

ICS 03.100.01

CCS A02

团 体 标 准

T/CSSC 001—2021

市场调查 术语

Market Research-Vocabulary

(征求意见稿)

2021 - XX - XX 发布

2021 - XX - XX 实施

中国商业统计学会 发布

目 次

前 言	II
市场调查 术语	1
1 范围	1
2 规范性引用文件	1
3 一般概念	1
4 调查启动	7
5 调查设计	8
6 数据收集	19
7 数据处理	22
8 调查报告	34

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件参考了ISO 20252-2019《市场调查 术语和服务要求》中市场调查 术语部分，并根据我国实际情况调整了部分术语。

本文件由中国商业统计学会提出并归口。

本文件主要起草单位：

本文件主要起草人：

市场调查 术语

1 范围

本标准规定了市场调查中常用的术语及定义。
本标准适用于从事市场调查的组织 and 人员。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

ISO 20252-2019 市场调查 术语和服务要求中的市场调查 术语部分

GB/T 26315-2010 市场 民意与社会调查术语

3 一般概念

3.1

调查 survey/research

基于特定目的，按照一定规范收集、整理调查信息，进行分析和呈现的过程。

注：research 在实际应用中也称为研究或调研，相对survey而言research还包含了分析的过程。

3.2

客户 client

要求调查服务提供者提供研究服务的委托人，包括公司、政府、非政府组织和个人。

3.3

资助者 sponsor

为调查服务提供者开展特定研究提供全部或部分研究经费或物资的机构或个人。

3.4

调查服务提供者 research service provider

按照客户或资助者要求，实施市场、民意和社会调查研究项目的机构或个人。

3.5

第三方 third party

独立的组织或个人，不受资助者和调查服务提供者的责任约束。

3.6

定性研究 qualitative research

通过非结构式研究技术，以非数值形式为主对目标群体的动机、思维方式、态度、预期或行为进行性质上的描述和解释。又称质性研究、质化研究。

3.7

定量研究 quantitative research

通过结构式研究技术，用数值形式对所观察的现象进行数量化描述和解释。又称量化研究。

3.8

自主研究 independent research

调查服务提供者出于自身需要，自行组织的研究。

3.9

委托研究 commissioned research

调查服务提供者根据客户要求，由客户提供经费而开展的研究。

3.10

资助研究 funded research

调查服务提供者根据自身需要或资助者要求，由资助者提供部分或全部支持而开展的研究。

3.11

现场服务 field service

为客户提供实地收集调查研究信息的服务。

3.12

分包 subcontracting**外包 outsourcing**

调查服务提供者将项目的部分工作转交其他机构或个人的行为。

3.13

研究人员 researcher

负责调查研究项目设计或分析的人员。

3.14

小组成员 panel member

已提供个人资料和适当的信息以验证身份，并明确同意根据调查服务提供者规定的条款和条件参与调查，并收集其数据信息。

3.15

现场工作人员 field worker

实地组织与收集数据的人员。通常包括(但不限于)督导员(6.1)、访问员(6.2)、预约员(6.4)、神秘顾客、观察员和审计人员。

3.16

受访者 respondent

接受访问调查的人员。

3.17

信息安全 information security

维护数据信息的机密性、完整性和可用性。

3.18

保密 confidentiality

调查服务提供者未经授权不得向其他机构或个人提供项目数据、信息和成果的强制性要求。

3.19

评估 appraisal

应用一定的方法科学测评调查过程的实施效果。

3.20

个人资料 personal data

与自然人有关的信息，可用于识别个人。

注：可以通过参考直接标识信息（例如姓名、电话号码、照片、声音、视频记录等）进行标识，也可以通过参考个人的身体、生理、心理、经济、文化或文化背景进行间接识别。

3.21

质量管理 quality management

调查活动中，指挥和控制组织在质量方面的协调管理活动。通常包括(但不限于)制定质量方针和质量目标以及质量策划、质量控制、质量保证和质量改进。

3.22

质量控制 quality control

为确保调查质量而采取的质量控制标准、流程及评估方法。

3.23

市场调查 market research

在市场活动中，以特定研究目的为指引，有计划地搜集、整理和分析市场的信息资料，提出解决问题的方案及建议的一整套科学方法、技术和规程。

3.24

专项调查 *Ad hoc research*

专门为解决某类问题或实现特定客户的某一目标而设计的调查。

3.25

多用户调查 *syndicated service research*

针对某类主题、采用统一方法为众多客户提供标准化产品而进行的调查。

3.26

探索性调查 *exploratory research*

对知之甚少的事物进行初步研究的调查方法与技术。探索性调查的常用方法有：专家咨询或调查、预调查（5.6.15）、个案研究（5.4.10）、二手资料（5.4.11）分析。

3.27

描述性调查 *descriptive research*

采用结构性方法描述总体（5.5.1）的特征或功能。描述性调查的常用方法有：二手资料（5.4.11）分析、观察法（5.4.3）。

3.28

因果性调查 *causal research*

为了确定研究对象中变量之间因果关系而进行的调查，是对变量之间作用机制的深刻认识。因果性调查的常用方法是实验法（5.4.7）。

3.29

预测性调查 *predictive research*

利用已有的市场经验和科学的预测技术，对市场未来的发展趋势进行估计和判断的调查方法。

3.30

文案调查 *desk research*

通过收集已有的资料、数据、调查报告、已发表的文献等有关的二手信息，加以整理和分析的一种市场调查方法。又称二手资料（5.4.11）调查，经常在探索性研究阶段使用。

3.31

实地调查 *field research*

在制定详细的调查方案基础上，由调查员直接向被调查者搜集第一手资料，再进行整理和分析，进而写出调查报告的市场调查方法。又称一手资料调查。

3.32

连续性调查 successive research

针对连续变化的调查总体，抽取调查样本，进行连续多期的调查。调查样本一般可分为固定样本、重复样本和轮换样本等形式。

3.33

搭车调查 Hitchhiking research

多个客户共同利用一个样本进行调查。

3.34

大数据 big data

所涉及的数据资料规模巨大到无法通过目前主流软件工具，在合理时间内达到摄取、管理、处理、并整理成为帮助企业经营决策更积极目的资讯，通常为非结构化数据（7.15.2）和半结构化数据（7.15.3）。又称巨量数据。

