**大数据会对数据收集产生什么样的影响?**

大数据，也称巨量资料，指的是需要新处理模式才能具有更强的决策力、洞察力和流程优化能力的海量、高增长率和多样化的信息资产。大数据时代的数据收集相较于传统数据收集发生了深刻的变革，大数据对数据收集来说，即是机遇，也是挑战。

**一、大数据时代下，数据收集发生转变**

1.收集到的数据类型更加丰富。传统数据收集得到的一般是结构化数据。大数据收集的对象为一切可记录、可存储的信息，不仅包括结构化数据，还包括非结构化和半结构化数据等形式，难以通过传统统计指标或者图表表现，借助现代技术可以挖掘其中的价值。

2.数据收集方式发生改变。相比传统数据收集的抽样调查方式而言，大数据采集技术通过传感器、社交网络、移动互联网等方式获得各种类型的海量数据。大数据来源更加多样且易于获取，内容也更加完整丰富，数据收集效率大幅提高。因此，从某种意义上讲，从大数据中收集数据就是识别、整理、提炼、汲取、分配和存储元数据的过程。

3.对于数据不确定性的认识改变。传统数据收集的不确定性来源于调查方案的设计、抽样的随机性、观测的偏差等。而大数据收集的数据几乎全覆盖，它的不确定性更多源自于收集数据时的数据形式多样性、内容真伪性、数据质量优劣性。如何有效地比较甄别，以提取出真正需要的数据是数据收集的主要任务。

4.更注重事物之间的相关关系。在统计预测时，传统统计偏爱因果关系，大数据时代倾向更简单的相关关系。另外，由于大数据的数据收集技术的应用，对于捕获相关关系也更容易。

**二、大数据时代的数据收集优势与缺陷并存**

大数据背景下，我们收集到的数据类型的更加丰富，当万物皆数据的时刻到来时，商业发展的更多新机会将会出现，数据将会帮助我们更好地做出判断。当有些表面上完全不相关的行业数据收集关联起来时，便会产生新的商业价值，行业之间的联系将会变得更加密切。但同时，大数据时代的数据收集也存在缺陷：

1.大数据存在覆盖漏洞，需用传统方法互补。大数据覆盖了数据的大部分区域，但由于信息的延展性，大数据集永远存在“数据黑暗地带”或“数据阴影区域”。

2.大量的无效信息往往与有效信息混杂，对研究者的统计推断产生了严重干扰。

3.准确度不够。大数据收集的主要途径是通过互联网中的关键词搜索进行相关统计分析，但是在数据的实际收集过程中，研究的复杂问题并不能仅用个别关键词就能真实反映。

4.法律法规不全面，导致暴力收集数据和隐私泄露。法律缺少关于明确数据收集主体资格与责任、限定数据信息收集的对象与范围等内容，避免不良商家过度挖掘隐私，恶意掌握个人行为与习惯，进一步可能演化为网络诈骗等违法行为。

**三、如何进行大数据时代下的数据收集**

考虑到传统数据和大数据收集的针对性、安全性和成本等比较问题，我们既要继续采用传统的方式方法去收集特定需要的数据，又要善于利用现代网络信息技术和各种数据源去收集一切相关的数据，并善于从大数据中进行再过滤、再选择，互相补充，以得到最为接近现实的数据。

对政府统计部门而言，应整合利用各类大数据资源，建立一套全国性大数据资源库，建立一套快速高效的统计调查平台；解决大数据与传统调查方法，特别是抽样调查的辅助、融合推断问题。政府统计系统在进行大数据采集时，要充分利用统计系统自身的大数据，开发利用好普查数据，加强、完善名录库建设；充分利用好其他政府部门的行政记录数据，加强沟通、协调与合作；与掌握大数据的公司战略协作，争取开放数据接口；利用网络爬虫等技术开展数据采集。

引用邱东老师的观点：如果说原来的统计实务的重心在于收集基础数据，难点在于如何获取数据，那么在大数据时代，则更在于如何选择有用数据。如果说原来重心是“做加法”—无中生有；那么现在重心则是“做减法”—“有中生用”。如果说大数据是一座大矿山，我们要做的就是在这座矿山中去伪存真、去芜存菁选取出我们想要的数据。

