



统计学原理(Statistic)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2022-03-26

西北农林科技大学

第二章 数据收集、整理和清洗

2.1 数据目标

2.5 数据质量

2.2 数据收集

2.6 抽样设计

2.3 资料整理和数据清洗

2.7 抽样分布和抽样误差

2.4 数据的数据库化

2.8 问卷设计技术

2.1 数据目标

数据收集的步骤

陈述现实难点 (problem)

提出研究问题 (question)



数据搜集的步骤：1定义研究目标

在开始数据收集过程之前，您需要准确确定要实现的目标。

您可以从编写现实难题（problems）陈述开始：您要解决的实际或科学难题是什么，为什么重要？接下来，提出一个或多个研究问题（questions），以精确定义您要查找的内容。根据您的研究问题，您可能需要收集定量或定性数据：

- 如果您的目的是检验假设，精确测量某些东西或获得大规模的统计见解，请收集定量数据。
- 如果您的目的是探索想法，了解经验或获得对特定环境的详细见解，请收集定性数据。
- 如果您有多个目标，则可以使用混合方法来收集两种类型的数据。



(示例) 如何定义研究目标

(示例)：您正在研究员工对大型组织中直接经理的看法。



- 您的首要目标是评估不同部门和办公室地点对经理的看法是否存在显著差异。
- 您的第二个目标是从员工那里收集有意义的反馈意见，以探索有关管理人员如何改进的新想法。



数据搜集的步骤：2选择数据收集方法

根据您要收集的数据，确定最适合您的研究的方法。

序号	方法	何时使用	如何收集资料
1	实验	测试因果关系。	处理变量并衡量其对他人的影响。
2	民意调查	了解一群人的一般特征或意见。	在线，亲自或通过电话将问题列表分发给样本。
3	访谈/焦点小组	深入了解某个主题的看法或观点。	在个别访谈或焦点小组讨论中以口头方式向参与者提问开放性问题。
4	观察	了解自然环境中的事物。	在不影响样本的情况下测量或调查样本。
5	人种学	直接研究社区或组织的文化。	加入并参与社区，并记录您的观察和思考。
6	档案研究	了解当前或历史事件，条件或做法。	从图书馆，图书馆或互联网访问手稿，文档或记录。
7	二次数据收集	分析无法直接访问的总体数据。	查找来自政府机构或研究组织等来源的现有数据集。



数据搜集的步骤：3规划资料收集程序

当您知道正在使用哪种方法时，您需要准确计划如何实现它们。您将遵循什么程序对您感兴趣的变量进行准确的观察或测量？

- 如果您要进行调查或访谈，请决定将采取何种形式的问题；
- 如果您要进行实验，请对实验设计做出决定。
- **实现可操作化。**可操作化意味着将抽象的概念转变为可测量的观察结果。在计划如何收集数据时，需要将要学习的内容的概念定义转换为实际要测量的操作定义。
- **设计采样方式。**您可能需要制定抽样计划以系统地获取数据。这涉及到定义总体，您要得出结论的组以及要从中实际收集数据的组的样本。
- **编写标准化程序。**如果涉及多个研究人员，请编写详细的手册以标准化研究中的数据收集程序。
- **制定数据管理计划。**在开始收集数据之前，您还应该决定如何组织和存储数据。



数据搜集的步骤：4动手收集资料

最后，您可以实现所选方法来测量或观察您感兴趣的变量。

为确保以系统的方式记录高质量数据，以下是一些最佳做法：

- 在获取数据时记录所有相关信息。例如，记下在实验研究期间是否或如何重新校准实验室设备。
- 仔细检查手动数据输入是否有错误。
- 如果收集定量数据，则可以评估可靠性和有效性，以表明您的数据质量。



陈述现实难点 (problem) : 基本内容

在开始数据收集过程之前，您需要明确难点并确定要实现的目标。

陈述难点，就是说明你要解决的实际或科学难点是什么，为什么重要？

明确的**陈述难点**需要具备如下几个要素：

- 将难点放在特定背景之中（我们已经知道什么？）
- 描述研究将要解决的确切难点（我们需要知道什么？）
- 显示难点的相关性（为什么我们需要知道它？）
- 设定研究目标（您将做什么以找出答案？）
- 提出研究难点：精确定义你要聚焦或解决的难点。