3.35

数据元 data element

用一组属性描述其定义、标识、表示和允许值的数据单元。又称数据元素或数据类型。

3.36

元数据 metadata

为了描述数据的数据，主要是描述数据属性的信息，用来支持如知识存储位置、历史数据、资源查找、文件记录等功能。又称中介数据、中继数据。

注：元数据的每一个元素都是一个数据元，用符合数据元标准的元数据属性和描述方法来说明元数据。

3.37

暗数据 dark data

来自大数据的阴暗面，未被发掘或理解的数据，即信息资产企业在日常业务活动中收集、处理和存储但不具备特定用途的数据。

3.38

数据集 data set

可以一种或多种格式访问或下载的可识别的数据集合。

3.39

数据变异 data variation

数据集的传输速率、格式或结构、语义或质量的变化。

3.40

数据科学 data science

从数据中提取有用知识的一系列技能和技术。

3.41

数据库 database

存储在一起的一系列相关数据的集合。

3.42

数据模型 data model

根据信息系统中的形式化描述和应用的数据库管理系统的要求，在数据库中构造数据的模式。

3.43

云计算 cloud computing

通过网络“云”将巨大的数据计算处理程序分解成无数个小程序，再通过多部服务器组成的系统进行处理和分析这些小程序得到结果并返回给用户的过程。

3.44

描述性统计分析 descriptive statistical analysis

对调查总体中所有统计指标的相关数据进行统计性描述的分析方法，主要包括数据的频数分析、集中趋势分析、离散程度分析、分布特征以及一些基本的统计图形。

3.45

推断性统计分析 inferential statistical analysis

通过随机样本的数量特征信息，对总体的数量特征进行以概率形式表述的推断的一种统计分析方法，它建立在描述性统计分析的基础之上。

3.46

市场调查提供者 market research provider

具体承担市场调查任务的机构或个人。

3.47

市场调查使用者 market research user

从市场调查提供者处有权获得相应的信息资料、调查报告和有关建议的机构或个人。

3.48

市场调查道德规范 code of market research morality

协调市场调查所涉及的各行为主体（市场调查提供者、市场调查使用者、受访者）之间关系的准则或章程。

3. 49

市场调查道德伦理 ethics of market research morality

指导市场调查所涉及的各行为主体之间关系的行为规范指南。

4 调查启动

4. 1

项目目标 research goal

对项目拟解决的问题, 以及为解决问题需要收集的信息的总体表述。

4. 2

项目需求 research requirement

项目目标的细则及任务清单。

4. 3

客户需求简述 client briefing

由客户向调查服务提供者提交的调查任务清单。通常包括所需获取的信息类型、调查的范围、对象、时间及其他限制性内容。通常也是调查建议书设计的依据。

4. 4

项目建议书 research proposal/research plan

调查服务提供者提交的调查设计及执行计划。

4. 4. 1

研究假设 research hypothesis

对所关注的现象之间的相互关系做出的推测性判断或设想。

4. 4. 2

概念化 concept

对调查对象和研究内容的属性、特征的抽象表述。

4. 4. 3

匿名化 anonymize

删除、模糊、汇总或更改标识信息, 以防止识别具体的调查对象。

4. 4. 4

操作化 operational definition

确定对抽象的概念进行具体测量、观察的指标。

4.4.5

抽样方法 sampling methods

调查中抽取样本的技术规则，包括概率抽样方法和非概率抽样方法。

4.4.6

数据收集方法 data collection methods

调查中获取数据资料的途径和技术规则，包括数据资料提供者、记录方式和采用的媒介。

4.4.7

项目进度表 project schedule

完成调查项目流程所需的时间进度计划。

4.5

项目协议书 research contract/research agreement

在委托调查中，客户和调查服务提供者书面达成的合作协议。

5 调查设计

5.1

市场调查方案设计 market research program design

根据调查目的、对象性质对调查工作总任务的各方面和各阶段进行通盘考虑和安排，形成相应的实施方案，制定出合理的工作程序。

5.2

市场调查目的 purpose of market research

通过对资料的收集、整理与分析，寻找利用机会的可能性及解决问题的可行方案。

5.3

市场调查问题 market research questions

帮助管理者解决决策问题所需要的具有一定广度与深度的信息。

5.4

调查方法 survey methods

针对不同目的和调查对象特点采取的收集数据的方法。

5.4.1

小组座谈法 group interview

主持人（6.3）组织一组具有代表性的受访者，采用小型座谈会的形式搜集专题信息的调查方法。

5.4.2

投影技法 projection techniques

采用无结构、非直接询问方式搜集潜在动机、态度或情感信息的调查方法。

5.4.3

观察法 observation research

通过观看、跟踪和记录调查对象的言行来汇集信息资料的调查方法。

5.4.4

结构性观察 structured observation

观察员按照明列的项目，记录和观察对象行为的调查方法。

5.4.5