**设计一份统计调查方案应该考虑哪些内容?**

统计调查方案设计就是统计调查组织和实施者在进行实际调查之前，根据调查研究的目的和调查对象的性质，对调查工作总任务的各个方面和分阶段进行的整体设计，提出相应的调查实施方案，制定出合理的工作程序。设计统计调查方案要遵循科学性原则、可行性原则和有效性原则。

设计一份统计调查方案应该考虑以下内容：

**(一)调查目的：**明确调查目的是确定调查的范围、内容和方法的前提。

**(二)调查对象和调查单位：**调查对象是根据调查目的和任务所确定的调查范围及所要调查的总体，由某些性质上相同的若干调查单位组成。调查单位是调查对象中的每一个具体单位，它是调查中要调查登记的各个调查项目的直接承担者。

**(三)调查内容：**调查内容指调查组织和实施者对调查单位进行调查的主要内容。确定调查内容，就是要明确调查组织和实施者需要向被调查者了解哪些问题。确定的调查内容应既是调查任务所需，又是便于取得调查答案的问题。调查内容的表达方式必须明确，备选答案应有明确的表示形式。确定调查内容时，应尽可能围绕某一需要调查的核心问题而设置一些相互关联的调查项目。调查内容的涵义明确、肯定，必要时可附加调查项目解释。

**(四)调查提纲和调查问卷：**调查项目确定后，即可将调查项目进行分类、排列，在此基础上形成调查提纲或调查问卷，以方便调查登记和汇总。

**(五)调查时间和调查工作期限：**调查时间指调查资料所要求的时间；调查期限指整个统计调查工作的开始时间和结束时间。

**(六)调查地点：**在调查方案中，还要明确规定调查地点。调查地点与调查单位通常一致，当不一致时，有必要规定调查地点。

**(七)确定调查方式和方法：**调查中所采用的调查方式和方法不是固定不变的，根据调查目的和其他一些约束条件的允许，调查方式和方法也可以进行灵活处理。调查方式包括普查、抽样调查、重点调查、统计推算等，调查方法包括直接观察法、通讯法、采访法、登记法等。数据收集方式和方法的选择，关系到我们能否及时、准确、完整地收集到所需的统计数据，也关系到我们在调查过程中要投入多少人力、物力、财力，因此，要根据研究的目的和我们的能力，选择合适的数据收集方式和方法。

**(八)调查资料整理和分析的方法：**调查资料的整理包括定性资料整理和定量资料整理两个方面。常用的分析方法有：频数分析、交互分析、回归分析、相关分析、聚类分析、因子分析等。每种分析方法都有其自身的特点、适用性和相关的技术要求，在分析时应根据调查的要求，选择最佳的分析方法，并在方案中进行规定。

**(九)调查报告内容和提交的方式：**调查报告内容主要包括调查报告书的形式和份数，报告书的基本内容、报告书中图表量的多少等。调查报告提交的方式指统计调查组织和实施者以何种形式向用户提交调查报告，一般有电子资料方式、光盘资料方式、纸质资料方式等。

**(十)调查的组织计划：**调查的组织计划指为确保实施调查而制定的具体工作计划。主要包括调查机构的设置、人员的选择和培训、工作步骤及其善后处理等。

**(十一)调查经费预算：**统计调查费用的多少通常应根据调查的目的、调查的范围和调查的难易程度而定。

**(十二)进行调查方案的可行性研究：**可通过逻辑分析法、经验判断法、试点调查法等方法。

**(十三)调查方案的总体评价：**调查方案一般都应满足以下三个要求：方案设计是否体现调查目的和要求；方案设计是否科学、可行和有效；方案设计能否使调查质量有所提高。

**你平时收集资料(数据)的渠道有哪些?**

**一、次级资料的收集**

次级资料也叫二手资料，是其他人或机构为了其他目的而收集、记录和整理出来的有关资料。次级资料的收集可以采用文案调查法。

文案调查资料有两类，来源不同。第一类是内部资料，指那些源自机构内部的数据，或者是由本机构记录的数据，其来源主要包括：内部记录（业务资料、统计资料、财务资料）；顾客反馈；内部数据库。外部资料是指其他机构而非调研人员所在机构收集或记录的资料，其来源主要包括：出版物（书籍、报纸、期刊等出版物）；计算机数据库（文献数据库、数据类数据库、指南性数据库）；中介机构数据；互联网等。