陈述现实难点 (problem) : 如何陈述I

步骤1: 将难点具体化。

- 对于实际研究难点，具体关注情况：难点何时何地出现？该难点影响谁？为了解决这个难点已经做了什么尝试？

例子：与该国其他地区相比，过去十年中X区的选民投票率一直在稳定下降。根据组织Y进行的调查，在25岁以下的人群和低收入人群中，投票率最低。已经进行了一些有效的尝试，以使这些团体参与其他地区的活动，在最近的两次选举中，政党A和B增加了在X区域的竞选活动，但是这些干预措施尚未对投票率产生任何重大影响。



陈述现实难点 (problem) : 如何陈述I

- 对于理论研究难点，具体关注和考虑科学，社会，地理或历史背景等方面：
关于该问题的已知信息是什么？难点是否仅限于特定时间段或地理区域？这个难点在学术文献中是如何定义和辩论的？

例子：在过去的十年中，“零工经济”（Zero hour）已成为劳动力市场中越来越重要的部分。30岁以下的年轻人更有可能从事自由、合同或零小时工作安排，而不是传统的全职工作（full time）。关于这种转变的原因和后果的研究集中在收入、工作时间和就业条件的客观衡量上，但是很少有研究探索年轻人对零工经济的主观经验。



陈述现实难点 (problem) : 如何陈述?

步骤2: 说明其重要性。

对于**实际研究难点**, 重要性往往与特定难点直接相关, 这个特定难点如何更广泛地影响组织、机构、社会团体或社会。因此, 可以问: 如果不解决难点将会怎样? 谁会感到后果? 难点是否具有更广泛的相关性 (例如, 在其他情况下是否也发现了类似的难点)?

例子: 低投票率与社会凝聚力和公民参与度之间存在负相关关系, 在许多欧洲民主国家中, 这一点正日益引起人们的关注。当特定的公民群体缺乏政治代表权时, 随着时间的流逝, 他们很可能被更多地排斥在外, 从而导致人们对民主制度的信任度下降。解决该问题将为X区域带来实际好处, 并有助于理解这一普遍现象。



陈述现实难点 (problem) : 如何陈述?

对于**理论研究难点**，有时理论难点会产生明显的实际后果，但有时它们的相关性并不那么明显。要确定难点为何重要，可以问：解决难点将如何增进对议题的理解？它对未来的研究有什么好处？这个难点对社会有直接或间接的影响吗？

例子：在零工经济的文献中，这些新形式的就业有时被称为灵活的积极选择，有时被视为剥削性的不得已而为之。为了更全面地了解年轻人为何从事零工经济，需要进行深入的定性研究。关注工人的经验可以帮助建立更稳健的灵活性和不稳定性的理论，同时也可以为未来的政策目标提供信息。



陈述现实难点 (problem) : 如何陈述3

步骤3: 设定目的和目标。

难点陈述应说明您打算如何解决难点。您的目标不应是找到最终的解决方案，而应找出难点背后的**原因**，并提出解决或理解难点的更有效**方法**。

目的 (aim) 是你研究的总体目的，通常以不定式形式编写：

- 这项研究的目的是确定.....
- 该项目旨在探索.....
- 我打算调查.....

目标 (objectives) 是你将要实现该目的的具体步骤：

- 定性方法将用于识别.....
- 我将使用调查来收集.....
- 使用统计分析，该研究将测量.....



陈述现实难点 (problem) : 如何陈述3

步骤3: 设定目的和目标。

示例1（实际难点研究的目的和目标）：这项研究的目的是调查有效的参与策略，以增加X区域的投票人数。它将通过调查和访谈来确定不参与投票的最重要因素，并进行实验以衡量不同策略的有效性。

示例2（理论难点研究的目的和目标）：该项目旨在更好地了解年轻人在零工经济中的经验。定性方法将用于深入了解各个行业中从事自由职业和零小时工作的30岁以下青少年的动机和看法。这些数据将通过对演出经济的最新文献进行回顾，并对劳动力中的人口变化进行统计分析，从而进行背景分析。