非结构性观察 unstructured observation

观察员对观察对象行为进行非格式化记录的调查方法。

5.4.6

单向镜观察 one-way mirror observation

观察者在单向镜后面记录观察对象行为的调查方法。

5.4.7

实验法 experiment

调查中有控制的改变一个或多个条件(变量),并观测这种变化对另一个变量的影响,从中发现因果关系的调查方法。

5.4.8

问卷调查 questionnaire survey

将需要研究的问题编制成题目,以书面提问方式搜集资料的一种方法。

5.4.9

深度访谈 depth interview

访问员和受访者采取一对一的方式围绕相关问题进行深入的讨论。

5.4.10

个案研究 case-study approach

对单一的人或事进行具体而深入的研究。

5.4.11

二手资料 Secondary data

其他人或机构搜集好的而且通常已经使用过的信息资料。

5.4.12

面访 face-to-face interview

访问员与受访者进行面对面的访谈,以搜集信息资料的调查方法。

5.4.13

拦截调查 mall intercept interview

在公共场所,随机拦截受访者当场或在访谈室中进行面对面的访问。

5.4.14

入户调查 door-to-door interview

访问员按照抽样(5.5.4)方案中的要求,到抽中的家庭或单位,按事先规定的方法选取适当的受访者,再依据问卷(5.6.1)或调查提纲进行面对面直接的访问。

5.4.15

邮寄调查 mail survey

通过邮寄问卷给受访者,由受访者自己回答问题并寄回问卷的调查方式。

5.4.16

电话调查 telephone interview

访问员通过拨打电话的方式对受访者进行调查。

5.4.17

计算机辅助电话调查 computer-assisted- telephone interview

访问员操作计算机提取问卷问题,通过电话询问受访者,并将直接将答案输入电脑控制中心的电话访谈方法。

5.4.18

计算机辅助面访 computer-assisted personal interview

访问员操作计算机提取问卷问题,通过面对面访谈方式询问受访者,并将直接将答案输入电脑的访问方法。

5.4.19

计算机辅助自填式问卷访问 computer-assisted self-administrated- interview CASI

受访者自行操作计算机提取问卷问题,并将直接将答案输入电脑的访问方法。多用于受访者隐私需高度保障的调查中。

5.4.20

计算机辅助在线调查 computer-assisted web interview

受访者在线填答发布在互联网网页上的问卷,其答案记录于网络服务器数据库的调查方法。

5.4.21

互联网调查 internet interview

借助互联网进行的调查,包括on line调查。

5.4.22

神秘顾客调查 mystery shopping

观察员扮成顾客,记录有关服务过程细节的情况,作为服务评价依据的参与性人员观察法。

5.5

抽样设计 sampling design

5.5.1

总体 population

由所关注的全部同质个体组成的集合,在市场调查中又称为目标总体。

5.5.2

个体 individual

组成总体的不可再分的最小单元。

5.5.3

抽样单元 sampling unit

构成总体的个体或个体集合。

5.5.4

抽样 sampling

通过抽样框(5.5.5)从总体中抽取一部分抽样单元构成样本的过程。

5.5.5

抽样框 sampling frame

关于抽样单元的完整目录。

5.5.6

样本 sample

从总体中抽取的抽样单元构成的集合。

5.5.7

样本量 sample size

样本中包含的抽样单元的个数(或个案的个数)。

5.5.8

个案 case

构成样本的个体,是样本中的最小单元。

5.5.9

个案数 number of case

样本中包含的样本单位个数。与样本数据矩阵行数相对应。

5.5.10

自愿样本 voluntary sample

非经抽样选择而自愿接受调查的人或机构的集合。

5.5.11

可访问样本库 access panel

由承诺将配合调查的潜在受访者组成的集合。不包括连续调查的固定样本组。

5.5.12

样本轮换 sample rotation

在连续调查中,每隔一定时期更换一定比例的抽样单元。

5.5.13

概率抽样 probability sampling

按照一定概率从构成总体的所有抽样单元中随机选择部分单元进入样本的抽样方法。

5.5.14

简单随机抽样 simple random sampling

从总体的全部单元中以等概率、不放回的方式抽取部分单元。

5.5.15

分层抽样 stratified sampling

样本来自于总体不同的层,且每个层至少有一个抽样单元入样的抽样。

5.5.16

分层随机抽样 stratified random sampling

每层都进行简单随机抽样的分层抽样。

5.5.17

整群抽样 cluster sampling

随机抽取群, 样本由被抽中群中的所有抽样单元组成的抽样。

5.5.18

系统抽样 systematic sampling

将总体单元按一定顺序排列, 先随机抽取一个或一组单元编号, 然后按某种确定的规则抽取其他单元的编号, 直至抽到所需样本量为止。

5.5.19

二阶段抽样 two-stage sampling

假设总体由若干个初级单元组成, 每个初级单元又由若干个二级(次级)单元组成, 在总体中按照一定的方法抽取部分初级单元, 在每个抽中的初级单元中再抽取若干个二级单元进行调查的方法。多于二阶段的抽样, 称为多阶段抽样。

