

谢宇教授
方法论系列
讲座

整理：杨志宏

社会科学与自然科学 的关系

Relation of social science to natural
science

主讲：谢宇教授（美国密西根大学社会学系教授）

主持：马戎教授（北京大学社会学系主任，博士生导师）

时间：2002年10月21日晚7:00 - 9:00

地点：五四体育馆体教电教厅

马戎主任：我们今天的讲座现在开始，这是北大社会学系为庆祝建系 20 周年举办的学术活动月系列讲座的一部分。今天我们非常荣幸地请到了美国密西根大学社会学系谢宇教授来给我们做他的第一次讲座。我想从事社会学研究的人对谢宇教授应该不会陌生，下面我把谢宇教授的生平简历给大家做一个简单的介绍。1982 年，谢宇教授在上海工业大学获得了本科学士学位；1984 年，在威斯康星大学获得了社会学、科学史两个硕士学位；1989 年，获得了威斯康

星大学的社会学博士学位，并且在这一年还被聘为社会学系正教授。之后，谢宇教授在美国许多著名的学校担任教职工作，并在很多方面做出了突出的成绩，成为美国很多著名杂志的主编或副主编，得到过很多教学奖和研究奖。2000年，谢宇教授出版了他的学术著作，另外在明年，谢宇教授还将出版他的另一部学术著作。而谢宇教授的学术论文则多得数不胜数，我就不一一列举了。下面，我们用掌声欢迎谢宇教授给我们做第一讲报告。

谢宇教授：首先，感谢大家来听我的讲座，感谢马戎主任和他的同事对我

的邀请，也感谢周浩博士花了很多时间和精力帮我把英文讲义翻译成了中文。我一共要做六个报告，前五讲是关于方法论的东西，最后一讲我讲一下关于美国社会和美国社会学的发展趋向。今天这一讲是第一讲，这一讲比较抽象，有很多哲学的东西，最重要的是思想的东西，是观念上的东西，而不是技术上的东西。假如你想学技术的东西，我有这方面的教材，在座的同行中也有很多是做这方面研究的。我讲的思想上的，观念上的东西可能和你以前学的不一样，这是我多年来对方法论的理解，对社会科学和自然科学的了解和自己的一些体

会。我从 1989 年到现在已经教了 13 年的书了，我教书喜欢用相互交流的方式，给这么多人做讲座对于我而言还是第一次。中国到底是人多，对社会学感兴趣的人也多。我习惯于在小课堂讲课，这么大的讲座不便于交流，因此我会采用两种方式来弥补，一是我会提一些问题，由大家来回答，另外，我住在勺园 6 楼 405 室，感兴趣的同学可以来和我一起交流。我现在问一下这里学本科社会学的同学有多少？不是学社会学的本科生有多少？好，谢谢！我会提一些问题来刺激你们的思想，调节现场气氛。

第一讲我讲的是社会科学和自然科

学的关系。其实，这个问题是很大的，任何一个搞社会科学的人都会接触到这个问题，就是社会科学与自然科学的关系是什么？你们当中有不少是学社会学的，很多老师也学过社会学方面的课，了解很多社会学的知识，所以我想你们对社会学有一定的认识，从一些课程中学到了很多社会学的知识，比如中国的现状，中国的人口，中国的家庭，中国的文化，中国的民族等等。但是我想提出这样一个命题，我觉得你可能学了很多东西，对社会有很多了解，有很多关于社会的知识，但是你可能还没有学到社会学的核心。社会学的核心是什么

呢？我认为社会学的核心并不是知识，并不是关于社会的知识，而是关于得到知识的手段和方法。并不是说你从一些学者的言论中得到多少知识，而是你对他们的言论的一些批判的见解，这是更重要的。也并不是说你一定要知道什么理论，对于我而言，知道什么并不重要，而是要知道你应该知道什么，什么是对的，什么是衡量知识的标准。所以就我个人看来，社会学的学习就是要培养一种批判性的思考。知识是变化的，社会是变化的，尤其是中国社会现在变化得那么快，你今天学的东西到明天可能就没有用了，但是你应该知道什么样的知

识是对的，什么是错的，就是你要对它作一个评价，他的结论是不是有道理，并不是说名人说的话就一定是正确的，不要跟着人家跑。所以我在训练学生的时候，特别强调学生的独立性思考，批判性思考，质疑他说得对不对，结论有没有道理。我为什么这样讲呢，我做过很多这方面的研究，知道的东西比较多，刚才马老师也讲了，我做过美国最重要的社会学刊物的副主编，还做过一些其它的工作。假如你仔细看一下社会学的东西，很多社会学的东西在方法上来是靠不住的，错误很多。有的人认为社会学就是技术性的东西，只是做简单描述；

也有的人认为社会学只是对政策表示赞同意见。其实，社会学是一门严格的科学，要做好并不容易。很多人在方法论上犯错误，是因为他的研究在方法论上站不住脚。一个好的学者，不但要发表自己的言论和观点，还要知道自己的言论和观点的缺陷和不足在什么地方。比如说现在他有一个假设，但是他拥有的证据不足，并不能说明他的假设是成立的，就应该知道自己的论点的缺陷性，尤其是在方法论上的缺陷性，假如不能认识到这点缺陷性，你的研究就可能出问题。我现在讲一个很简单的例子，这是一个在假设上犯错误的例子。我有一

个研究生，他提出了这样一个假设，说一个人的数学成绩取决于他的语言能力，语言能力强，数学成绩就高。这就是一个假设出来的理论模型。他有了这个理论模型之后，就想做一个验证，他就制定了这样一个研究方案。他去了两个班，一个是数学 101 班，这个班是一个基础班，有 100 个学生。另外一个班是数学 501 班，是给研究生上的高级班，这是一个小班，只有 10 个学生。这个研究生到了这两个班上收集了这样一个数据，他问这两个班上的同学，你认为你的语音能力怎么样，数学成绩怎么样，强还是不强，强就用 1 表示，不强就用

0 表示。在 101 班上，100 个学生有 90 个学生说他的语言能力强，有 94 个学生说他的数学能力强。在 501 班上，10 个学生有 8 个说他语言能力强，有 9 个学生说他的数学能力强。这个研究生收集好数据之后，第二天就很高兴地跑到我办公室来，说他收集的数据证明了他的假设，因为一个班上的同学语音能力强，数学能力也将，另一个班上，学生的语言能力低，数学能力也低，这就证明了他的理论假设，让我给他答辩。假设你是我，你会给他提出什么样的批评？这样的数据存在什么毛病？我想听听你们的意见。

学生一：1 和 0 是定类变量，不足以客观描述出具体的强弱状态；另外，他是通过一种主观测评来获得数据，是被测者自己说的，因此数据的客观性不强。

学生二：我认为他在抽样上存在问题，他得出的数据不具有统计上的意义。

学生三：两种能力的相关可能是虚假的。

学生四：他是通过集体的语言能力和数学能力来判断个人在这两种能力上的相关性。

学生五：样本太小，调查的数量不够。

学生六：他调查的学生是两个不同层次的，他们之间不具有可比性。

好，很不错。这个数据存在很多问题，一个是样本规模太小，其中的差异在统计学上未必是显著的。

二是这两种能力并不存在因果关系，这种相关是一种伪相关。

第三个是天花板效应，这是一个测量的问题，在做好与差两种选择时，大家都感觉很好，这是美国社会的一个普遍现象，这就显示不出其中的差异性。

还一个是选择性偏差，他收集数据时是以班级为单位，而他选择的班级存在级别上的差异，他们之间不具有可比

性。

第五个是测度问题，他的数据是由被测者自己报的，学的简单的就感觉好，自我感觉好的就报好，这是由主观因素主导的，缺乏客观性。

第六个是他的验证存在一个生态学谬误，他的理论假设是以个人为单位的，是说个人的语言能力高会导致数学成绩好，这是对于个人而言的，而他做的数据和得出的结论中是以班级为单位的，但是在班级中有些人可能数学成绩好，但是语言能力差，有些人数学成绩差，但是语言能力强，他们的关系可能是正的，可能是负的，也可能不存在关系。

这就是说，当你的数据层次和你的理论层次不吻合的时候，你的数据可能说明不了你的理论，甚至有可能和你的理论是相反的。这就是为什么以团体为单位来研究问题经常会出错误，因为这反映不了个人层次上的现象。

这一点我在以后的讲座会详细讲到。

下面，我开始讲今天的主题：社会科学与自然科学的本质区别。我想从三个方面来讲社会科学与自然科学的差异。第一个是讲社会科学与自然科学在本体论上的差异；第二个是讲社会科学与自然科学在认识论上的差异；第三个

是讲社会科学与自然科学在方法论上的差异。我讲前面两个差异的目的是为了讲第三个差异，因为我侧重的是对方法论的研究。我只有先把社会科学与自然科学在本体论和认识论上的差异讲清楚，才能讲他们在方法论上的差异。当然有很多人并不同意我的观点，我也讲过这是我自己总结的观点，大家可以各持己见。有的人认为社会科学就是自然科学的延伸，社会科学就要按照自然科学的模式去做，这是我不赞同的。我认为社会科学与自然科学是有根本的区别的，这一点在后面我会讲到。

那我们先从本体论上来讲社会科学

与自然科学的差异。大家都知道什么是本体论，本体论是研究世界本质的一个哲学分支。就是什么是世界的本质，什么是世界最重要的东西。换句话说，哲学也好，科学也好，我们到底要研究什么，什么是重要的，什么是世界最重要的部分。刚才马老师也介绍了，我本科是学工科的，研究生最开始是学科学史的，当时一个很热的话题就是为什么古代中国没有科学，科学为什么起源于西方而不是起源于东方？中国的技术很发达，但为什么没有科学？很多人都考虑过这个问题，提出过这方面的见解。我认为这个问题提得并不是很好，为什么