次级资料收集的实际渠道包括：

1.官方机构：各地统计局--统计数据、分析报告

2.学术单位：学校图书馆、地方图书馆--统计出版物、期刊、数据库

3.行业资料：行业报刊、杂志等、行业专业人士访谈或研讨会、行业专项调查、现成的研究报告、其他市场研究公司资料服务研究等

4.网上资料：搜索引擎(google)、网络数据库(资讯行)、专业论坛(人大经济论坛)、开放式文库(百度文库)等

**二、原始资料的收集**

原始资料是调研人员为满足特定目标而收集的第一手资料。收集原始资料的调查是实地调查，实地调查分为定性调查和定量调查。

定性调查主要应用于探索性调查，调查的目的是考察被调查者的态度、感觉、动机、反应等。在定性调查中，调查员根据讨论提纲或讨论指南进行调查，所收集的数据大多是难以量化的定性数据。具体方法包括：观察法（人员观察法、机器观察法），小组座谈法，深层访谈法，德尔菲法和投影技法等。

定量调查主要应用于描述性、因果性和预测性调查，调查的目的是对总体的数量特征进行推断。在定量调查中，调查员一般根据调查问卷进行调查。具体方法包括：电话调查（传统电话调查、计算机辅助电话调查），面访调查（入户面访调查、街头/商城拦截面访调查、计算机辅助面访调查），邮寄调查，网络调查（网站/网页调查、电子邮件调查、弹出式调查、网上固定样本调查、微信推送调查、一对一的网上深层访谈、网上小组座谈会、网上观察、网上文献资料分析），实验法（经典设计、统计设计）。

**调查方法与调查组织形式有区别吗?各有哪些?**

调查方法是指统计机构和统计人员搜集统计资料的方法，样本单位确定之后，对这些单位实施调查，得到所需要的数据，可以采用不同的方法，这里的方法就是统计调查方法。调查组织形式是指组织统计调查，搜集统计资料的方式方法，统计调查组织形式取决于调查对象的特点以及调查的目的和要求。

**两者存在区别：**调查组织形式是抽象的，相当于一个理论；而调查方法是具体的，相当于一个实际活动方法。例如，我们对某一类情况进行调查，首先要选择一个合适的调查组织形式，从而在一个总体中选出一个样本；其次，对于这个样本中的每个个体，我们要选择合适的调查方法，以得到想要的数据。

**调查方法**包括以下几类：

1.直接观察法：调查人员亲临现场对调查单位的调查项目进行清点、测定、计量，并加以登记，以取得第一手资料的一种方法。

2.报告法：调查单位或报告单位利用各种原始记录、基层统计台账和有关核算资料作为报告依据，按照隶属关系，逐级向有关部门提供统计资料的一种方法。

3.采访法：由调查人员对被调查者进行采访，根据答复来收集数据的方法。

4.通讯法：把要调查的问题邮寄或电子传送给被调查者，填好后返回。

5.文献法：根据调査目的，浏览著作、报告、论文、统计或业务报表等，获得所需要的研究信息。

**调查组织形式**包括以下几类：

1.统计报表：按照国家相关法律的规定，自上而下统一布置，自下而上地逐级定期提供基本统计资料的一种调查方式。

2.专门调查

（1）普查：根据特定的统计研究目的而专门组织的一次性的全面调查，用以收集所研究现象总体的全面资料。

（2）重点调查：是一种非全面调查，是对数据收集对象总体中的部分重点个体进行观测的统计调查方式。重点单位是标志量占标志总量比重较大的少数个体。

（3）典型调查：根据调查目的，在对研究对象总体进行全面分析的基础上，有意识地从中选取若干个总体单位进行系统周密调查研究的一种非全面调查。

（4）抽样调查：一种非全面调查，从总体中抽取样本，以样本推断总体。根据抽取样本的方式不同，抽样调查可分为概率抽样和非概率抽样两类。

数据收集方式和方法的选择，关系到我们能否及时、准确、完整地收集到所需的统计数据，也关系到我们在调查过程中要投入多少人力、物力、财力，因此，要根据研究的目的和我们的能力，选择合适的数据收集方式和方法。