5.5.20

二重抽样 double sampling

先从总体中抽取一个大样本, 通过简单的预调查筛选满足条件的对象, 从中再抽样进行下一步的主调查, 也称为二相抽样。

5.5.21

不等概率抽样 sampling with unequal probabilities

按照不等概率抽取样本的方法。最常见的不等概率抽样是PPS (5.5.22)。

5.5.22

与规模成比例的概率抽样 PPS sampling with probability proportional to size

每个单元被抽中入样的概率与其单元规模的大小成比例的有放回的概率抽样方法。

5.5.23

设计效应 design effect

用于衡量抽样设计方案的好坏, 一个常用的指标是将设计方案的抽样方差与简单随机抽样的抽样方差进行对比, 其值大于1或小于1。

5.5.24

非概率抽样 non-probability sampling

样本不是按照一定的概率抽取, 而是由抽样者主观抽取或任由受访者自愿进入样本的抽样方法。

5.5.25

滚雪球抽样 snowball sampling

根据研究对象的稀有特征, 选定一个或多个受访者, 然后根据受访者提供的信息寻找其他满足条件的受访者的抽样方法。

5.5.26

主观抽样 judgment sampling

根据研究人员的主观意愿、经验和知识, 从总体中选择具有典型代表性的样本单位构成样本, 作为调查对象的一种非概率抽样方法。

5.5.27

便利抽样 convenience sampling

根据调查者的方便与否来抽取样本的一种非概率抽样方法。

5.5.28

配额抽样 quota sampling

将总体按一定标准划分为若干个类别, 在每个类别中按照事先规定的比例或数量抽选受访者的抽样方法。

5.5.29

区域抽样 area sampling

以构成总体的个体所在的地理区域或经济区划为抽样单元的抽样方法。

5.5.30

目录抽样 catalogue sampling

以构成总体的个体目录清单为抽样框的抽样方法。

5.5.31

抽样误差 sampling error

由于用样本数据对总体特征进行估计所引起的代表性误差。

5.5.32

非抽样误差 non-sampling error

除抽样误差以外, 由其他原因引起的样本统计量与总体真值之间的差异。

5.5.33

抽样框误差 Sampling frame error

抽样框中的信息存在缺陷或者错误而产生的误差。

5.5.34

计量误差 metering error

对一个问题所做的回答记录与真值不同而形成的误差。

5.5.35

无回答误差 non-response error

调查时未能从指定的受访者处获得有效的回答而形成的误差。

5.5.36

准确度 accuracy

估计值与相应总体参数之间的接近程度。

5.6 问卷设计

5.6.1

问卷 questionnaire

为了搜集人们对某个特定问题的态度、价值观、观点或信念等信息而设计的一系列问题。

5.6.2

自填式问卷 Self-filled questionnaire

需要受访者自己填写的问卷。

5.6.3

访问式问卷 Interview questionnaire

由受访者在各种方式的访问中所使用的一种问卷。

5.6.4

发放式问卷 Distribution questionnaire

在调查现场由调查员将问卷直接发给选定好的目标群体，待回答完问题后再进行统一收回的问卷。

5.6.5

网络问卷 Online questionnaire

发布在网络上需要受访者填写并在线提交的问卷。

5.6.6

甄别问卷 screening questionnaire

用以筛选合格受访者与否的问卷。

5.6.7

测试问卷 pre-testing questionnaire

为了确保调查项目预期目的的实现,在正式调查之前使用的问卷。

5.6.8

问卷长度 length of questionnaire

完成每一份问卷所用的平均时间。

5.6.9

问题 question

问卷中对受访者关于调查内容的提问。

5.6.10

封闭性问题 closed-ended- questions

答案只能从事先给定的选项中进行选择的问题。

5.6.11

开放性问题 open -ended- questions

事先未给定答案选项,受访者可以自由回答的问题。

5.6.12

敏感性问题 sensitive questions

内容涉及个人隐私或不便于公开表述的问题。

5.6.13

过滤性问题 filter

用以识别合格受访者的问题。

5.6.14

回答 answer; response

受访者对问卷中问题的回应,或者是针对事先给定可能应答的选择。

5.6.15

预调查 pre-investigation

正式调查之前的测试性调查。

5.7

问卷评估 questionnaire evaluation

围绕信度和效度,对问卷尤其是问卷中的量表进行的质量评价。

5.7.1

信度 reliability

问卷调查结果的一致性、稳定性和可靠性程度。

5.7.2

再测信度 retest reliability

用同一份量表对同一组人员进行前后两次测量,然后用两次测量所得结果的相关系数评价量表信度的方法。

5.7.3

复本信度 parallel-forms reliability

对被测量的问题使用两个平行量表分别进行测量,然后用这两个量表上回答结果之间的相关系数评价量表信度的方法。

5.7.4

分半信度 split-half reliability

将用来测量问题的一组量表分成两半,然后根据两半部分的测量结果计算出的相关系数来评价量表信度的方法。

5.7.5

内部一致性信度 internal consistency reliability

用测量同一概念的多个量表之间的一致性 or 同质性来评价量表信度的方法。

5.7.6

效度 validity

问卷能否正确衡量研究者所关注属性的程度。

5.7.7

内容效度 content validity

实际测量到的内容与所要测量的内容之间的吻合程度。

5.7.8

准则效度 criterion validity

采用的测试量表与准则量表之间的一致性程度。

5.7.9

结构效度 construct validity

采用量表测量结果体现出来的某种结论与关于所测属性的理论命题之间的对应程度。

5.8

测量 measurement

按照特定的规则将数字或符号分配给研究对象的特征将其量化的过程。

5.8.1

态度测量 attitude measurement

运用测量工具对受访者的认识态度进行测量。

5.8.2

行为测量 behavioral measure

运用测量工具对受访者行为习惯或行为特征进行客观的测量。

5.8.3

背景测量 background measurement

运用测量工具对受访者的有关背景信息进行测量。

5.8.4

测量尺度 measurement scale

用来测定总体属性的指标或变量的类型,一般分为定类尺度、定序尺度、定距尺度和定比尺度。

5.8.5

量表 scale

通过一套事先拟定的用语、记号和数目,来测定人们状态、观念、意见、态度、行为或心理活动的度量工具。

5.8.6

分等式量表 fractional equation scale

无参照前提下直接确定被测试对象对某事物或现象的态度等级的量表。

5.8.7

排列式量表 Permutation scale

对若干个受测项目进行比较后,再按其偏好程度对这些项目排出先后顺序的量表。