呢？因为科学是偶然产生的，科学的产业受到很多因素的影响。技术是人人都需要的，要生存，要发展都需要技术，科学和技术在本质上是不一样的，有技术不一定有科学。科学很重要的一个来源就是柏拉图对真正知识的定义。柏拉图对自然哲学提出了很高的要求，他提出真正的知识并不是对你看得见摸得着的现实社会的认识，而是对抽象的，超现实的理念世界的了解，这样的知识才是真正的知识。因此，他把世界分成两种，一个是“永恒世界”(the world of being) 和“变化世界”(the world of becoming)。柏拉图讲的这两个世界是有

很大区别的，我们一般人接触的都是 the world of becoming，是现实的东西，是具体的东西。而柏拉图说，你要做一个哲学家，就不应该研究这些具体的东西，抛开你接触到的东西，而提出一种永恒的，放之四海而皆准的真理性的东西。举个例子，我们日常生活中会接触到很多圆，但是柏拉图说，如果你要研究圆，你就不应该研究你说接触到的圆，因为你接触到的只是圆的一种具体形态，并不是真正的圆，不是完美的圆。完美的圆不存在于现实，而存在于哲学家的思想中。你要研究圆就必须抛弃具体的圆，而想象一个圆，这个圆是没有形状，不

占空间的，既画不出，也得不到。科学就是要了解这个圆，你知道了这个圆的特征，你就可以把它推广到任何具体的圆，这就是科学的来源。它不是要你去寻找一种具体的东西，而是要找到一种永恒的，到处适用的真理性的东西。柏拉图讲的 the world of becoming 是什么呢，就是现实的东西，是我们人为的仿造品。所以，如果你要得到知识，就要透过现象看本质，要看到那个抽象的圆。柏拉图的理论是科学产生的一个很重要的基础。追求真理是科学家和科研活动者一个永恒的动力。而这些东西在中国古代都是没有的，中国古代的东西都是

很实用的，都是为了生存而设计的，并没有这种对真理性认识的追求，这是中国古代没有科学的一个很重要的原因。所以，我认为自然科学是以发掘“永恒世界”(the world of being) 中的真理为最终目的，也是其精华所在。而社会科学是以了解“变化世界”(the world of becoming) 为最终目的。历史上很多人想在社会科学领域找到一种真理，能够适用于各个方面，并且做过这方面尝试。我认为社会科学不应该是这样，在社会科学中，我们的目的是要去了解现实社会，而不是要去挖掘永恒的真理，这可能和你们想象的不一样。

社会科学与自然科学本体论上的区别也导致了它们在认识论上的差异。认识论是什么，认识论是探索人类知识的起源、本质、方法及局限的一个哲学分支。你到底能知道什么，你怎样认识世界，这是认识论的东西。自然科学的关注点是什么，它追求的是永恒的真理性的东西，因此他的关注点在于一些典型的现象，你知道了一个圆就可以把它应用到很多地方，你知道一个分子也可以应用到很多地方，甚至知道一个疾病也可以应用到很多地方。所以，你需要的是一个典型的例子，通过这个典型的例子可以使你了解一个类别的东西。因此

你的重点并不在于了解具体的现象，而在于了解典型的现象。而社会科学则不一样，社会科学的关注点在于所有个案组成的总体的状况。你了解一个典型的例子并没有很大的用处，你最好是能够了解所有的个体，对所有个体的了解就是对总体的了解。你最终要了解总体，但你是通过了解每一个个体来了解总体。这是一个辩证的关系，我会讲到为什么会是这样，怎样才能做到这一点。

社会科学与自然科学在方法论上也有差异。自然科学在可能的情况下，都希望用实验的方法来证明他的结论。虽然有些情况也是不可能的，但是至少他

们是希望用实验的方法来隔离外来因素的影响。比如，你想知道温度的变化，在实验时你就可以对压力等其他因素进行控制。通过实验，隔离外来因素的影响，就可以得到一个很纯的规律。而社会科学不一样，我们社会科学家只能运用一定社会环境下的数据（被称为观察数据），而观察数据必然受到外来因素的影响。我们也可以做一些实验，但是很多情况下是无法实现的，它有很大的局限性。社会科学之所以复杂，是应为我们的数据是通过观察所得，而观察所得的数据必然受到外来因素的影响，这些外来因素都可能能够解释你的数据。你

可以用统计的方法排除一些外来因素，但你不可能排除所有因素。我下一讲讲因果关系就是讲为什么这些外来因素会给你一种假的现象，致使因果关系不一定成立。所以说我们的数据都有问题，如果一个人说，我的数据非常好，那么他在说假话，要么他根本就没有去收集数据。收集数据和分析数据的人都知道数据来之不易，但也不能回答所有的问题，你应该知道这个局限性，因为他来自于现实生活，受到其它因素的干扰太多。学过统计的人就知道，统计上两个东西很重要，一个是平均值，一个是差异。我在这里把柏拉图的两个观念“永恒

世界”和“变化世界”用统计学的东西来进行对照。这是科学史上发展演变过来的东西。关于平均值，我举一个例子，对于一根一米长的棍子，在测量时就会碰到问题，你让几个人去测量，如果数字很精确的话，有几个人测量就会得到几种结果。你需要得到的长度只有一个，而你得到的数据却有一串，这时候你怎么办？取平均值，平均值在统计上的理论基础是大数定律和中心极限定理。大数定律就是讲，测量的平均值会随着测量次数的增加而越来越可靠和稳定。它可能永远不能达到真正的值，但它存在这种倾向。所以你要得到一个比较准确

的平均值，你就要增大测量次数。这是平均值在测量上的一个很好的依据。另外一个中心极限定理，随着样本数量的增加，其平均值越来越趋向于成正态分布。大数定理和中心极限定理这两个定理对于平均值而言具有十分重要的意义，他给我们的测量提供了帮助。但这两个定理也有条件，就是差异的来源必须是独立的、非系统的微小因素，如果没有这个条件，这两个定理并不适用。比如是由你的疲劳，气候等微小因素造成的差异就可以通过取平均值的方法得到基本消除。这两个定理对社会科学很重要，有一位数学家、天文学家叫

Quetelet，他认为社会科学很简单。柏拉图认为科学是要找到永恒的真理性的东西，但是柏拉图讲的科学不包括社会科学，因为社会科学研究的是现实的东西，而现实的东西是不断变化，因此柏拉图的科学中没有社会科学，只有自然科学和哲学。但是 Quetelet 说社会科学中也可以得到永恒的，一成不变的东西。怎么办呢？我们就取平均值。当样本足够大的时候，我们就能很准确的通过平均值来预测一些现象。因此，他提出了“社会物理学”，要获得永恒性的东西就取平均值，这就得到他提出的“平均人”概念。“社会物理学”是研究“平均人”的，“平均

人”是永恒不变的，是能够符合柏拉图对科学的定义的。这是他的一段话，“我们人类多悲哀啊，我们每年可以知道将有多少人的双手会沾满他们同胞的鲜血，多少人编造谎言，多少人死于毒药……”这些都是可以通过“平均人”来预计的。因此他就用统计的方法，用取平均值的方法来研究社会学，他的社会学就是取平均值的社会学。自然科学是研究典型现象，而社会科学的“平均人”也是一个典型，但这个典型不是一种个别现象，而是一个平均数。这就是 Quetelet 的观点，他这种观点是用柏拉图的传统意义上的科学对社会科学研究的一个尝试，

就是通过统计的方法得到一个不变的东西，这个东西是抽象的，不是具体的。但这种思想在 19 世纪受到很大的冲击。我下面要讲的是达尔文的人口思维对我们社会科学的影响。我认为社会科学起源于达尔文，虽然我们今天记住达尔文主要是因为他的自然选择学说，但是他对社会科学的影响是非常大的，你想一下，他的自然选择学说是哪里来的，自然选择的基础是差异，假如每一个孩子和他的父亲或母亲是一模一样的，这个社会就不会有变化，更不会有进化。进化的前提是差异。同样的原因会出现不一样的结果，相同的父母会生出不同的

孩子，同样的家庭环境会给孩子造就不同的性格。这里就有一个随机的因素，这是一种自然想象。而这种差异对于达尔文来讲就是世界本质，而不是人为的差错。既然差异是世界的本质，那差异就应该是研究的对象，我们研究的对象不应该只是一个简单的平均数，虽然这对我们也很重要，但我们不能仅仅局限于此，我们不能光讲平均，而更要讲差异，比如经济上北京和上海等大城市很发达，但内地怎么样呢，这里就有一个地域差异，因此差异是很重要的。对于达尔文而言，差异并不是一种人为造成的错误，而是一种现实的必然，是精华

所在，是值得研究的东西，而不是说你知道了-一个平均值了，差异就可以扔掉。这是很大的一个思想上的转折。因此，今天我讲的最重要的东西就是差异的重要性。我把达尔文的这种思想叫做“人口思维”(Population Thinking)，而前面讲的柏拉图的思想就叫“典型思维”(Typological Thinking)，我在后面会详细对比这两种思维。“人口思维”(Population Thinking)是对科学的传统定义的一个挑战。传统科学是要求你得到一个典型的规律，而达尔文的人口思维是要你得到一种分布，得到所有由个体组成的一个整体。达尔文是一个生物

学家，他并没有做什么社会科学的研究，那是谁把这些东西介绍到社会学呢？是他的一个表弟，叫 Galton，是他将人口思维引入社会科学。Galton 去的地方多，见的世面广，他认为平均值并没有多大意义。他觉得有的人聪明有的人笨，平均值根本没有意义。美国有一个笑话，有人问一个统计学家把一支手放进冰箱，一支手放进火炉里感觉会怎样，他说从统计上来讲两者平均一下刚刚好，这就是平均值的毛病。平均值在很多情况下并不能反映真实的现象。Galton 还认为 Quetelet 的社会物理学的用处不大，“平均人”并不是万能的。我想现在不论