**统计调查与实验设计有什么区别及联系?**

统计调查是指根据统计研究的目的和任务，运用科学有效的调查方法，有计划有组织地向调查对象搜集原始资料和次级资料的活动过程。实验设计是对实验进行科学合理的安排，已达到最好的实验结果，根据内容的不同，可以分为专业设计与统计设计。

统计调查与实验设计在以下几个方面存在**区别：**

1.应用范围：统计调查更加广泛地应用于社会经济现象，几乎所有与社会经济现象有关的统计数据都是观测数据，如GDP、CPl、房价等。实验设计更多的是应用于理工农医各领域，在实验中控制实验对象以及其所处的实验环境，收集到的是实验数据，如一种新药疗效的数据。

2.研究方式：统计调查只是研究者“被动”地观察研究对象的实际情况，而未施加任何干预。实验设计则是研究者“主动”地对观察对象施加某种人为干预，即对一些外来变量进行控制，而改变自变量来研究其对因变量的影响，而不是消极被动地等待研究对象的发生或改变。

3.因果关系：实验设计可以通过对外来变量的控制，改变自变量而研究其对因变量的影响，并通过与对照组的对比，验证两个变量之间存在相关关系和前后相继关系，且不存在其他的原因，即检验两个变量之间的因果关系，这是一般统计调查所不能实现的。若采用统计调查来探究因果关系的话，只能对所获取的数据进一步进行分析，如进行各种回归分析。

4.研究结果：相比较而言，实验设计的科学性更强，精确度更高。实验设计通过有效地控制实验环境，进行反复研究，取得的数据比较客观，具有一定的实用性和可信度，也提高了调查的精确度。

统计调查与实验设计也存在**联系：**

总体来说，统计调查与实验设计都是科学研究经常使用也非常实用的两种数据收集方法，二者都是通过对资料的收集、分析，解决一定的问题。二者相辅相成，为研究提供数据支撑，保证研究的顺利进行。

**大数据时代还需要抽样调查吗?**

大数据时代仍需要抽样调查。

大数据正多方位改变着人们的认识、思维和行为。维克托﹒迈尔﹒舍恩伯格提出大数据时代要全体不要抽样，诸多学者和大众都表示赞同，认为以往因无法获取总体所以需要抽样，大数据时代“样本＝总体”，要分析与某事物相关的所有数据，毕竟抽样就意味着有抽样误差的存在。进入大数据时代，我们能对全体数据进行分析，不一定要抽取样本，但这并不意味着抽样就要退出历史舞台。

**（一）大数据并非完美，仍需要抽样**

当所获取的“大数据”总体不能完全代表目标总体时，有一部分个体被遗漏在外，大数据分析可能得到有偏结果。首先，目前来看并非所有数据都可以通过网络信息系统获得，因为并不是所有产业都已实现智能化，还有很多数据只能通过传统的抽样方式获得；其次，即使是网络数据，某些情况下对总体进行分析也并非最优选择，例如当面临均匀度很大的总体时，随机抽取部分单位作为样本开展分析就可以得到理想结果，此时并不需要去费时费力分析总体。大数据带来的既有信息也有噪声,通常噪声居多，使得数据分析容易为假像所迷惑，造成规律丧失和失真。大数据的生成与采集在人为设计的框架之下，也可能存在系统性偏差，譬如社交网络中的人群数据与真实总体之间可能存在明显偏差。所以,大数据条件下随着众多缺失、含有噪声甚至是错误的数据进入到数据库中,从中抽取部分样本数据更能有效进行数据清洗，以挖掘出数据背后的真规律。

