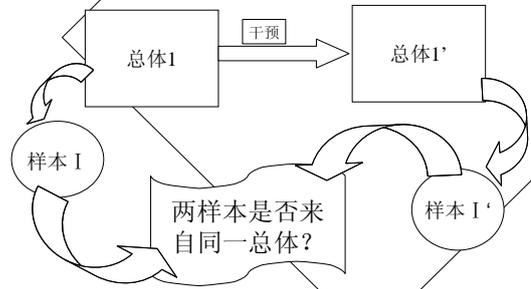
定量研究中的样本量估计

需考虑的因素:

- (1) 容许误差 (δ)
- (2) 调查人群中某种情况的预期率 (π)
- (3) 所调查总体的变异 (标准差 σ)
- (4) 调查要求的精密度 (α)
- (5) 把握度 ($1-\beta$)

2008-9-11
19

干预研究: 前后自身对照



2008-9-11
20

干预前后自身对照研究

$$n = \left[\frac{(Z_{\alpha} + Z_{\beta})S}{\delta} \right]^2$$

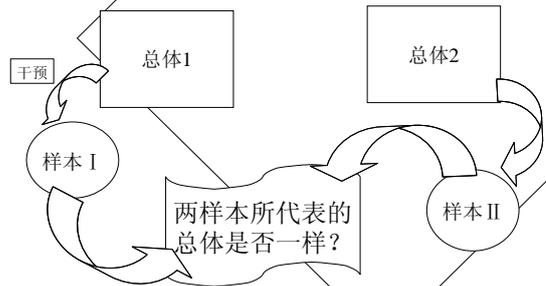
$$\delta = \bar{X} - \mu \quad \text{或}$$

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2 [\pi(1-\pi)]}{\delta^2}$$

$$\delta = p - \pi$$

2008-9-11
21

干预研究: 另设对照



2008-9-11
22

另设对照的干预效果研究

$$n = 2 \left[\frac{(Z_{\alpha} + Z_{\beta})S}{\delta} \right]^2 \quad \text{或}$$

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2 2p(1-p)}{(p_1 - p_2)^2}$$

n为每组样本
所需例数

2008-9-11
23

另设对照的干预效果研究

- 假设对照人群的艾滋病相关危险行为发生率为20%，干预人群在进行健康教育干预后，发生率降低至15%， α 水平为5%， $(1-\beta)$ 为90%，本研究为双侧检验，请估计干预人群和对照人群各需的样本数。

解: 双侧 $u_{0.05} = 1.96$, 单侧 $u_{0.10} = 1.282$ 。

$$n = \frac{(u_{\alpha} + u_{\beta})^2 2p(1-p)}{(p_1 - p_2)^2} = \frac{(1.96 + 1.282)^2 2(0.475)(1-0.475)}{(0.2-0.15)^2} \approx 1214(\text{人})$$

- 即各组需要1214人。

2008-9-11
24

多项指标时的样本量估计技巧

限制性指标——“瓶颈指标”

规则：

- 同时有平均数、率和比时，率和比为限制性指标。
- 同一类指标，误差要求最小的为限制性指标。
- 平均数类指标，标准差最大的为限制性指标。
- 率和比类指标，率和比的值最小的为限制性指标。

2008-9-11
25

调查问卷 questionnaire

□ 一般构成

- 指导语（目的意义，知情同意，填写说明）
- 问卷编码
- 调查项目（包括问题题干及答案）

□ 附件：

- 问卷说明（保证正确理解，统一标准）
- 问卷编码说明
- 数据库建立及数据库录入说明

2008-9-11
26

调查项目

□ 分类：

- 分析项目
- 备查项目

□ 项目设计原则：

- 尽量精简
 - ✓ 原则上：与调查目的和调查指标无关的、不必要的项目坚决不要
- 定义明确
 - ✓ 尤其是疾病分型、吸烟者或饮酒者划分、文化程度等，不能模棱两可
- 陈述简单明了
 - ✓ 通俗易懂，使人不致误解，尽量做到不加说明或少加说明也能标准统一。
- 能用定量指标时尽量采用定量指标。

2008-9-11
27

答案设计

□ 分类：

- 封闭式选择题
- 开放式回答

□ 封闭式选择题的组织和排列

- 定性变量：直接列出答案。如，性别：男、女
- 等级变量：
 - ✓ 升序排列，如：文化程度：文盲、小学、初中、高中、大学及以上、不详。
 - ✓ 降序排列，如：满意程度：很好、好、一般、不好、很不好。
 - ✓ 图表方式列出，如百分尺度线。
- 定量变量：
 - ✓ 直接填入数字；
 - ✓ 转为等级变量，如：年人均收入：1000元以下、1000元—5000元—10000元—。
- 整个问卷尽量选择相同的升序/降序排列方式。
- 有效合理利用纸面。

2008-9-11
28

封闭式选择答案

- 答案标准化，易于整理归类和分析；
- 容易询问，容易回答，节省时间；
- 拒答率较低；
- 文化程度较低者也能填写，节约调查时间，加快调查进度。
- 有时选择答案太少，不能概括所有实际情况；选择答案太多，会增加调查时确切判定的困难；
- 可能诱导调查对象随便选答而失真。
- 不容易觉察到调查对象对问题的误解；
- 填写时容易圈错答案；
- 无法获得出选择答案以外的信息。