是中国社会学还是美国社会学都还存在这种传统，就是对数据做平均值，不管什么都平均一下，这只能反映一个很小的方面，还有很多方面被忽略了。Galton说科学的探索必须关注差异与协同差异。差异有个体之间的差异，比如有的人聪明，有的人笨，有的人勤奋，有的人懒散，而且不仅仅是一个变量存在差异，多个变量之间也有差异，我这里把它翻译成协同差异，就是什么东西和什么东西一起变化。他说这才是重要的东西。这里我用图表来解释 Galton 的一个理论，这个图可以用来解释很多东西，比如教育和收入，经济发展程度和消费

水平的相关性。Galton 首先发现这种相关性是通过父亲的身高与儿子的身高的关系发现的。他把一些父亲与儿子的身高数据画成图表后发现这些点在一个椭圆之内，但这个椭圆不会变成直线，因为直线就意味着完全相关。这里的相关性意义就是父亲的特征会影响到儿子的特征，比如父亲聪明，儿子也偏向于聪明，父亲高，儿子也偏向于高，这和达尔文的自然选择也是相对应的。除了相关还有一个回归，回归就是说这种相关关系不是决定性的，不是说父亲高，儿子就一定高。回归就是一种倒退，就是没有原来那么好，比如你父亲很高，但

你可能没有你父亲那么高，但这种倒退不会倒退到平均一下，如果倒退到平均值一下，就意味着没有相关关系了。因此相关关系并不是决定性的，偶然性起着非常重要的作用，回归是一种往均值方向的倒退。社会变迁有随机的因素，再好的人家也会出败类，再不好的人家也会出凤凰。所以，相关与回归似乎是矛盾的，因为相关是讲前者对后者的影响，可能是一种因果关系，而回归是把这种关系往回拉，就是弱化这种关系。相关和回归是一个问题的两个方面，相关是存在的，但这种关系不是决定性的，这就要通过回归来把这种关系拉回来。

相关和回归的关键在于差异，没有差异，就不会有相关，也不会有回归。相关和回归是 Galton 发现并命名的。Galton 是一个很优秀的科学家，但由于他叙述不清，致使相关系数的发明权落入他人手中。Galton 发起过一个优生运动，因为通过统计，他发现有的人聪明有的人笨，他认为通过科学的方法可以促进优生，以改进社会的质量。下面，同学们有问题的话，可以先提几个问题。

问题一：“平均人”的概念是不是一个“永恒世界”的概念？

谢教授：对，Quetelet 是用统计的方法来满足柏拉图的“永恒世界”的要求。

他并不研究具体的东西，他认为通过“平均人”的办法可以找到永恒不变的真理。

问题二：谢教授，您刚才说差异是世界的本质，您讲社会科学与自然科学存在差异，而差异是本质的东西，那么是不是说共性的东西就不是本质，就不需要研究呢？

谢教授：我讲社会科学和自然科学的差异并不是要给某个具体的研究定性，也不是说只要研究差异，不需要研究共性，而是因为以前我们讲差异讲的太少，对差异不够重视，所以我在这里突出的讲一下差异。当然平均也很重要，但在平均的基础上要知道差异，要知道

整个分布的情况，而不是说了解了平均就足够了，就能够了解社会。

问题三：您讲的父子特征的关系到底是一种统计上的相关性，还是一种遗传的结果，这是自然科学的东西还是社会科学的東西？

谢教授：当然，这种相关性可以通过基因遗传来解释，但这种遗传也是有偶然性的，受一定的随机因素的影响的。这一点，统计学与自然科学中的生物学有一定的相似性。统计学有三个奠基人，一个时搞人类学的加尔顿，一个是搞农业的皮尔逊，另外一个埃·菲舍，他把这一套统计的东西理论化，提出了一套

系统的统计理论。后两个学者都是以研究农业为主，但是我们也用他们的研究方法，因为他们的方法对于我们的研究也很适用。

问题四：谢教授，您讲社会学的核心是一种批判性的思考，但这种批判性思考是其他学科也需要的，因此我想问您一下，您讲的社会科学与社会学是一种什么关系，是包容，还是等同？

谢教授：可能是我开始讲得概括性强了一些，我讲的核心并不仅仅是批判性的思考，还包括人口思维，包括后面我会讲的事实性。像心理学，它比较重视个案的研究，而比较少的考虑社会规

律性的东西，我们社会学考察的是社会事实，注重对社会现象的解释和社会规律的探讨。政治学讲政治体系，经济学讲经济实体，这些跟我们社会学都是有区别的。

问题五：我们求相关系数是为了发现一种规律，而规律是一种共性，而您却强调差异性，我想问一下您的解释。

谢教授：我理解你的问题，相关系数确实反映一种共性，但这种相关是会变化的，因为相关是受到偶然因素的影响，产生这种偶然因素的原因就是差异，有差异才会有偶然，有偶然才会有变化，只有在变化中才能寻找相关。好，我们

接着讲下去，到后面有时间我再回答一些问题。下面我讲一个反例，这是一个数据“太好”的例子。社会科学用的是观察数据，而观察数据是有规律的，如果有人伪造数据的话，我们就可以分辨出来，因为伪造的数据一般可信度比较低。下面是一个真实的例子，讲的是一个英国著名科学家 Sir Burt，Sir 是爵士，这也反映了西方文化和东方文化对科学的不同看法。在中国，书读得好，学问做得好的人要给官，而西方是给荣誉，这个 Sir 就完全是一个荣誉，非常优秀的科学家都会得到这样一个称号。Burt 是一个心理学家，其主要贡献在于对智商代

际遗传的研究。为了得到更科学的数据，他增加了样本数量，但是得到的相关系数却是不变的，不变到第四位小数都是相同的。按统计学规律，随着样本的增加，相关系数也会随之发生适当变化，而他的数据太“完美”了，真像柏拉图讲的那样永恒不变，这就好得不可信。但是由于当时他是非常有名的科学家，因此没有人提出质疑，直到本世纪 80 年代初，才有人提出质疑。我们现在知道他的数据是伪造的，这是科学界一个很大的伪造事件，他发表的文章中的合作者都是虚构的，他做的访问也是捏造的，得出的数据都是自己编的，因此我们说

任何学科都要有批判性，没有一种批判性的反思，这门学科就很难发展。我这里总结一下“典型思维”(Typological Thinking)和“人口思维”(Population Thinking)的区别。“典型思维”(Typological Thinking)来源于柏拉图的“永恒世界”(the world of being)。而这种典型思维在社会科学中的运用，就是Quetelet的社会物理学，用统计的方法来满足柏拉图对科学的定义。再具体而言就是用实验的方法，把外来因素排除，通过控制变量来达到目的。在这方面，心理学做得比较多，经济学也有，最近经济学出来一个分支叫实验经济学，但

是实验的方法还是有很多缺陷的，我在最后一讲我会讲到实验方法的缺陷性。

另外，“人口思维”(Population Thinking)有什么特点呢？它来源于达尔文的进化论。这种人口思维并不是社会学独有的，在很多其它学科也用到这种思维，但在社会学里，这种思维特别重要，我下面会讲到为什么他对社会学特别重要，我有自己的一些看法。把这种“人口思维”(Population Thinking)应用于社会学的主要是 Galton，他提出了差异和协同差异。社会科学之间虽然有差异，但是他们都会用到达尔文和 Galton 的人口思维，并且他们都要应用到统计的方

法，我们一方面要把事实描述得准确、全面，另外我们又要重视整体下面的每一个个体，把整个过程、事实、变量描述得准确是我们的责任。但是，我们在做这件事情的过程中又会遇到一个问题，就是我们的思考能力，分析能力是有限的，当样本数量增加时，为了使我们对事实的描述更准确，我们就要运用到统计方法和统计工具。真正的科学是简单的，相同的理论，越简单的越可信。很多有名的科学家都信教，最典型的例子是开普勒和牛顿，他们都相信上帝。上帝一定是一位很聪明的数学家。因此越简单的东西越可信，哥白尼的“日心

说”之所以能够说服后人，是因为他的解释比托勒密的“地心说”简单很多，省去了很多繁琐的公式。

我是搞统计的，而统计是要求实证的，因此很多不了解我的同事就认为我是实证主义者。实际上，我不是实证主义者。实证主义是一种哲学观点，它认为自然科学与社会科学没有本质的区别，换言之，社会科学可以做得和自然科学一样严格，而且社会科学应该照着自然科学去做。这里就有两种提法，一种是说社会科学还不够完善，还不够科学化，需要通过采用实证的方法把它提高，使他向自然科学靠近；另外一种是

要把自然科学降低，这种说法讲自然科学也受到社会的影响，受到阶级的影响，在这方面和社会科学一样，因此可以把自然科学降低，和社会科学靠拢。这两种提法在本质上是一样的，就是社会科学和自然科学没有本质的差异。但我是不同意这种观点的，所以认为我是实证主义者是很不公平。实证主义受到很多方面的挑战，包括马克思主义、后现代主义、后后代主义等方面的挑战和冲击。我虽然不属于实证主义者，但是我承认我的思想属于一个流派，这个流派我把它翻译成人口学“流派” (Demographic Approach)，这个流派在美国社会学界有

很大的比重和影响力，他是以 Duncan 为代表的。“人口学”是一个包容性很强的综合性学科，它和社会学，人类学，经济学，心理学都是有关系的。中国的人口学太闭塞，还没有发展为一个综合性的学科。Duncan 对“人口学”的一个定义是人口学是对人口数量、人口构成及其变化的研究。这个定义是很广的，人口数量就包括出生、死亡和迁移，而人口结构就包括家庭、学校、教育、职业等等。所以 Duncan 对人口学的定义下得很广，很多人问我是人口学的还是社会学的，我说我既是人口学的，也是社会学的，我是用人口学的定义研究人口学和