(示例) 如何有效地陈述现实难点

现实难点：与该国其他地区相比，在过去十年中，X区的选民投票率一直在下降。

背景情形（已知的内容）：根据组织Y进行的调查，在25岁以下的人群和低收入人群中，投票率最低【引用具体数字】。有关Z国投票模式的文献表明，这反映了更广泛的趋势，但是该地区的人口统计信息使其成为一个更重要的问题【请使用来源进行扩展和解释】。已经进行了一些成功的尝试来提高其他区域的投票率，但是类似的干预措施尚未在X区域产生任何重大影响【引用来源】。需要就该地区参与的具体障碍以及影响青年和低收入人群的有效战略开展更多研究。



(示例) 如何有效地陈述现实难点

相关性（为什么重要）：低投票率与社会凝聚力和公民参与度之间存在负相关关系，在许多欧洲民主国家，政党和公民社会组织正在日益关注这一领域【提供实例和引用来源】。当特定的公民群体缺乏政治代表权时，随着时间的推移，他们很可能被更多地排斥在外，从而导致对民主机构的信任受到削弱，并在治理上带来困难【扩大并提供资料解释】。解决这一难点将使各政党有洞察力来调整其政策和竞选策略，改善X地区居民的民主包容性，并有助于对选民行为的当前趋势有更细微的了解。



(示例) 如何有效地陈述现实难点

目的和目标（您将要做什么）：这项研究的目的是调查有效的参与策略，以增加X区投票者的投票率。它将通过调查和访谈来确定不参与投票的最重要因素，并进行实验以衡量不同策略对投票意图的影响。



提出研究问题 (question) : 基本要求

一个好的研究问题对于指导您的研究、项目或论文至关重要。它将精确地指出了你要查找的内容，并为你的工作提供了明确的重点和目标。

所有研究问题应为：

- 专注于单个问题
- 可使用主要和/或次要来源进行研究
- 在时限和实际限制条件下可行回答
- 具体到足以彻底回答
- 足够复杂，可以在论文或论文的范围得出答案
- 与你的学习或社会有广泛相关性



提出研究问题 (question) : 基本要求

下面给出两个示例：

(案例1) 学校教育：



- **现实难点：** X学校的老师没有能力识别或正确指导教室里的有天赋的孩子。
- **研究问题：** X学校的老师可以使用哪些实用技术来更好地识别和指导有天赋的孩子？

(案例2) 零工经济：



- **现实难点：** 30岁以下的年轻人越来越多地从事“零工经济”而不是传统的全职工作，但是很少有人研究年轻人在此类工作中的经历。
- **研究问题：** 影响年轻人参与零工经济的决定的主要因素是什么？工人认为它的优缺点是什么？年龄和受教育程度对人们体验这种方式有影响吗？



提出研究问题 (question) : 常见类型

下表显示了一些最常见的研究问题类型。

研究问题类型	公式
描述性研究	X的特征是什么？
比较研究	X和Y之间有什么区别和相似之处？
相关研究	变量X和变量Y之间有什么关系？
探索性研究	X的主要因素是什么？ Y在Z中的作用是什么？
解释性研究	X对Y有影响吗？ Y对Z的影响是什么？ X的原因是什么？
评估研究	X的优缺点是什么？ Y工作得如何？ Z有多有效或理想？
行为研究	如何实现X？ 改善Y的最有效策略是什么？



提出研究问题 (question) : 什么是好的研究问题?

好的研究问题应该: 专注性和可研究

- 专注于单个主题和问题: 您的主要研究问题应从您的研究问题开始, 以使您的工作重点突出。如果您有多个问题, 那么所有这些问题都应与该中心目标明确相关。
- 不要求主观价值判断。避免使用诸如“好/不好/更好/更糟”的主观用语, 因为这些主观用语没有给出回答问题的明确标准。如果您的问题正在评估某些内容, 请使用具有更可衡量的定义的术语。
 - (bad) X或Y是更好的策略吗?
 - (good) X和Y策略在降低Z率方面的效果如何?



提出研究问题 (question) : 什么是好的研究问题?

好的研究问题应该：专注性和可研究

- 可使用主要或次要数据。您必须能够通过收集定量和/或定性数据或通过阅读有关该主题的学术资料来论证来找到答案。如果无法访问此类数据，您将不得不重新考虑您的问题，并提出更具体的问题。
- 不问为什么。为什么问题通常过于开放而不能充当良好的研究问题。通常有很多可能的原因导致研究项目无法给出详尽的答案。尝试询问什么或如何提问。
 - (bad) 为什么会出现X?
 - (good) 造成X的主要因素是什么?
 - (good) X如何受到Y的影响?