2008-9-11
29

封闭式选择答案

- 在答案的类型或数量较少、且能互相区别时使用；
- 一般测量定性变量，或等级变量；
- 答案应尽可能包括主要的、全部的答案，即使在没有指导的情况下，调查对象也能独自回答，不至于产生误解；
- 通常在多种可能的答案外，增加“不详”和“其他”两项答案，以涵盖所有出现的可能情况。

2008-9-11
30

开放式回答

- 在不知道答案的情况下，可使回答者自由发挥，收集到设计者不能预料到的信息资料；
- 让回答者以喜欢的方式回答问题并加以论证说明，充分反映出被调查者的观点；
- 可用于有多种答案的问题。
- 容易离题；
- 不易于进行统计处理、分析；
- 拒答率较高。尤其对于不习惯用文字表达看法者在信访调查时更是如此；
- 调查时间花费较多；
- 适用范围有限，对文化者可能适用。

2008-9-11
31

开放式回答

- 适用于那些少数无法概括的错综复杂的答案的问题，此时设计者无法肯定问题的各种可能答案；
- 要求详细讨论或回答问题，以发现调查对象的特殊意见与观点。

2008-9-11
32

调查问卷的组织

- 项目的排列顺序符合逻辑，循序渐进；
- 问答有条不紊，防止遗漏；
- 敏感性问题 and 开放式问题宜安排在问卷的后面；
- 设计时首先应考虑有关调查对象的基本信息。必要时，可首先考虑列入有关备择项目；
- 检验资料可靠性的问题应分隔开来询问；
- 问卷的编码；
- 调查问卷的填写说明、编码说明。

2008-9-11
33

调查问卷的确定

- 调查问卷的预调查——修改——最终定稿
- 调查指南及建立数据库、录入说明等
- 调查问卷的信度、效度检验

2008-9-11
34

调查问卷的信度 *reliability*

- 问卷的可靠性：测量结果的可靠或可重复程度。
- 可用一致性分析。
 - 随机抽取10%~15%调查对象，进行重复调查；
 - 比较重复调查结果与初次调查结果。
 - 在尽可能短的时间间隔内完成两次调查，并保持调查标准的一致性和调查对象经历的相对稳定。

2008-9-11
35

调查问卷的信度检验

- 分类变量资料：**Kappa**值
 - **Kappa**值越大，表示两次调查的一致性越好，问卷的可信度越高。
 - 一般**Kappa**值在0.75以上为可信度高，在0.40和0.75之间为可信度较好，低于0.40表示可信度差。
- 数值变量资料：**Pearson**积差相关系数 / 简单相关系数 (**r**)
 - **r**越大，表示两次调查的一致性越好。
- 同质性分析

2008-9-11
36

调查问卷的效度 *validity*

- 问卷的正确性：问卷调查结果符合实际情况的真实程度。
 - 效度越高，表示调查结果的真实性愈好。
- 效度分类：
 - 内容效度：问题选择的适当性
 - 构想效度：内容设计的系统性和有效性
 - 效标效度/实证效度：测量结果对所研究内容或行为的预测性。

2008-9-11
37

调查问卷的效度检验

- 内容效度：专家评定法、经验法
- 构想效度：
 - 比较复杂，没有单一指标（内容效度、同质性分析）；
 - 因子分析法
- 效标效度/实证效度：
 - 效度系数（相关系数）
 - 数值变量资料：Pearson积差相关系数 / 简单相关系数 (r)
 - 信度系数的平方根是效度系数的最高限度。
 - 不同组间差异性检验

2008-9-11
38

定量调查的设计 *design*

- 确定调查目的
 - 了解目标人群的特征（描述）
 - 提供评价的本底资料
 - 研究有关影响因素（比较）
 - 了解活动的效果及其影响因素评价性研究

2008-9-11
39

确定调查指标

- 调查目的一定具体化为指标
- 指标选择：
 - 少而精的原则；
 - 尽量选用客观性强、灵敏度高、精确性好的定量指标；
 - 少用定性指标。

2008-9-11
40

确定调查内容

- 调查范围和目标人群
- 调查方法
- 调查对象的数量以及选择方法
- 调查资料的收集方法
- 调查问卷的设计
- 调查的组织实施
- 资料的整理分析方法以及预期结果等等

2008-9-11
41

调查内容的充实

1. 确定调查范围和目标人群、调查对象和观察单位
 - 划清调查总体的同质范围
 - 地点及对象
2. 调查方法及调查对象的数量
 - 确定样本量大小和抽样方法
 - 普查（全面调查）；
 - 抽样调查；
 - 典型调查（案例调查）；
3. 搜集资料的方式
 - 调查资料的收集方法
 - 直接观察法；
 - 面对面的口头询问调查；
 - 自填问卷/信访
4. 确定调查项目和调查问卷
 - 调查问卷的设计

2008-9-11
42

5. 制定调查的现场组织和实施计划

- 组织动员、分工联系;
- 调查员培训;
- 预试验 (预调查);
- 调查问卷及有关材料的准备;
- 调查资料的检查制度 (逻辑检查、计算检查);
- 经费预算等等。

2008-9-11
43

6. 资料的整理及分析计划

- 数据的计算机录入和清理
 - 录入前资料的检查和编码, 录入程序的设计;
 - 录入时的质量控制: 同一资料的重复录入, 逻辑检查;
 - 录入后的抽查: 通过简单的统计描述, 发现异常值等。
- 设计分析表和资料的分组
- 分析方法的选择及使用

7. 调查研究中偏倚的控制

2008-9-11
44

谢谢!
Thanks!

qianlingzh@126.com

2008-9-11
45