社会学的。我还给人口学流派一个名字，叫做实用经验主义（ pragmatic empiricism ），这里的经验是以你从现实世界收集的资料为基础的，而不是你的主观经验。“实用”是说你收集的资料可能对你的想法很有说明力，但是它并不是十全十美的，只是它是可用的，对你的研究有可用性。我要举的例子是 Lieberman ，他是一位很优秀的社会科学家，他讲到，“社会学的重要贡献之一，就在于它提供信息的能力。这些信息包容了社会普遍关心的内容，有关种族平等、强奸、贫困、无家可归者、代际流动的统计。我选择这些问题是因为他们

的答案提供了有关社会的有用信息，即使它们有时候并不一定为某种理论服务（顺便说一句，谁说多了解一些事实是一件坏事呢？）”。社会学的重要贡献是它能够提供信息，在这方面的能力上，社会学比其它学科要强，比如说人类学、经济学、心理学、政治学。因此，在提供信息方面，我们的责任更多。在美国几乎所有的对社会信息的描述都是社会学家提供的。对于家庭，家庭的破裂，小孩的生育，就业等比较概括性的，描述性的东西，都是我们提供的。因为其它的学科往往重视了一方面而忽视了全面的東西。美国的趋势是怎么样，家庭

是怎么样，父母离婚对小孩的影响怎么样，贫富差别怎么样，上代人对下代人的职业影响怎么样，犯罪率是增还是减，离婚率是增还是减，这些东西都是由社会学家来做，所以我们的学科和其它学科是有很大差别的。因此社会学家的使命是很重要的，有时我们讲社会学家就是历史学家，只不过我们描述的历史是发现在现在，而不是发生在过去。我们是把现在发生的过程和现象系统的表现出来。Lieberson 的这段话其实是对纯理论学家提出批评，他批评理论学家常常忽略了经验性的结果，他们是空谈理论而不重视实际，理论和研究是没有根本

区别的，理论是知识的全部内容，理论应该超越事实，但它必须包含已知的信息。一个忽略了已知事实的理论，只能是一个片面的、错误的理论。他认为爱因斯坦的理论是好的理论，他的理论是尊重过去的经验事实的，而我们一些社会学家却不去了解社会事实，而只是空谈理论。他批判了三个例子，一个是吉登斯，他写了一本关于美国社会分层的书，书里对美国分层的描述根本不符合美国社会的真实情况。第二个是研究文化社会学的哈德，他认为社会参与性在降低，而他这个结论是建立在猜想的基础上。另外一个就是布尔迪厄，他提出

一个品味阶级差异理论，Lieberson 说他的理论缺乏论证的严密性。因此，我讲的人口思维就是理论要建立在事实的基础上。

社会学有两种，一种是先写方法，再写结果，再写结论；还有一种是先写结论，再写方法，再写开头。我是第一种人，我习惯先摆出我使用的方法，再描述结果，然后总结出我得到的结论。我就讲这些，还有十分钟，由大家来提问题。

问题一：我想请您谈一谈社会学的科学性和它的发展未来。

谢教授：我在第六讲就会讲到这个

问题，社会学的科学性和社会学的未来，特别是美国社会学，我觉得还是有令人担忧的地方。这个问题有几个方面，第一是什么是科学性，我认为科学性就是要以事实为基础，这个问题我会在十一月一号详细的讲。

问题二：您讲到自然科学是以发掘“永恒世界”(the world of being) 中的真理为最终目的，而社会科学是以了解“变化世界”(the world of becoming) 为最终目的，这是您个人的归纳和演绎推理的结果还是有现实证据的支持？

谢教授：这是我个人的判断，所以我说不一定所有人都同意我的观点，有

的社会学家就追求一些与自然科学一样的规律性的东西，这只是我个人的观点，也是我要强调的，是我认为从达尔文以来的社会科学的一个特定的东西，并不要求其他人都接受我的看法。

问题三：您在讲社会科学和自然科学的差异的时候讲到自然科学重视典型现象，而社会科学要研究全部个体，但是在韦伯的“理想模型”中提出一种理论分析方法，就是对经验现象进行理论概括和抽象，从经验事实中抽象出主要特征，以之与经验中各具体类型作比较。这也是一种典型性分析，您怎么看待这个问题呢？另外，社会学和自然科学

都存在一种相互借鉴，相互使用对方的一些适用的研究方法和研究工具的现象，您怎么解释这种现象呢？

谢教授：社会科学和自然科学确实 在方法和思想是有相互的借鉴，这是对的。现在科学越来越走向跨学科的趋势，自然科学和社会科学之间的界限也不是很严格的，但是他们之间还是有差别，这种差别是一个侧重点的问题，是侧重点的不同，而不是说他们之间在方法和思想上完全没有相似性。关于典型性，韦伯讲自然科学的定律越普遍越好，而社会科学的定律越普遍，它的实用性越弱，社会学理论越具体越深刻，越宽泛

越肤浅。人与人之间虽然有差异，但是他总有一个主要的倾向，根据这个趋向的不同，我们可以把人分成不同的组，先比较组与组之间的差异，然后再来比较同组的人之间的差异，这样就可以解释很多现象。比如在美国，黑人与白人存在种族差异，另外还有性别差异，还有年龄差异，受教育水平的差异，我们就可以通过分组，通过组与组之间的比较来解释这种差异。

问题四：您开头讲过古代中国没有科学，您能不能讲一下为什么。

谢教授：这个问题可以从很多学科角度去看，有很多国家有很好的技术，

比如阿拉伯国家，印度、中国，但是他们都没有形成科学。我认为这是因为他们的需求起点不一样。在中国，很多知识是很注重实用的，功用性心理很强，这就缺少了对追求科学的思想起点，科学是为科学而科学，而不是为实用而科学。因此，中国没有能够形成科学是由中国传统的文化背景，思想观念决定的。

问题五：谢教授，我想问一下，您认为社会学研究要强调差异，是不是意味着可以忽视对共性，对规律的研究，或是这种研究不重要？

谢教授：我可能过多地强调了差异，而对共性讲的不够，但是我认为共性也

是重要的，但是仅仅知道了共性是不够的，你还要知道他们的差异在什么地方。相对来讲，心理学和经济学对共性强调的比较多，社会学也需要研究共性，但是只知道共性也是不够的，共性是建立在个性的基础上，所以我们要研究共性，也要强调对个性，对差异的研究。因为一般的人对共性研究的比较多，而讲差异讲的比较少，所以今天我更侧重讲差异。

马戎主任：由于时间关系，我们现在提最后一个问题。

问题六：从本体论上来讲，社会是分为很多层次的，对社会的分析也应该

是从多个层次来分析，您直接把两种不同的思维合并在一个层次上，不是有不妥；另外，实验是一种观测的方法，统计是一种分析方法，把这两者混在一起是不是有些不恰当？

谢教授：我并不是说社会科学就不能有典型思维，也可以有，但是存在一个侧重点的差异。另外，实验是一种方法，它可以排除外因干扰，而我们得到的数据是观察的数据，这种观察是以事实为基础的，而现实社会中的事实是受到外部因素干扰的，我们不能够强制性的排除外部因素，而通过实验就可以对变量进行控制，但是社会学主要是通过

观察来得到数据，而不能通过强加条件来排除外来因素。好，谢谢大家！

马戎主任：今天晚上的讲座就到这里，谢谢大家！

发表时间：2004-12-9 12:53:00

主题：因果推理

Causal Inference

主讲：谢宇教授（美国密西根大学社会学系教授）

主持：马戎教授（北京大学社会学系主任，博士生导师）

时间：2002年10月23日晚7:00
- 9:00

地点：五四体育馆体教电教厅

马戎教授：今天这个讲座是谢宇教

授方法论系列讲座的第二讲，主题是因果推理。下面我们用掌声有请谢宇教授给我们做报告。

谢宇教授：谢谢大家光临来听我的讲座，我今天要讲的是方法论的第二讲因果推理。上一讲我们讲的是比较抽象的哲学比较。这一讲我会讲得更具体，更深入一些。要讲因果推理，我们首先引入几个因果关系的问题（Causal Question），第一个是9.11事件会使美国人感到恐怖吗？如果会的话，9.11事件就是因，美国人感到恐怖就是果，是9.11事件导致了美国人的恐怖感。第二个例子是你们来参加我的这次方法论系列讲

座有收获吗？如果有收获的话，你们参加讲座就是原因，收获知识就是结果，如果你们不来参加这个讲座可能就没有这种收获，但可能会有别的收获，是你参加这次讲座造成你在方法论方面有了这种收获。因此，因果关系问题是一个形式上简单的问题，它包含两个理论性概念之间的关系：原因（cause）和结果（effect）。是不是有原因就会导致结果？如果把原因变量定义为 X ，把结果变量定义为 Y ，是不是有 $X \Rightarrow Y$ ？这就要求我们正确的理解因果关系。因果关系是所有科学研究的基本目标。通过研究因果关系，我们可以准确的预测未来，

为政策干预提供科学根据，还可以验证和丰富认识客观世界的理论知识。

为了理解因果关系，我们先引入简单比较的概念。简单比较是一种简单的方法，就是比较试验组和控制组。下面我举一个例子来说明这种简单比较的方法。在一个社区内， X_1 个儿童参加了一个启蒙教育项目 (Head Start)，这些儿童就是试验组，而另外 X_2 个儿童没有参加这个项目，这些儿童就是控制组。27 年后再来测量这两组人各自得到的受教育水平。我们用 Y_1 来表示参加了启蒙教育项目的试验组儿童的受教育水平；用 Y_2 来表示没有参加启蒙教育项目的控制组

儿童的受教育水平。结果是受过启蒙教育的儿童的受教育水平比没有受过启蒙教育的儿童的受教育水平要低。那我们是否可以得出这样的结论，启蒙教育项目对受教育水平有负向作用。60年代前期，Westinghouse 就报告了这个结果。从观察到的数据来看，参加启蒙教育项目与受教育水平的关系好像是负向的。但是，我们这里忽视了一个变量的作用，就是社会经济地位。参加启蒙教育的儿童一般是来自贫困家庭，需要通过政府资助才能上学，而那些没有接受启蒙教育的儿童来自比较富裕的家庭，他们父母能够供他们上学，并且能提供较好的