5.8.8

文字式量表 verbal scale

测量工具上的刻度直接以文字加以表述的量表。

5.8.9

图式量表 schema scale

测量工具以图示的形式表现出来的量表。

5.8.10

表式量表 table type scale

测量工具以表格形式表现出来的量表。

5.9

示卡 show card

访问现场经常使用的一种实物卡片,用以形象地呈现较多内容的题目或说明。

5.10

访谈指南 interview guide

在定性访谈中所涉及的话题概要。

6 数据收集

6.1

督导员 supervisor

负责组织、管理数据收集过程的人员。

6.2

访问员 interviewer

按照调查方案与受访者进行联系,并执行数据采集过程的人员。

6.3

主持人 moderator

负责推动参加小组座谈人员之间互动的人员。

6.4

预约员 recruiter

确定并邀请受访者参加研究项目的人员。

6.5

基础培训 basic training

组织执行人员掌握一般执行技能的训练活动。

6.6

项目培训 project training

组织执行人员掌握项目执行文件要求的活动。

6.7

执行文件 executive document

对研究项目执行过程要点的文字描述。

6.8

记录 record

对已发生的执行过程的文字描述、数字描述或客观证据。

6.9

调查轮次 wave

连续性项目中,在目标和项目总体设计没有改变的情况下,每次进行的连续性的重复。

6.10

陪访 accompaniment

当访问员对受访者进行访问时,由督导员在现场进行观察以便对访问员的访问过程进行评价。

6.11

试访 trial interview

在正式访问前,访问员按照要求对受访者进行的测试性访问。

6.12

数据收集复核 validation of data collection

检查数据是否按规定进行收集,并对过程或结果做出符合性质量判定的过程。

6.13

数据复核 validation of data

检查执行过程生成数据的文件,并对过程或结果做出符合性质量判定。

6.14

回访复核 backchecking

由检查人员再次联系受访者,以确认访问过程的真实性及过程或结果的符合性并做出质量判定。

6.15

监控复核 monitor to validation

通过视、听手段,对访问过程或结果的符合性做出质量判定。

6.16

监控 monitoring

通过视、听手段或通过视听设备,对现场执行过程或结果的符合性进行判定的活动。监控可以直接控制执行过程的进程,也可以事后对执行过程进行符合性判定。

6.17

卷审 questionnaire editing

由接受过专门训练的人员对受访者回答的问卷进行审核(6.24),以检验访问过程与执行要求的符合性。

6.18

无效样本 cases of ineligibility

在调查过程中发现的不属于总体的个案集合。

6.19

废卷率 nonconforming rate

质量判定受检问卷不合格的个案数量占该调查项目(或批次的)受检问卷个案总数量的百分比。

6.20

酬谢 incentive

为了提高回答率(6.30)而给予受访者的礼品、金钱或其他。

6.21

原始记录 primary records

原始数据包括未编辑的已经作答的问卷、定性研究的记录和其他类似信息。

6.22

执行过程评估 executive process appraisal

监控个体或群体工作执行能力的过程。

6.23

执行报告 executive report

执行负责人员就项目执行情况、质量情况向项目研究人员提交的说明文件。

6.24

审核 audit

系统、独立并有书面记录的过程,目的是为了获取并客观评估审核证据,以确定符合审计标准的程度。

6.25

审核标准 audit criteria

用作判定所依据的政策、流程或要求。

6.26

审核员 auditor

具备执行审核能力的人员。

6.27

不符合 nonconformity

未满足要求。

6.28

结果不符合 nonconformity product

结果形成过程未满足要求。

6.29

过程不符合 nonconformity process

过程的输出未满足要求。

6.30

回答率 response rate

完全或部分完成问卷的样本数量占全部样本数量的百分比。

7 数据处理

7.1

编码 coding

给每一个问答题的每一个可能答案分配一个代号，通常是一个数字。

注：代号可以是数字、字母、符号或其组合。

7.1.1

事前编码 pre-coding

在设计问卷时进行编码，通常给每个答案选项一个代号，数据文件用的记录格式常常放在最右边或放在某处的括弧内。

7.1.2

事后编码 post-coding

在数据整理阶段给某个没有事先编码的答案分配一个代码，通常需要事后编码的有：封闭式问答题的“其他”项、开放式(含填空)问答题。

7.1.3

自动编码 automated coding

由计算机编码或计算机辅助编码，用以全部或部分取代手动编码的一种编码形式。

7.1.4

开放式编码 open coding

对阅读资料 and 实际记录过程中获取的编码主题进行顺序编码，进而将大量零散、混杂的资料变为不同类别。

7.1.5

轴心式编码 axial coding

在开放式编码的基础上，形成类属、属性和维度，深度分析其中每一个类属并检验各类属之间关系。

7.1.6

选择式编码 selective coding

在所有类属中选择一个核心类属，然后通过一个整合图式或故事线，将各种理论要素(类属、属性、假设)整合起来。

7.1.7

编码框 code frame

对问题答案进行分类的相关编码的类别列表。

7.2

域 domain

总体中具有某种特殊性质、需要单独给出估计的子总体，也指地理上相对总体区域范围较小的区域，也称研究域。

7.3

均值 mean

描述一组数据集中趋势的测度值，又称平均数，主要有算术平均数、调和平均数和几何平均数三种计算方法。

7.4

众数 mode

一组数据中出现次数最多的数值。

7.5

分位数 quantile

将一个随机变量的概率分布范围分为几个等份的数值点，常用的有中位数（即二分位数）、四分位数、百分位数等。

7.6

极差 range

数据集中最大值与最小值之差，又称范围误差或全距。

7.7

标准差 standard deviation

方差的算术平方根，反映了一个数据集的离散程度，又称均方差。

7.8

方差 variance/deviation var

每个样本观测值与全部样本观测值的平均数之差的平方值的平均数，用来衡量随机变量或一组数据离散程度。

7.9

变异系数 coefficient of variation

标准差与平均值之比，用来度量一组数据的离散程度，又称离散系数。

7.10

协方差 covariance

衡量两个变量的总体误差。

7.11

残差 residual

实际观察值与估计值(拟合值)之间的差，也称误差。

7.11.1

绝对误差 absolute error

测量值与真实值之差的绝对值。

7.11.2

绝对误差限 absolute error limit

绝对误差的上限。

7.11.3

相对误差 relative error

测量所造成的绝对误差与被测量真值之比，以百分数表示。

7.11.4

相对误差限 relative error limit

相对误差的上限。

7.11.5

标准误 standard error of mean (SEM)