**（二）大数据与抽样的选择至关重要**

如果通过抽样能够显著降低数据处理的复杂程度，同时解决问题的效果没有太大的下降，那么应该采用抽样的方式。如果随着采样率的降低，解决问题的效果会快速下降，则应利用大数据。如果问题的处理效果随着数据量上升有一定提升，但当数据大到一定规模后效果提升并不明显，则应该选取一个有较大规模但并非全体的数据集来处理。实际当中也应该考虑研究问题的性质，对大概率事件通常选择抽样的方式，如新政策的支持率；对小概率事件应使用尽可能多的数据，如信用卡诈骗问题，只有掌握了全体数据才能找出异常情况。

**（三）大数据与抽样相互补充**

大数据时代既要善于利用现代网络信息技术和各种数据源去收集一切相关的数据，又要采用传统的抽样方式去收集与处理特定需要的数据。大数据拓宽了抽样调查数据采集渠道，为抽样框的及时更新维护提供了便利条件。统计机构通过抽样调查所获取的数据具有权威性，可作为对照基础和验证依据与大数据进行对比，开展抽样数据对互联网数据的校正与调整。

大数据与抽样并非相互排斥，一定程度上存在着互补关系，大数据条件下合理利用抽样技术能够更有效地开展数据收集和分析，更好地释放大数据的能量，挖掘大数据的价值。未来，大数据与抽样调查将相互补充，携手前行。

**你认为选择调查组织形式最重要的依据是什么?**

统计调查组织形式是指组织统计调查、搜集统计资料的方式方法，可分为统计报表和专门调查，其中专门调查又可以分为普查、典型调查、重点调查和抽样调查。选择何种调查组织形式会受到多项因素的影响。

**1.调查经费：**调查组织形式的选择首先会受到调查经费多少的影响，若调查经费不充足，则某些调查将无法正常开展。

**2.调查目的：**若不受调查经费制约，则应根据调查目的选择调查组织形式。如果是国家为了取得国民经济和社会发展情况基本统计资料而进行调查，那么应该选择统计报表形式，由各级企事业、行政单位按规定的表格形式、内容、时间要求报送。如果一个国家或地区要详细调查某项重要的国情、国力，如农业、人口情况等，则应该选择普查。

**3.调查对象本身的特点：**如果调查对象中含标志量占标志总量比重较大的少数个体，则可选用重点调查。对抽样调查而言，如果对调查对象的资料进行分类整理后，类间方差较大，类内方差较小，则用分类抽样；如果不仅有详细的分类资料，而且有详细的按有关标识大小排队的资料，则宜采用等距抽样方式；相反如果调查对象的单位较分散，且经过分析群间方差较小，群内方差很大的情况，则用整群抽样方式。

**4.对调查对象的了解程度：**如果对调查对象有详细的了解，则可根据需求选取典型，采用典型调查方式。如果对调查的总体毫无了解，一般用纯随机抽样法。

**5.效率原则：**有些情况下，从一般的直观观察很难判断用哪一种方式较好，那就要具体计算各种方式的抽样误差的大小和抽样数目的多少。一般来说,应当采用抽样误差较小，抽样数目最少的抽样方式。各种抽样方式不同，所需要的人力、物力和财力等也不同。有些抽样方式虽然精确度高，但耗费人力、物力、财力太大，而且实际工作也不一定要求那么高的精确度。这就要求在一定的抽样容许误差的范围内，尽量采用节约人力、物力、财力的组织形式。