学习环境。因此这两组人根本就不具有可比性，因此在这个试验中把家庭条件较好的儿童定为控制组就是不合理的，因为在家庭经济地位的影响下，我们根本就看不出参加启蒙教育与不参加启蒙教育对贫困儿童受教育水平的影响。因此在这个试验中不应该把家庭经济地位较好的儿童设为控制组，而应该选择两组都是来自贫困家庭的儿童，一组作为试验组，另一组作为控制组，再来测评启蒙教育对他们受教育水平的影响，这才有意义。另外一个例子是，加州大学伯克利分校研究生录取的性别比例问题。我们先来看一下总的录取数据。

	申请人数	入学比例
男性	8442	44.2%
女性	4321	34.58%

表面上看，好像男性的录取率要比女性要高，但实际上是不是这样呢？我们来看一下具体某个专业的录取率，从专业的录取率来看，女性集中报考在录取率较低的热门专业，并且在这些专业中女性的录取率还比男性要高，在其他相对冷门的专业里，女性的录取率都不必男性低。但是，为什么总的录取率中，女性会比男性低很多，致使当时掀起了一股女权运动。主要是因为报考的女性

基数比男性要小很多，只有男性的一般左右，当大部分女性集中报考那些录取率比较低的热门专业时，其总录取率必然较低，但男性虽然也有不少人报考了录取率较低的热门专业，但是他们还有很多人（相当于女性的总人数）报考了其他录取率较高的专业，因此平均下来，男性的总录取率会高于女性。因此这里根本不存在性别歧视的问题，只是因为一些人只看到了表面数据，而没有深入分析。因此在这个案例中，性别与专业选择有关系，性别影响专业选择，专业与录取率又有关系，不同的专业有不同的录取率，由于录取率在各专业之间有

很大的差异，女性多的专业录取率低，男性多的专业录取率较高。这就导致了总录取率上男性要高于女性。

下面，我们来看一下其它三个社会学案例。第一个是从受教育水平来看，是兄弟姐妹多好还是独生子女好？是独生子女受教育水平高还是有兄弟姐妹的孩子受教育水平高？我想听一下大家的意见。

学生一：我认为这个问题要从两个方面来看，一个是当经济发展水平较高，家庭收入相对较高，受教育机会均等时，多子女之间由于有相互协作，他们的受教育水平就会比独生子女要高。当经济

发展水平较低，家庭普遍比较贫困时，独生子女的受教育水平就会相对比较高，因为这个时候受教育的机会不均等，子女多的家庭难以支付所有子女上学。

谢教授：你分析地很好，但是我要补充一点，就是事实上当机会不均等时，独生子女享有更多的受教育机会，但是这并不能说明独生子女的受教育水平会比非独生子女要高。由于非独生子女受教育机会难得，他们往往会非常珍惜这种机会，因此他们有很强的学习动力，这往往时他们在学习时候比独生子女要刻苦，而独生子女因为家庭条件优越，往往没有什么学习进取心，当然我这里

是讲一般情况，不乏有特例，因此就学习成绩而言，有的非独生子女往往比独生子女成绩要好，但是由于家庭经济条件的限制，非独生子女获得本科和本科以上的受教育机会很少，因此总的而言，独生子女的受教育水平会高于非独生子女。

我们的第二个案例是婚前同居使离婚的可能性增加了还是减少了？我还是先听一下你们的看法。

学生一：我认为婚前同居会增加婚姻的稳定性，因为婚前同居会使相互之间加强了解，增进认识，形成较稳定的认同，这会使他们的婚姻更加稳定。

学生二：我认为婚前同居不利于婚姻的稳定。从观念上考虑，选择婚前同居的人一般是观念比较开放的人，他们在结婚以后如果婚姻遇到挫折的话也会很轻易的选择离婚；而那些受传统观念影响的人不经过婚前同居而结婚，他们就会很看重婚姻的稳定，而不会轻易的离婚。

谢教授：从理论上来分析，结论和第一位学生比较接近，因为同居关系的确立会增进相互之间的了解，导致婚姻更加稳定。但是统计数据表明婚前同居会使离婚的可能性增加。为什么呢？原因就跟第二个同学说的差不多。由于选

择同居和部选择同居的人在观念上存在差异，选择同居的人一般观念比较开放，他们会轻易的选择同居的话，他们在结婚以后如果相互之间出现矛盾的话，他们也会很容易选择离婚。而不通过婚前同居而直接结婚的人，他们的观念相对比较传统，他们会比较重视婚姻的维系。因此婚前同居并不一定会增加婚姻的稳定性。

案例三是接受了大学教育后有什么回报？你们都上了北大，将来都会挣很多钱，但是当你挣钱的时候，你是不是会认为你能挣这么多钱是因为你在北大受教育的结果？在考虑这个问题的时

候，我们不能通过比较在上大学的人和没有上大学的人来获得认识，因为我们不知道在上大学的人如果他现在没有上大学的话，他会是什么状况。而那些没有上大学的人，他们如果上了大学后又会有什么状况，这些都是我们无法知道的，因此我们直接把在上大学的人和没有上大学的人放在一起比较是不科学的。那么，我们能不能通过试验来获得对这个问题的认识，就是我们能不能让一个人既去上大学又不去上大学，然后再来比较上大学与不上大学的区别呢？这显然是不行的。当一个同学在上大学时，我们就不知道如果他不上大学会是

什么样，如果一个人没有上大学，我们也不会知道如果他上了大学会是什么样，这是无法通过试验来获得对这个问题的了解的。因此，归根结底，因果关系问题实际上是一个反事实问题（Counter-Factual），就是你要反过来想一想你在做某一件事情的时候，你没有做的事情。因此，在做因果推理的时候我们必须考虑反事实的问题，对于那些接受了“试验”的人，你要想如果他们没接受这种试验会是怎样的情况。比如那些受过启蒙教育的儿童，假如他们没有受到这种启蒙教育，他们会是怎样呢？就是对相同的人来讲，得到了

treatment，和没有得到 treatment，会有什么区别。对于那些没有受过试验的人，假设他们接受了试验，那么又可能是怎样的状况呢？比如独生子女的孩子没有受过非独生子女的待遇，非独生子女的孩子也没有独生子女的经历，婚前同居的人不知道婚前不同居的人的感受，婚前不同居的人也不会了解婚前同居的人的体会。因此我们在想问题的时候不仅要想组与组之间的差别，更要想同一组人在两种不同情况下的差别。因为这是一个反事实的问题，我们根本不可能通过试验得到验证，因此这里就有一个很现实的问题，就是一个缺乏数据的问题。

对于一组人，我们只能知道他在一个时间内进行一件事情的情况，而不可能知道在这个时间内如果他去做其他事情的话，他会是一种什么情况。比如对一个上大学的学生，我们不可能获得他不上大学的情况的数据。一个婚前同居的人不可能知道如果他婚前不同居会是什么情况。但是，这种数据的缺乏并不能阻止我们通过逻辑思维来对这个问题进行思考。因此这里就需要对无法验证的现象进行假设。但是我们引入假设也是有代价的，假设是否合理会直接影响到结果的正确与否，就是你必须从最牢靠，最基本的现象来看问题，必须把你的假

设建立在事实的基础上，你的结果错了就完全有可能是你假设错误，因此假设也是有代价的。但是由于我们无法获得反事实现象的数据，我们不得不通过逻辑思维进行假设，这就引入了简单比较所需的假设问题。

如果试验组对象与控制组对象大体上是相当的，那么可以用这样的假设来简化问题。我这里为什么要说大体上是相当的，因为他们之间的差异还是有的，比如性别上的差异，年龄上的差异，但是通过大样本取平均值就可以忽略不计。就是假如组与组之间只有接受 treatment 和不接受 treatment 之间的差

异，而没有其他的差异的话，我们就可以简化问题。对于第一组，我们知道了他接受了 treatment 的情况，但我们不知道如果他们没有接受 treatment 会是什么情况，我们就可以假设他们和控制组是差不多的，他们没有接受 treatment 的情况会和控制组没有接受 treatment 的情况差不多。同样，对于控制组，他们没有得到 treatment，我们不知道他们得到了 treatment 后会是什么情况，我们也可以假设控制组和实验组是大体相当的，实验组接受了 treatment 的情况就会和如果控制组也接受 treatment 的情况差不多。当这两个假设都成立时，第一组的处理

效应减去第二组的处理效应就会等于第二组的处理效应减去第一组的处理效应。一般我们做简单比较时会有代价，这种代价就是因为我们认为被处理过的和没有被处理过的组大致是相同的，在大致相同的假设下才可以用简单比较的方法。比如就我前面讲的例子，假设婚前同居的人和婚前不同居的人在观念上、性格上大体上是一样的，如果这个假设成立的话，我们就可以对同居和不同居的人进行简单比较。另外一个例子是上大学的人和不上大学的人，假如他们之间在能力，年龄，机会和把握机会的能力上是基本相同的话，我们也可以

对他们进行简单比较。但是如果假设不成立的话，你得到的结果就会有很大偏差，而当你的结果出现偏差的时候，你也会知道你是在假设上出了错误。所以在假设成立的时候，就可以进行简单比较，但是简单比较也是有代价的，当假设不成立时，你得到的结果就有可能是错误的。在现实情况下，这种假设一般是不成立的，你只能在试验的时候对外部因素进行限制和排除才能使这种假设成立，但是这在现实社会中是不可能的。比如婚前同居的人和婚前不同居的人他们在观念上，生活习惯上，个人性格上就会存在很大的差异，他们之间是