多个样本平均值的标准差，是衡量均值抽样误差大小的度量，描述了均值抽样分布的离散程度。

7.11.6

均方误差 mean-square error (MSE)

衡量估计量与被估计量之间的差异程度。

7.12

相关系数 correlation coefficient

反映变量之间相关关系密切程度的统计指标，常用的相关系数有皮尔逊相关系数和斯皮尔曼等级相关系数。

7.13

判定系数 coefficient of determination

检验回归方程对样本数据拟合紧密程度的统计量，记作 R^2 ，也称拟合优度。

7.14

算法 algorithm

对解题方案做出的准确而完整的描述，是一系列解决问题的清晰指令。

7.15 数据与数据库

7.15.1

结构化数据 structured data

根据预先定义的按某种规则构成的数据。

7.15.2

非结构化数据 unstructured data

数据结构不规则或不完整的，数据模型没有预定义的，不能够用数据库二维逻辑表来表现的数据。

7.15.3

半结构化数据 semi-structured data

没有以传统的方法进行格式化的数据，例如图、表、XML文档以及电子邮件。

7.15.4

关系型数据库 relational database

根据关系模型存放数据的数据库。

注：关系模型是由一组数据结构为二维表的关系组成的数据模型。

7.15.5

非关系型数据库 non-relational database

不遵循关系模型来存放数据的数据库。

7.15.6

分布式数据库 distributed database

利用计算机网络将物理上分散的多个数据库单元连接起来组成的逻辑统一的数据库。

7.15.7

分布式文件系统 distributed file system

能够把大量数据存储多个存储设备上并减少存储成本和复杂度的、跨多个网络系统管理文件和文件夹的系统。

7.15.8

结构化查询语言 structured query language (SQL)

用于存取数据以及查询、更新和管理关系数据库系统的数据库查询和程序设计语言。

7.15.9

数据库仓储技术 extract-transform-load (ETL)

将业务系统的数据经过抽取、清洗转换之后加载到数据仓库的过程。

7.16

数据处理 data processing

管理和转换数据从其原始状态到所需的输出的过程。

7.16.1

数据录入 data entry

将收集的数据转换成计算机可读形式或其他类型的标准化形式的处理过程。

注1：简单数据录入是不包含内置逻辑检查的数据录入。

注2：逻辑数据录入是一个数据录入过程，它使用自动检查元素的数据逻辑，例如数据类型、过滤器、问题跳过和响应选项。

7.16.2

数据编辑 data editing

验证收集的数据和必要时纠正数据的方法。

7.16.3

数据清洗 data cleaning

为达到提升研究质量的目的而识别、纠正或删除不需要的或不准确的数据。

7.16.4

数据核查 data check

尽可能地发现错误的或不合理的数据并进行一致性检查。

7.16.5

批处理 batch processing

处理大量数据(如在一段时间内收集到的一系列的交易数据)的有效方法。

7.16.6

分布式数据处理 distributed data processing

连接多个数据库单元对数据进行处理。

7.16.7

集群计算 cluster computing

使用多个服务器资源来计算或者处理数据的一个形象化术语。

7.16.8

联机分析处理 online analytical processing(OLAP)

从多角度对信息进行快速一致、交互地存取，从而更深入了解数据的一类软件技术。

7.16.9

联机事务处理 online transaction processing(OLTP)

利用计算机网络，将分布于不同地理位置的业务处理计算机设备或网络与业务管理中心网络连接，以便于在任何一个网络节点上都可以进行统一、实时的业务处理活动或客户服务。

7.16.10

离群值 outlier

看起来与数据集中的其他观察值不一致的一个观察值或一组观察值，也称异常值。

7.16.11

去重 de-duplication

删除在研究数据集中出现多次的记录或者与参与者对应的数据记录。

7.16.12

加权 weighting

为了细化或纠正样本估计的代表性，重新获得不同的单位的计算过程。

7.16.13

插补 imputation

为了解决在审核过程中检查出来的缺失数据、无效数据和不一致数据的修补过程。

7.16.14

分类 classify

对观测值按特性划分为几个子组的过程，或是从数据库到一组预先定义的、非交叠的类别的映射。

7.16.15

无量纲化 nondimensionalize/dimensionless

通过某种变换方式消除各个观测指标的计量单位，使其转化为统一、可比的变换过程。

7.16.16

标准化 standardization

把一组数据转化成均值为0、标准差为1的新数据的过程。

7.16.17

归一化 normalization

将一组数据缩放到 $[0, 1]$ 或者 $[-1, 1]$ 区间内，仅由变量的极值决定。

7.17

数据分析 data analysis

用适当的统计分析方法对收集来的大量数据进行分析，并加以汇总、理解和消化，以求最大化地挖掘数据的潜力，发挥数据的效用。

7.17.1

数据分析任务书 data analysis task book

详细列出全部必需的统计工作的可操作的清单，一般应包括：变量标识和变量值标识的编制说明、基本统计表、统计图的形式与数量、基础统计分析以及高级统计分析或模型五个方面的内容。