**在进行抽样设计时，样本容量如何确定?**

进行抽样设计时需要确定样本容量，影响样本容量的因素有很多，应综合考虑。

**（一）根据理论依据确定样本容量**

**1.单位标志变异程度：**单位标志变异程度一般用方差或成数方差*P*(1−*P*)的大小来表示，在其他条件不变的情况下，总体单位标志的变异程度大，样本容量应越大。

**2.抽样极限误差：**抽样极限误差是指在一定的把握程度下保证样本指标与总体指标之间的抽样误差不超过某一给定的最大可能范围，一般用Δ表示。在其他条件不变的前提下，所允许的抽样极限误差越小，即抽样估计的精确度要求越高，样本容量应越大。

**3.抽样推断的可靠度：**抽样推断的可靠度是指总体所有可能样本的指标落在一定区间的概率度，用表示，一般简写为t。在其他条件不变的情况下，抽样估计所要求的可靠程度越高，所需的样本容量就越大。

**4.样本抽选方法和抽样组织形式：**在同等条件下，不重复抽样比重复抽样需要的样本容量小一些；采用分层抽样和系统抽样要比简单随机抽样需要的样本容量小一些。

例如，在简单随机重复抽样条件下，用样本平均数和成数去推断总体相应指标值时所需要的必要样本容量为：

在简单随机不重复抽样条件下，用样本平均数和成数去推断总体相应指标值时所需要的必要样本容量为：

在抽样方案设计中确定必要的样本容量时，会出现所需数据资料缺乏等应引起我们注意的问题，主要有：第一，公式中总体方差是未知的，可用样本数据、经验数据、试验性或试调查数据代替，但在有几个数据资料可以选择时，应选择方差大的。特别是有关成数方差的资料完全缺乏时，可用成数方差的大值0.25来代替。第二，在通过一次性抽样调查来实现对总体平均指标值和成数P进行估计的预期目标时，如果用它们各自样本容量的计算公式得出的结果不一样，就应选择两个数中较大的一个为必要样本容量，即这两个数中的较大者是应抽取的样本单位数目的下限。有时也可以考虑设计效应或依据经验和一些基本规则来确定样本容量。

**（二）考虑现实因素确定样本容量**

因为市场研究的复杂性和多样性，根据公式计算出的数值只是作为一个参考基础，样本量的确定还需要综合考虑以下几个方面的因素。

**1.无回答情况：**在无回答率较高的调查项目中，样本容量应大一些。

**2.调研的目的：**实际的市场研究中，有些调研目标比较单一，较小的样本量就可以满足要求；而有些调研需要了解全方位、多方面的信息，所需要的样本量就较大。

**3.调研的产品：**一般而言，对于消费人群比较广泛的产品，如饮料、化妆品等，所需要的样本量较大；而对于消费人群比较集中的产品，如汽车、信用卡等，所需要的样本量相对较小。

**4.子群分析的需要：**市场研究中，经常要通过市场细分研究了解不同群体的消费习惯和消费需求。因此，设计样本量时，必须保证各子群有足够的样本量可供分析。一般为了显著性检验的需要，子群分析的样本通常不少于30个。

**5.调研的区域：**调研区域越大，可能需要的样本量越大，如一般中心城市的调研，样本量可以设计在300-600个；而对于小城市、县城的调研，样本量可以设计在100-200个。但样本容量不宜过大或过小，以免造成浪费或推断价值不足。

**6.调研成本和时间：**时间和费用较为宽松的情况下，可以适当向调查精度方面倾斜，所需要的样本量较大。反之，则设计样本量较小，但是必须保证能够满足调查精度的最低要求。

最终样本量的确定需要在精度、费用、时限和操作的可行性等相互冲突的限制条件之间进行协调。它还可能需要重新审查初始样本量、数据需求、精度水平、调查计划的要素和现场操作因素，并作必要的调整。

**你认为是调查的具体实施重要，还是方案更重要?**

调查方案和调查的具体实施同等重要，设计调查方案是实施统计调查的前提，科学、周密的调查方案，是统计调查得以顺利实施的可靠保证，二者相辅相成，共同推动统计调查达到最终的成功。

调查方案是指在正式调查之前，根据调查的目的和要求，对调查的各个方面和各个阶段所作的通盘考虑和安排。调查总体方案是否科学、可行，将影响到整个调查工作能否顺利进行。首先，只有明确调查目的和任务，才能确定调查的对象、内容和方法，保证调查具有针对性。其次，只有正确、科学确定调查对象，才能划清要研究的总体界限，这对于保证调查资料准确反映实际情况是十分重要的。再者，不同的问题适用调查方法和组织方式不同，所以选择合适调查方法、方式是很重要的。并且，合理的调查经费预算能够避免经费不足等问题对调查进度的影响。最后，调查方案经过可行性研究，尽可能地避免后续实施及调查结果的低效性。总之，合理的调查方案设计是调查实施的前提。