不会大体相同的。另外，上大学的人和不上大学的人在能力、个人兴趣等方面也是有很大差异的，我们也不能把他们完全等同。而这些东西又是我们不能通过试验来实现的，我们不能强制一些人婚前同居，一些人婚前不同居，我们也不能强制一些人去上大学，一些人不去上大学。所以用简单比较的方法计算回报率得到的结果会偏高。

为什么是偏高而不是偏低，这就引入了一个比较具体的问题，就是忽略变量偏差 (Omitted Variable Bias)，这种偏差是不可避免的，但是我们必须把这个偏差的方向搞清楚，就是我们要知道为

什么会有偏差，而且偏差是正的还是负的。如果试验组与控制组对象在观察到的有关特征上具有明显差异，那么这种选择性就被称为“观察到的选择性”（Observed Selectivity）。我们社会科学方法论上最头痛，最难解决的问题就是选择性的问题。人的行为是理性的，他做一件事情的时候会有他的道理，结婚也好，同居也好，这都是他们自己的选择。所以假如这里所有的选择性我们都能够观察到，那我们还能够解决，关键是有很多选择性是我们观察不到的。比如第一个启蒙教育项目的例子，这里的选择性假如是由于家庭经济地位造成

的，那么我们就应该对这些家庭的经济状况进行测量。所以选择性是一个潜在的东西，要解决选择性的问题，有一种方法是把这些所有可能的危机全部都找出来，因为这些危机是可能导致我们得出的因果关系是假的。就是假如你知道某些变量是重要的，但是你忽略了这些变量，就有可能导致忽略变量偏差。在多元分析当中，我们可以通过统计控制，使这两个组具有可比性，从而解决这个问题。实验组和控制组是有差异的，是不相同的，但是我们可以通过统计的方法来对它们进行控制，使他们达到基本相同。比如，如果我们不考虑年龄的

因素，我们会发现佛罗里达州的死亡率会大大高于其他州，为什么呢？当我们从年龄结构上去分析时就会发现，佛罗里达州的老年人人数大大多于其他州，这就直接导致它的死亡率很高。所以当我们忽视了年龄这个变量的时候就会导致忽略变量偏差。因此，你在做统计的时候，就要尽可能地考虑到所有能够产生影响的变量。这就是我们要做多元分析的基本点。在因果关系里我们想知道的就是两个东西，一个因，一个果，由于组与组之间存在差异，所以我们在针对一个结果去找原因的时候我们就要想到多个因，通过多元分析来掌握其中的

因果关系。这就是为什么统计很难做，为什么问卷要设计很长，就是因为我们要尽可能的考虑到可能对某个结果造成影响的各个变量。但是并不是所有的比较都会造成忽略变量偏差，很多时候并不会出现忽略变量偏差，因为忽略变量偏差要成立的话，它必须满足两个条件。这两个条件中缺少其中的一个都不会出现忽略变量偏差。第一个是有关条件，就是你忽略的变量要对因变量产生影响，也就是你忽略的变量要与原因有关系。第二个是相关条件，就是你忽略的变量要与主要的自变量相关，第一个是要和因有关系，第二个是要和果有关系，

只有这两个条件同时成立时才可能会出现忽略变量偏差。在伯克利分校录取率的案例中专业与性别也是有关的，不同性别的学生集中在不同的专业，另外专业与录取率是有关的，不同的专业录取率不同，忽略了专业这个变量时就会造成忽略变量偏差，假如说专业与性别没有关系，或者专业与录取率没有关系的话，就不会出现忽略变量偏差。只有当专业与性别有关系并且跟录取率也有关系时，如果你忽略了专业时就会造成忽略变量偏差。另一个就是同居与离婚率的关系问题，如果观念与同居没有关系，或者与结婚后的选择离婚没有关系，就

不会出现忽略变量偏差。正是因为人的观念既影响了婚前的同居，也影响了婚后对离婚的选择，因此忽略观念在这两者中的作用就会导致忽略变量偏差。

下面我们来看一个例子，我们这个社会对两个东西最感兴趣，一个是钱，一个是婚姻。我要举的这个例子把这两样人们最感兴趣的东西都包括了。这个案例就是在美国社会有一个很奇特的现象，就是婚礼花费的费用对婚姻的稳定性有正向作用。这一结果是否是忽略了有关变量造成的。我想问一下你们为什么婚礼的花费会对婚姻的稳定性产生正向影响？

学生一：应为婚姻花费越高，沉没成本越高，而这种沉没成本是无法回收的，因此在他们通过巨额的开支而结了婚后，如果要离婚的话，一是意味着他们投入的沉没成本没有获得收益，第二他们就会考虑下一次再结婚又要投入很大的沉没成本，因此他们更可能会选择维系现存的婚姻。

学生二：我觉得有两个原因，一个是如果他的婚礼办的比较大，就会牵动较多的关系，当他们要选择离婚时，要面对比较大的舆论压力。另外一个是他的婚礼花费高意味他的经济条件比较好，而经济条件好的家庭更容易维系婚

姻。

谢教授：你们讲的很好。在美国，婚礼的费用一般是由女方家庭来出的，你们讲的成本，舆论压力，经济条件都是影响婚姻稳定性的因素。因此婚礼本身的稳定性并不是由钱这一方面的因素造成的，他还通过很多中间变量在起作用，因此简单地说婚姻花的钱多，婚礼的稳定性就高，也是忽略变量偏差导致的结论。因此这不是一对真正的因果关系。

现在，我们来讲忽略变量偏差的可能情况，和偏差的方向。忽略变量偏差有三种情况。在第一种中，忽略的变量

用 Z 表示，原因用 X 表示，结果用 Y 表示， Z 分别是 X 、 Y 的决定因素， Z 对 X 的关系是 C ， Z 对 Y 的关系是 B ，如果 B 和 C 有相同的符号，那么忽略变量 Z 导致的偏差是正的，如果 B 和 C 符号相反，那么忽略变量 Z 造成的偏差则是负的。

比如假设 X 是受教育水平， Y 是职业， Z 是智力水平。智力水平是受教育水平和职业的决定因素，他会在这两者都产生影响，智力水平对职业的影响是正向的，对受教育水平的影响也是正向的，因此如果你忽略了智力水平在这两者之间的关系，造成的偏差就会是正的。

第二种情况是 Z 和 X 相互影响，不是 Z 导

致了 X ，而是 Z 和 X 相互影响，并且 Z 是 Y 的决定因素。如果 Z 与 Y 的关系 B 和 Z 与 X 的关系 C 具有相同的符号，那么忽略变量造成的偏差是正的，如果 B 与 C 的符号相反，那么忽略变量造成的偏差则是负的。第三种情况是 Z 是 X 作用于 Y 的一个中间变量或是解释变量，如果 B 和 C 有相同的符号，那么对直接作用而言，忽略变量 Z 导致的偏差是正的，如果 B 和 C 符号相反，忽略变量 Z 造成的偏差则是负的。比如女性的数学成绩比较低，假设数学成绩对职业选择有很大的关系，如果忽视了数学成绩在这两者之间的关系，就会造成负的偏差。

下面我们来讲一个比较重要的东西，这一段我没有翻译出来，主要是讲 Propensity Score. $P(T)$ =probability of treatment, if observed selectivity is present, $P(T)$ is a function of observed variables vector Z , we can estimate $P(T)$ through a logit model, under the assumption of no other omitted relevant factors, group T and group C are comparable, within levels of the estimated propensity score. Adjustment is thus made through a single dimension of propensity score. Propensity Score 是 1984 年发现的，发现之初并没有很多人使用，但是现在用的人比较多。

最后我要讲概率性的因果关系。我们讲的因果关系都不是绝对的，都是概

率性的因果关系。概率性的因果关系就决定了我们只能做相对的平均值分析，做趋势的分析，而不能做绝对的分析。在这种分析中我们要讲到嵌套的原因，这种概率性原因对个案的分析是不适用的，另外我们还会讲一下回归和相关在因果关系分析中的作用。我们第一个要讲的是同因不同果，因果关系的经典注释是决定性。我们讲的这种绝对性是指概率性和随机性，因为变异是社会科学的本质，相同的原因会导致不同的结果。变异又分为总体变异和时间变异，总体变异是不同个体之间的变化，比如在座的同学们，你们同样是听了 2 个小时的讲

座，但是你们对这个讲座的接收程度就不一样，有的人接收得较快，有的人较慢，这就是总体之间的个体差异。时间变异是相同个体在不同时间的变化。比如不同的时候，人的态度，人的反应是不一样的。概率性的原因就是强调平均值和趋势。随着样本规模的扩大，因果关系渐趋近于平均值，但是这种平均值并不是针对个案而言的。这个概率性原因有两层含义，一个是有原因不一定有结果，比如子女多的家庭未必子女的教育水平就低。相反，另外一层含义是结果可以在没有某个原因的情况下发生。比如吸烟与肺癌之间的关系，一个

人得了肺癌不一定就是因为他吸烟造成的，虽然在统计上而言吸烟对肺癌有直接因果关系，但是这种因果关系并不是决定性的。第三个要讲的是嵌套原因，嵌套原因是必要条件而不是充分条件，这事实上是一个风险问题。另外，概率性原因对个案是不适用的，只有在有重复事件时，概率性的因果关系才会有意义，概率论的基础是重复，假如没有重复，就不可能有概论，也不会有概率性的因果关系。假如你要研究很具体的事件，我建议你不要用统计的方法，这基本上是行不通的。

第五个要讲的是相关和回归，它们是发现因果关系的工具，这个我在以后的讲座中还会着重讲。相关和回归是一种分析的方法，但是两个事情相关并不代表他们之间存在概率性的因果关系，因为相关的关系可能有很多种，不仅仅是因果相关，因此大家不要把相关当作因果相关。由于时间的关系，我今天的讲座就到这里，谢谢大家！