7.17.2

定量分析 quantitative analysis

对社会现象的数量特征、数量关系与数量变化进行分析。

7.17.3

定性分析 qualitative analysis

依据预测者的主观判断来推断事物的性质和发展趋势，常见的分析方法有专家意见、销售人员估计和市场测试等等。

7.17.4

规模分析 scale analysis

通常以频数或频率来描述样本规模的分析方法。

7.17.5

比较分析 comparative analysis

把两个相互联系的指标数据进行比较，从数量上展示和说明研究对象规模的大小、水平的高低、速度的快慢以及各种关系是否协调。

7.17.6

集中度分析 concentration ratio analysis

测度一组数据的集中趋势，反映各调查数据向其中心值靠拢或聚集的程度，其目的是寻找一组数据的代表值或中心值，以此衡量全体数据的一般水平。

7.17.7

变异性分析 variability analysis

用来测度数据的离散趋势，反映各调查数据远离其中心值的程度，同时反映各变量值之间的变异状况。

7.17.8

参数估计 parameter estimation

在满足一定精度和把握程度的条件下，利用样本信息来估计总体特征的统计分析方法，主要有点估计和区间估计两种形式。

7.17.9

列联分析 contingency analysis

通过原始数据结构，揭示品质型变量（定类变量和定序变量）之间以及品质型变量各种状态之间相关关系。

7.17.10

相关分析 correlation analysis

研究两个变量之间相关关系密切程度的统计方法。

7.17.11

回归分析 regression analysis

研究因变量对自变量依赖关系的一种统计分析方法。回归分析按照所涉及变量的多少,分为一元回归分析和多元回归分析;按照因变量的多少,可分为简单回归分析和多重回归分析;按照自变量和因变量之间的关系类型,可分为线性回归分析和非线性回归分析。

7.17.12

显著性检验 significance test

事先对总体分布形式或总体(随机变量)的参数做出一个假设,然后利用样本信息来判断该假设是否合理,即判断总体的真实情况与原假设是否有显著性差异,有参数检验和非参数检验两种。

7.17.13

参数检验 parameter test

在总体分布已知的情况下,先对总体参数提出假设,然后再利用样本信息去检验该假设是否成立。

7.17.14

非参数检验 non-parametric test

在总体分布未知的情况下,先对总体提出假设,然后再根据样本资料对假设的正确性进行判断。

7.17.15

正态分布检验 normal distribution test

判断总体分布是否为正态分布的检验。

7.17.16

独立性检验 test of independence

利用样本数据来检验两个变量的独立性,即样本资料所显示的两个分类变量的相关性在总体中是否存在。

7.17.17

协方差分析 analysis of covariance (ANCOVA)

对两个或多个样本均数有无显著差异的分析方法,也称共变量(数)分析。

注:模型的方差分析部分和回归部分不能同等对待,主要的还是方差分析部分,而回归部分只是因某些变量完全人为地控制而不得已引入的。

7.17.18

方差分析 analysis of variance (ANOVA)

推断各种因素状态对观测变量影响效应的一种统计分析方法,它通过分析数据中不同来源的变异对总变异的贡献大小,来确定可控因素对研究结果影响力的大小,可以分为单因素方差分析和多因素方差分析。

注:在方差分析中,因变量称为观测变量,自变量或控制变量称为因素。

7.17.19

因子分析 factor analysis

把具有错综复杂关系的变量综合为数量较少的几个因子,以再现原始变量与因子之间相互关系的一种降维、数据简化技术,包括探索性因子分析和验证性因子分析。

7.17.20

对应分析 correspondence analysis

利用加权主成分分析法来描述两个或多个分类变量各水平间相关性的分析方法,又称相应分析。

7.17.21

多维标度分析 multidimensional scaling analysis

将多维空间的研究对象(样本或变量)简化到低维空间进行定位、分析和归类,同时又保留对象间原始关系的数据分析方法。

7.17.22

联合分析 conjoint analysis

在已知受测者对某一受测对象整体评估结果的情形下,通过分解受测者的整体反应,来估计其偏好结构的一种分析方法,其中,受测对象由研究人员事先根据某种因子结构加以设计。

7.17.23

多指标综合评价 multi-index comprehensive evaluation

根据研究的目的,首先建立综合评价指标体系,然后基于评价指标体系,对现象的各个方面进行定量分析,得出一个整体性的评价结果。

7.17.24

德尔菲法 delphi method

由调查者拟定调查表,按照既定程序,以函件的方式分别向专家组成员进行征询而专家组成员又以匿名的方式(函件)提交意见;经过几次反复征询和反馈,专家组成员的意见逐步趋于集中,最后获得具有很高准确率集体判断结果。又称专家评估法。

7.17.25

层次分析法 analytic hierarchy process(AHP)

在将与决策有关的元素分解成目标、准则、方案等层次的基础上,进行定性和定量分析的决策方法。

7.17.26

主成分分析 principal component analysis(PCA)

通过线性组合的形式,从多个指标中尽可能多的提取有效信息,以较少主成分保留大部分指标信息的数据分析方法,也称主分量分析。

7.17.27

判别分析 discriminant analysis

以过去划分类别的有关资料和某种最优准则为依据,确定一种判别方法以判定样本归属类别的数据分析方法。

7.17.28

时间序列分析 time-series analysis

通过对时间序列的长期趋势、季节变动、循环变动和不规则变动四部分进行观察和研究,然后总结时序数据的内在规律,对未来的发展趋势进行预测。

注:时间序列(或称动态数列)是将同一指标的数值按其发生的时间先后顺序排列而成的数列。

7.18

机器学习 machine learning

研究如何使用计算机模拟或实现人类学习行为,并重新组织现有内容的知识结构来提高学习效率。

7.18.1

监督学习 supervised learning

利用一组已知类别的样本调整分类器的参数,使其达到所要求的性能的一种机器学习方法。

7.18.2

无监督学习 unsupervised learning

利用无标签的数据,学习数据的分布或数据与数据之间关系的一种机器学习方法。

7.18.3

半监督学习 semi-supervised learning

使用大量的未标记数据,以及同时使用标记数据,来进行模式识别工作。

7.18.4

逻辑回归模型 logistic regression model

通过某一逻辑函数将自变量的线性回归结果转换成概率值的一种广义线性回归分析模型。

7.18.5

支持向量机 support vector machine (SVM)