调查的具体实施是调查方案设计的应用。无论方案设计多么完美，具体实施工作若做不到位，将会功亏一篑。在调查前应组建分工明确的调查团队，对调查员进行培训，做好必要的前期准备工作，准备必要的辅助工具。有时还应该进行试调查，以检验调查方案设计是否合理。在实际调查过程中，会产生随机误差和系统偏差，要遵守工作计划一步步进行，还要根据实际情况进行灵活变通，尽可能围绕调查目的以“不变”应“万变”。同时，要时刻调整调查进度，注意成本控制。

统计调查是一项复杂细致的工作，一项统计调查，往往需要组织许多的人员参加设计详细的调查工作方案与实施计划，有步骤、有组织地开展起来。因此，没有科学、严密的调查方案和有效的的实施，要取得预期的效果是不可能的。所以，在组织调查之前以及调查过程中，必须周密地设计调查方案并做好实施，以保证统计凋查任务的完成。好的调查方案能够降低具体实施过程中产生的误差，调查的具体实施效果直接关系到调查数据的质量。调查方案和具体实施同等重要，在调查过程中都应该得到重视。

**调查阶段要提高数据质量需注意哪些方面?**

在统计数据来源多样化的情况下，统计数据的可靠性和有效性直接影响到统计分析和预测能否得出正确的结论。所以统计数据质量的提高显得尤为重要。

影响数据质量的因素有调查的内容（调查内容的敏感性、问题的难度、问题的兴趣），调查的条件和环境（调查的开展时间、调查地点、被调查社区对调查的态度），应答者的特征（应答者的社会特征、应答者对入户访问的态度和合作程度、应答者对提问的理解能力、应答者的回答能力及方式），调查者的特征（调查员的社会特征、调查技术的掌握程度、调查员的主观能动性和工作责任心）等，可从这些因素入手来提高数据质量。

（一）在**调查组织阶段**应注意以下方面：

1.抽取有代表性的样本。

2.根据研究目的设计调查问卷：有填表说明；问题的意思应具体明确，没有歧义；问题尽量口语化，通俗易懂；问题尽量采用封闭式的选择题；敏感问题排在后面。

3.调查员的选择与培训：对调查员进行严格的选择和培训，须使每个调查员都能准确的理解调查的目的和要求，且在调查实施之前须对被调查群体的特征有初步了解，比如调查的背景等。

4.组织试调查：可检验调查方案的可行性。

（二）在**调查实施阶段**应注意以下方面：

1.调查初期，要及时小结，讨论遇到的问题，统一解决的标准。

2.与被调查者建立密切融洽的关系；调查员态度要端正，要让被调查者明白你的来意及其个人信息将会得到保护；做好被访者的工作；保持中立，避免诱导和臆测；尽量询问每个成员及家庭主要能说清情况者，减少无回答和回答出现偏误的误差，设法提高应答率。

3.做好初审、复审，及时发现错、漏项，予以改正、补充。比如在访问调查中，调查员应掌握好面访的时间，合理安排每天的面访次数，并在两次面访之间留出足够时间，补全访问中未完成的记录，以及完成所必须的审核工作。入户调查中，可以采取电话复核和实地复核等做法。

4.注意减少环境因素带来的影响。

5.现场数据收集质量控制四点。

（三）在**整理资料阶段**应注意以下方面：

1.人工审核问卷上的数据，检查出明显错误及数据不一致。

2.编码，把填写的文字信息转换为数字代码，以便于分组处理；发现有缺项与错项应要求所属地区负责核查更正，如无法更正者，则将其作废。

3.数据采取两次录入，完全一致的数据才能进入推算。

4.计算机逻辑检错，进行一致性审核、有效性审核与分布审核。

（四）**大数据背景下**应注意以下方面：

1.流程维度：建立科学的统计数据质量评估体系；建立基于数据仓库的统计数据集成平台；尊重客观统计规律，拓新统计技术与方法。

2.技术方面的解决对策：研究高效率、低成本的数据存储方式；将数据质量检测技术贯彻到数据整个生命周期；将大数据分析技术运用到统计工作的各个方面。

3.管理方面的解决对策：树立科学合理的统计观念；加强统计法制建设；完善数据标准；完善统计制度体系。