马戎主任：由于时间的关系，我们就不提问了，今天的讲座到此为止。

主题：研究设计和抽样理论的基础

Basics of research design and sampling theory

主讲：谢宇教授（美国密西根大学社会学系教授）

主持：马戎教授（北京大学社会学系主任，博士生导师）

时间：2002年10月25日晚7:00

- 9 : 00

地点：交流中心新闻发布厅

马戎主任：今天，我们欢迎谢宇教授给我们做方法论系列讲座的第三讲。

谢宇教授：这是我的方法论讲座的第三讲，在第一讲里，我强调了社会科学与自然科学的差异，它们之间有很不同的支点，那就是个体与个体之间的差异和由个体组成的整体之间的差异。第二讲我讲的是做因果关系分析时经常遇到的问题，这个问题的来源就在于个体与个体之间有差异。你比较两个人或者两组人，两个国家，两个文化，除了你看到的支点以外，还有隐藏的你看不见

的东西，这就是我讲的选择性。假如这些选择性是我们观察到的，能用统计方法测量到的，我们就能够用统计的方法来控制。我在第四讲和第五讲会详细的讲一些方法上的问题，就是假如这些选择性没有被观察到的话，我们该怎么办。另外我还讲了造成忽略变量偏差的两个条件，就是在什么情况下，你忽略了变量会导致偏差。这些都是因为人与人，一件事情和另外一件事情，个体与个体之间存在差异造成的。

今天我要讲一些相对比较传统的一些概念以及方法上的研究设计，还要讲一点抽样理论的基础，抽样理论基础这

部分今天讲不完，我在第四讲里会接着讲，在这一讲我会把我认为最重要的东西强调一下，有的可能你们已经学过了，我只是给你们把重要的东西强化一下，系统的讲一下。一般讲研究设计是从分析单位讲起，然后会讲到分析单位的层次问题，再讲分析单位的总体和几种重要的分析单位。我们先来给分析单位下一个定义，分析单位就是被研究的社会实体，这些实体可能是个人，可能是家庭，可能是组织，任何我们社会学研究的对象都可以叫做研究实体。他们是研究者需要观察、了解、描述的对象。所以分析单位是很简单的，就是你想研究

的东西。分析单位是有层次的，一个较高水平上的分析单位可能包括多个其他较低水平的分析单位。比如个人是生活在家庭中的，所以一个家庭有很多个人，很多家庭又组成一个社区，多个社区组成一个县，再组成地区，再组成国家。我前面讲的社会现象存在差异，而差异就表现在分析单位的不同层次上，差异所在的地方，就是你要解释的地方，假如你不知道差异在什么地方，你就不可能找到合适的解释。比如说文化，如果你找不到文化之间的差异，你就很难解释两个国家之间的差异。因此，解释的差异和结果的差异要吻合，要不然你的

解释就是失败的。所以，你知道了你的研究对象在什么样的水平，你解释的差异和结果的差异就要在同一层次上。因为我要讲的内容很多，我会讲的比较快，如果你们有问题的话可以提出来。我讲课有一个假设，就是如果你们不提问题的话，就说明你们听懂了，如果你们没有听懂就要提问题。

你一旦建立了你的分析单位就要知道你的分析单位的总体是什么，确定分析单位的总体是非常重要的，因为你要解释的是所有现象，而不是一两个现象。我们举一个例子，如果我们要评估北大社会学系 20 年来的教学质量的趋势时，

什么是总体？我想听一下同学的回答。

学生一：是社会学系。

谢教授：社会学系不是一个变量，假如有几个系的话，我们可以做一个比较，但是只有一个系的话，他不成为一个变量。

学生二：是不是北大社会学系每年的招生质量。

谢教授：很好，你讲的是生源的质量，我的答案是所有社会学系 20 届的的大学生。以每一届学生为一个观察点，你想知道 20 年来教学质量的变化变化的话，就对每一届学生进行一个比较分析，从一届到另外一届，他的教学质量有什么差

别。因此这里的分析单位是一届学生，而分析单位的总体就是整个 20 届北大社会学系的学生。所以你选定一个研究对象后要知道他的分析单位是什么，知道了分析单位后要知道分析单位的总体是什么，这是做研究的第一步。在具体的研究中，我们会用到那些具体的分析单位呢？我这里讲几种重要的分析单位。一个是个体（Individuals），个体在社会学中基本上是指个人。个体是社会学最基本的分析单位，其他的分析单位都是由个体组成的，没有比个体更小的分析单位。第二个分析单位是人群（Groups），人群也可以作为分析单位，

人群的很多特征是由个人的特征汇总而成的，对人群测量是通过对个体的测量汇总而来。比如种族，社区，家庭等以个体集中起来的一个群体。再比如我们讲男性工作者的工资要比女性高，工资本来是以个人为单位的，我们这里讲的两性的比较就是通过对不同个体的汇总得来的。下一个概念就是组织（Organization）组织不仅仅是个人的集体，它重在于组织成员之间的关系。关系是一个人没有的，没有学生就没有老师，就没有师生关系。组织的重要性在于它强调了关系，涂尔干就强调了社会学研究的是社会现象，社会现象不是由

单个人造成的。当人们参加组织和团体活动时，个人的力量就被组织力量所掩盖。社会学之所以重视对组织的研究，是因为组织虽然是由人组成的，但是它已经超越了人本身而形成了一种关系，这种关系就是一种社会事实，而这种社会事实就是我们社会学要研究的重要部分。第四个分析单位是社会艺术品，比如书籍，建筑，绘画等。这些东西你们都熟悉，我在这里就不多讲。

下面我要讲的是时间性（The Time Dimension）。时间性是一个很重要的概念，假如我有更多的时间，我会多讲一点关于时间的东西。你要研究社会现象，

你就要关注时间，时间从方法论角度来讲是一种信息；从理论角度来讲是研究的对象。因为我们要研究社会变化，我们也要研究人的行为，人的行为是随着人的经验、经历而变化的。以前的经历会对你以后的经历产生影响，你这方面的经历对你那方面的经历会有影响，这都是与时间有关系的。你在考虑社会变化和个人变化时，都要考虑到时间性。我们做方法论的也要注意时间性，什么事情先发生，你先观察到什么，什么东西是后发生的，是你后观察到的。因此时间性是一个很重要的概念。

根据时间性，我们可以把研究设计

分成两大类，一类是横向分析，另一类是纵向分析。横向分析是对一个时点上的分析，是社会发展历程中的一个“快照定格”，要了解一个社会就要从横向分析开始，这是最基本的分析。虽然横向分析给你的信息是有限的，但是这些信息还是很有用处的，这是研究社会的第一步。比如在美国每十年进行一次人口普查，中国第五次人口普查也刚刚做完，数据刚刚出来。一次普查就是一次横向分析，是对社会的一次解剖，解剖的时间是人为定的。在美国还有很多横向分析，我这里举两个，一个是“现时人口调查”(Current Population Survey)简称CPS，

它起源于美国的大萧条，美国政府为了监测经济发展情况，防止大萧条的再度发生，由政府出资，每个月都对经济发展的一些指标进行统计分析，通过这些指标来了解经济发展的状况，看当前的经济状况是有发展还是面临危机。我们经常跟学生开玩笑，问他们知不知道 CPS，知不知道 CPU，知不知道 CPI，CPI 是通货膨胀的指标。通过调查 CPS，算一算 CPU 得出来的就是 CPI。另外一个横向研究是“监测未来”(Monitoring The Future)。这是我们密西根每年都做的一个很大的调查。这个调查是对中学生做的，美国青少年有很多不良现象。

这个调查就是监测青少年的不良行为，包括吸烟、喝酒、吸毒等，这是美国每年都在密西根做的一个项目。但是横向分析给你的信息是有限的，因为他忽略了动态的东西，为了知道动态的现象，我们就要采用纵向分析的方法。

纵向分析有助于推倒因果关系。因为从纵向分析中你会很清楚地看到什么事情先发生，什么事情后发生，从道理上来讲先发生的是原因，后发生的是结果。比如美国一个很重要的发现就是父母离婚对小孩的教育，成长状况，心理健康以及以后的行为都是不利的。如果你要研究这个项目的話，你首先要观察

到什么样的家庭是离婚的，父母离婚前小孩的行为和父母离婚后小孩的行为有什么不一样，父母离婚后，小孩会改变他的行为，离开他原来的生活轨迹。所以说纵向分析对社会学研究有很大的帮助。关于纵向分析怎么具体的来推倒因果关系，我会在后面的讲座中讲到。纵向分析有两种类型，一种是趋势分析，一种是追踪分析，其中趋势分析其实不是真正的纵向分析，我们把他们放在一起是要你们知道他们的区别。趋势分析是在不同时间点上对同一总体进行独立重复的抽样研究。我要强调的是趋势分析是对同一总体的研究，研究对象可以是

不同个体，但必须是来自同一总体，而且第一次研究和第二次研究是独立的重复研究。因此，从方法论角度来讲，趋势分析实际上不是纵向分析，他可以称作是汇集的横向分析（pooled cross-sectional studies），就是把不同的横向分析在不同的时点汇总起来。我们做横向研究并不是一次就做成的，比如人口普查，每十年就要做一次，你把一次次的的数据汇总起来就可以看到社会的变化。我们对国家，对社会，对任何总体在不同的时间抽不同的样，就是想对总体的趋向有一个了解。比如说人口的增长，教育对人口素质的影响，这些趋势

都是我们需要研究的。

这些都是通过在不同时候抽不同的样来反映总体的变化。因此趋势分析也叫做重复的横向分析。如果你在不同的时间，多次重复抽不同的样本，就能看出研究对象的发展趋势。趋势分析的例子就有前面讲过的美国的 pooled census, pooled CPS 等。你要了解美国社会的变化，你只要把这些资料汇总在一起，就能看出美国社会的很多变化。我为什么说这些趋势分析不是真正的纵向分析呢？因为真正的纵向研究是追踪研究。追踪研究是对同一人群，同一样本在不同时点上做的重复观察。前面的趋势研