按监督学习方式对数据进行二元分类的广义线性分类模型,其决策边界是对训练样本求解的最大边距超平面。

7.18.6

决策树 decision tree

呈树形结构，且基于一系列规则对样本进行分类的算法。

7.18.7

随机森林 random forest

利用多棵决策树对样本进行训练并预测的一种分类算法。

注：决策树算法随着数据集变大，往往会过度拟合模型，而随机森林是一种旨在减少过度拟合的决策树算法。

7.18.8

人工神经网络 Artificial Neural Network (ANN)

由大量的节点(神经元)之间相互联接构成的运算模型，模型中每个节点代表一种特定的输出函数(激励函数)，每两个节点间的连接都代表一个对于通过该连接信号的加权值。

7.18.9

聚类分析 cluster analysis

依据某种准则对个体(样本或变量)进行分类的一种无监督学习方法，又称群分析或类分析，包括Q型聚类分析和R型聚类分析。

注：分类结果满足：每一个类别至少包含一个数据记录，每一个数据记录属于且仅属于一个类别。

7.18.10

典型相关分析 canonical correlation analysis

通过识别并量化两组变量之间的联系，将两组变量相关关系的分析转化为两组变量的各自线性组合之间相关关系分析方法，常用于研究两组变量间的内在联系。

7.19

数据挖掘 data mining

从存放在数据库、数据仓库或其他信息库中的大量的数据中获取有效的、新颖的、潜在有用的和最终可理解的模式的非平凡过程，又称知识发现。

注：非平凡指它已经超越了一般封闭形式的数量计算，将包括对结构、模式和参数的搜索。

7.19.1

关联规则 association rule

形如“由于一个事件的发生而引起另外一个事件发生”之类的规则。

7.19.2

购物篮分析 market basket analysis

通过支持度和置信度这两个值来对顾客所购买的商品组成情况进行分析。

7.20

情感分析 sentiment analysis

研究人们针对实体(包括产品、服务、组织、机构、事件、话题等)表达出来的主观观点、情感、评价、态度和情绪的分析方法,也称观点挖掘。

7.21

决策支持系统 decision support system

辅助决策者通过数据、模型和知识,以人机交互方式进行半结构化或非结构化决策的计算机应用系统。

7.22

数据可视化 data visualization

关于数据的视觉表现形式的研究。

7.23

商业智能 business intelligence (BI)

用现代数据仓库技术、线上分析处理技术、数据挖掘和数据展现技术进行数据分析以实现商业价值,又称商业智慧或商务智能。

8 调查报告

8.1

数据发布 data dissemination

通过各种媒介将调查收集到的数据及调查报告等资料向用户提供或向公众公开。

8.2

报告演示 report demonstration

正式向用户提交调查报告(发布调查信息)之前,调查机构以演示会的形式向用户演示调查报告的详细内容。

8.3

调查报告 research report

根据调查目的,将调查资料通过文字、图、表格等形式组织而成的主题鲜明、逻辑清晰的有观点和建议的文字材料。

8.3.1

书面调查报告 written investigation report

详细而具体的调查报告,主要用于市场调查最终成果的汇报交流,具有“证明调查过程的科学性与真实性”、“阐述调查结果和提出建议”以及“文献功能”等三个功能。

8.3.2

口头调查报告 oral investigation report

书面调查报告的浓缩版本，其主要作用是作为一个正式的交流文本，在向调查委托方进行正式汇报时使用。

8.4

一般性报告/执行报告 general report / executive report

报告内容只包含调查结果，仅提供给对调查结果感兴趣的对象的调查报告。

8.5

技术报告/科学报告 technical report / scientific report

内容包含调查中详细科研细节的调查报告。

8.6

数据质量报告 data quality report

对整个市场调查项目，从数据的精确性和可靠性等方面对调查的最终产品进行评估，形成的书面材料。

8.7

数据分析报告 data analysis report

用户展示对调查数据进行描述性分析和推断性分析的分析结果的书面报告，一般需要包括关于分析方法的简述、对抽样误差和非抽样误差的描述和讨论以及调查结果的统计意义。

8.8

统计表 tables

以纵横交叉的线条所绘制的表格来表现统计资料的一种形式。

8.9

频率分布表 tables of frequency distributions

把总体中各个类别及其相应的频数、频率及累计频率等指标用汇总表格的形式展现出来的统计表。

8.10

列联表 contingency tables

对两个或两个以上的分类变量进行交叉分类所形成的复合频数分布表。

8.11

统计图 charts

用于直观地显示数据、集中强调数据的相对模式以及反映数据的分布形状、结构和数量级别等的图形或图形组合。

8.12

泄露 disclosure

根据已公布的信息能够直接地或间接地、精确地或近似地推断出自己提供、或他人提供、或还未公布的信息。

8.13

泄露控制 disclosure control

为避免数据公布侵犯相关方的隐私或商业秘密而采取的措施,在确保被调查者的隐私受到尊重的同时,最大程度地保持数据的效用。

8.14

身份泄露 identity disclosure

在公布数据时将某项调查结果与特定的被调查者(家庭、个人或公司)联系在一起,这类信息受相关法律保护。

8.15

公用微观数据文件 public-use microdata files

对外公开的包含所有(或部分)受访者回答记录的数据文件,但不包含受访者的身份识别信息。

8.16

完整调查资料集 all relative research files

包含在调查启动、调查设计、数据收集、数据处理、数据发布等各个调查实施阶段所形成的书面或电子文件,一般包括调查计划书、调查方案、培训资料、数据处理手册、各种调查报告和各种数据文件等。

参考文献

- [1] 《市场调查理论与方法》，高等教育出版社2018版。