提出研究问题 (question) : 什么是好的研究问题?

好的研究问题应该：可行而具体

- **在限定条件下完成：**确保您有足够的时间和资源来进行回答问题所需的研究。如果您认为您可能难以获得足够的数据访问权限，请考虑缩小问题的范围，使其更加具体。
- **使用明确的特定概念：**您在研究问题中使用的所有术语均应具有明确的含义。避免使用模糊语言和广阔的思路，清楚你的问题指向是什么、谁、哪里和何时出现？
 - (bad) 社交媒体对人们的思想有什么影响？
 - (good) 每天使用Twitter对16岁以下青少年的注意力范围有什么影响？
- **不要求最终的解决方案/政策或行动方案：**研究是在告知而非指导。即使您的项目专注于实际问题，它也应该旨在增进理解并提出可能性，而不是寻求现成的解决方案。
 - (bad) 政府应如何处理投票率偏低的问题？



提出研究问题 (question) : 什么是好的研究问题?

好的研究问题应该：复杂而有争议

- 无法回答“是”或“否”：封闭的是/否问题太简单，无法像优秀的研究问题一样工作——它们没有提供足够的调查和讨论范围。
 - (×) 在过去的十年中，英国的无家可归人数有所增加吗？
 - (✓) 在过去十年中，经济和政治因素如何影响英国的无家可归者模式？
- 无法用容易找到的事实和数字来回答：如果您可以通过Google搜索或阅读一本书或一篇文章来回答问题，那么它可能还不够复杂。一个好的研究问题需要原始数据，多种来源的综合，解释和/或论据才能提供答案。
- 提供辩论和审议的范围：问题的答案不仅应是简单的事实说明，还需要有空间供您讨论和解释所发现的内容。在论文或研究论文中，这一点尤其重要，因为对于您的问题的答案通常采取有争议的论文陈述的形式。



提出研究问题 (question) : 什么是好的研究问题 ?

好的研究问题应该：与现实相关且具有原创性！

- **解决与您所在领域或学科相关的问题：**研究问题应基于围绕您的主题的初步阅读而提出，并且应专注于解决现有知识中的问题或差距。
- **有助于进行话题性的社会或学术辩论：**这个问题的目的应该是促进现有的辩论——理想情况下是您所在领域或整个社会当前的辩论。它应该产生将来的研究人员或从业人员可以依靠的知识。
- **问题尚未得到回答：**您不必问以前从未有人想到的突破性问题，但是这个问题应该具有独创性（例如，通过专注于特定位置或对长期辩论进行新的思考）。



(示例) 提出好的研究问题 (question)

示例1：社交媒体

- (bad) 社交媒体对人们的思想有什么影响？
- (good) 每天使用Twitter对16岁以下青少年的注意力范围有什么影响？



第一个问题还不够具体：什么类型的社交媒体？哪个人有什么样的效果？
第二个问题更清楚地定义了其概念。它是可研究通过定性和定量数据收集。



(示例) 提出好的研究问题 (question)

示例2：荷兰住房危机



- (bad) 为什么荷兰出现住房危机？
- (good) 大学国际化政策对荷兰住房的可用性和可负担性产生了哪些影响？

以“为什么”开头通常意味着您的问题不够集中：可能的答案太多，没有明确的研究起点。通过仅针对问题的一个方面并使用更具体的术语，第二个问题为找到答案提供了一条清晰的途径。



(示例) 提出好的研究问题 (question)

示例3：医疗保健系统



- (bad) 美国或英国有更好的医疗保健系统吗？
- (good) 美国和英国如何比较慢性病低收入人群的健康结局和患者满意度？

第一个问题过于笼统和过于主观：什么才是“更好”没有明确的标准。第二个问题要研究得多。它使用明确定义的术语，并将其关注范围缩小到特定人群。



(示例) 提出好的研究问题 (question)

示例4：地区投票率



- (bad) 政党应该如何应对X区的低投票率?
- (good) 在X区30岁以下的年轻人中，提高选民投票率的最有效的交流策略是什么?

对于学术研究来说，回答关于“应该做什么”的广泛问题通常是不可行的。
第二个问题更具体，目的是了解可能的解决方案，以便提出明智的建议。

本节结束