究是对同一总体取不同的样本，通过不同的样本来反映总体在不同的时候的状况，从而看出总体在一段时间内的变化趋势。而追踪研究是对同一个样本的研究，我们的样本没有发生变化，第一次取样后，我们就一直针对这个样本进行研究。由于样本有时候是流动的，因此我们做追踪研究的成本很高，花费很大。我们要跟着原来的样本流动，核实样本的真实性等等，因此做追踪研究的成本是很高的。但是花费高成本来做追踪研究是有回报的，因为追踪研究要比汇集的横向研究提供更多的信息，得到的资料更加有效，科学价值更大。利用这些

资料，我们可以做出更好的，更科学，更可信的研究，这就是我们花这么多钱和精力去做追踪研究的原因。例如从1972年开始，美国做了一个“全国高中生纵向研究”的项目，研究者抽查了一部分在1972年读高二的学生，然后于1973、1974、1976、1979和1986年分别对第一次抽样调查的那批人进行追踪调查。这是由政府出资，由大学学者做的，做出来的资料在全国范围内共享，谁要用都可以，我就用这个数据做了我的博士论文，现在我还在用这个数据做关于教育方面的研究。为了让你们更明白趋势分析和追踪分析得到的数据的差异，我

给你们举一个例子。

这是性别与劳动参与情况的表格

	1970 年		1980 年	
	有工作	无工作	有工作	无工作
男性	90	10	85	15
女性	50	50	60	40

在所抽样本中，1970 年男性有 90% 有工作，10% 无工作；女性 50% 有工作，50% 没有工作，而到了 1980 年，男性 85% 有工作，15% 没有工作，女性 60% 有工作，40% 没有工作，可见在这 10 年内，女性的就业率有所增加，也表明女性的社会地位有了提高。这种研究是趋势研究中比较典型的，他是通过横向

分析的数据汇总而来。但是在 1980 年的数据中有工作的人到底是在 1970 年就有工作，还是原来没有工作后面找到工作，而 1980 年没有工作的是 1970 年有工作的还是 1970 年就没有工作的呢？这些问题我们从这个表中就得不到解决。因此，趋势分析只是能够看清研究对象的发展趋势，但是它有很大的缺陷，就是无法知道具体个体的变化。这种缺陷要怎样来弥补呢？下面我们来看一下第二个表格。

男 女

80 年有工作 80 年无工作 80 年有工作 80 年无工作

70 年有工作 80 10 15 35

70 年无工作 5 5 45 5

从这张表中，我们就能看到男性中 1980 年有工作的人有 80% 在 1970 年就已经有工作了，有 5% 在 1970 年没有工作，到了 1980 年就已经找到了工作。另外有 10% 的人在 1970 年有工作，因为某种原因如退休，失业等，到了 1980 年就没有工作了。而有 5% 的人在 1970 没有工作，到了 1980 年还是没有工作。在女性中，1970 年有工作的只有 15% 在 1980 年还有工作。而在 80 年有工作的女性中有 45% 在 1970 年没有工作，可见女性做临时的工作比较多。从这张表

中我们就能很具体地看出具体的人群的就业情况的变化。这种研究是追踪调查中比较典型的，他包含的信息量就比趋势分析得到的信息量要多得多。这就是我们为什么花费很高的成本去做追踪调查的原因。追踪调查比趋势调查能提供更多的信息，并且根据追踪调查你可以推断趋势，因为你能掌握不同个体在不同时间的状况，把这些个体汇总后就是总体的化趋势，而趋势分析仅仅能够知道研究对象大致的变化趋向，但是不知道总体中具体的个体的变化。我马上就要接着讲下一个题目，现在有什么问题大家可以问。

问题一：谢教授，我想问一下追踪调查是不是一定要以个人为单位？

谢教授：不一定，这要根据你的分析单位具体而定。如果你的分析单位是组织，是学校等，就不是以个人为单位。因此追踪调查的单位要视具体的总体的分析单位而定，并且在调查中分析单位不能变。你要根据你的分析单位来抽样，抽样好了之后就要追踪这些样本，收集他们的信息和资料。

问题二：追踪调查的时间是怎么选定的？

谢教授：这是一个很好的问题，这要根据你对这个问题的了解。在这里，

经验研究很重要，理论也很重要，理论就是你对这个世界的了解，你对全部现象的了解综合而成就是你的理论。因此对时间的选择关键在于你对这个事件的了解，如果他变化很快，你定的时间就要频繁一点。比如教育，教育在70年代变化特别大，因此那个调查项目在73年、74年连续调查了几年，你上不上大学，上大学又上什么专业，毕业后有什么打算，这都是变化很快的，因此要比较频繁地进行调查。而有些东西的变化就比较慢，如人口的增长，退休后老年人生活状况变化也变化较慢，像这些现象就可以间隔较长一段时间做一次调

查。所以选定追踪调查时间要看你所要调查的社会现象的变化速度，变化速度较快，前后调查间隔的时间就要较短，变化速度较慢，间隔时间就可以相对长一些。

问题三：如果选择的分析单位是一个群体和一个组织，如果在一段时间内组织内部发生了人员变动，这会不会是追踪调查退化成趋势调查？

谢教授：这个一个比较复杂的问题，这个问题的关键还是分析要看分析单位是什么，假如分析单位是机构的话，你设计的观察项目和指标就会跟观察个体有所不同。因此你在做机构和组织的追

踪调查的时候，你不需要考虑到组织下面个体的变化。比如研究企业的，我们就追踪企业的发展变化，看它什么时候诞生，什么时候转型，什么时候破产等。虽然在追踪组织的时候，组织内部的人员结构，人员个人能力等都会发生变化，但是这些都不是我们要追踪的范围。

问题四：追踪研究可以推断出趋势的变化，那它是不是可以代替趋势研究？

谢教授：这个问题很有意思，很多人都认为可以，所以美国可能成为较早的不用趋势研究而直接用追踪研究的国家，因为如果你能够对每一个个体都进

行追踪的话，就可以不需要趋势研究了。这就有两个问题，一个是成本的问题，追踪研究的成本很高，花费很大，它要比趋势研究贵很多。第二个是追踪研究的对象会有损失，由于很多原因比如研究对象不愿意被研究，或者是你找不到原有的研究对象等，这些原因就会导致一些追踪调查无法进行下去。所以，你假如要用从留下的人那里得到的资料做一个推广的话，就一定要做一个假设，就是没有被追踪的人可以用已经被追踪的人来代替，可以用统计的方法来推论他们的情况。因为这里需要做一个假设，所以很多统计学家对这个不满意。还有

一种方法就是不断地补充，像一个渠道一样，一定要有源头活水，这样才能保证样本的新鲜性，从而使调查资料更具有科学性。

问题五：这种追踪研究能否反映总体的结构变化？

谢教授：应该是可以的，虽然有一些东西通过追踪调查还是观察不到，但是就趋势调查和追踪调查来讲，趋势调查能做的，追踪调查基本上都能做到，只要总体不发生变化，保证了这一点，就可以通过追踪研究来观察总体结构的变化，因为你不仅能够知道总体的变化趋势，你还能知道具体的个体的变化情

况，因此通过对个体的全面了解，你就能够把握总体的结构。下面我们接着讲下一个题目。这个题目叫概念化（Conceptualization），概念化实际上是把理论的东西变成可以测量的东西，我会先给它下一个定义，再讲一下概念化的三个测量维度，然后再比较一下这三个维度，再给你们一个例子。先给你们一个定义，概念化是建立并澄清概念的过程，也就是用语言和例子来说明一个概念，以期得到一个准确的定义。只有对概念给了一个准确的定义以后，我们在研究和讨论问题时才会有一个相同的起点，否则公说公有理，婆说婆有理，

这就很难形成一个共识。因此概念化就是要将不同的观点放到同一个平面上，给他一个定义，这个定义就是一个共同的起点，有了这个起点，其他的研究才可以进行。比如什么是教育，上学是教育，实践也是教育，但是当我们研究教育时，我们就必须对教育下一个很明确的定义，给出一个概念。这就是我们经常把教育定义为在学校中得到的知识和培训。这样就把教育限定在学校中，你在学校之外得到的知识和培训就不算教育。有了这个概念以后对教育的研究才有了一个共同的起点，一个出发点。为了让大家进一步了解概念的含义和测

度，我要问你们一个问题。什么是社会地位，划分社会地位的标准是什么？

学生一：社会地位有很多种，一种是根据财富划分的社会地位，一种是根据政治职位，还一种是根据职称，比如教授等，还有可能是根据年龄和血缘关系等，这都可以用来划分社会地位。

谢教授：你讲的都是具体的标准，但不是概念性的。假如我们是讲理论上对社会地位比较重要的概念，你会想到什么？

学生二：职业及其社会影响。

学生三：个人在不同社会阶层中的地位。

学生四：社会地位的不同，关键在于划分的标准不同。

谢教授：那什么标准比较重要呢？

学生五：占有资源的标准。

谢教授：资源，还有呢？

学生六：声望。

谢教授：好，讲的很好。从我自己来看，衡量社会地位时，有三个概念很重要，你们讲的东西都可以用这三个来概括。一个是财富，一个是声望，一个是权力。有财富就有社会地位，但是有钱也不是万能的，如果他只是有钱，而没有声望和权力也不行，哈佛的教授工

资不是很高，但是他们很有声望，因此声望也是一个很重要的标准。有的人他的工资不高，声望也不高，甚至有时候还有点坏，但是他很有权力，这也能给他带来很高的社会地位。但是这三者不是一样的，有的人有财富但没有声望和权力，有的人有声望但没有权力和财富，有的人有权力但没有财富和声望。当然，很多人都想同时拥有这三样东西，但这是很难做到的。

这个例子讲的就是一个理论上的概念可能会有不同的测量维度。维度就是根据不同的含义区分成不同的类型，这些类型就被称为维度。一个概念可以包

含几个维度，因为有的概念无法通过一个维度来测量。在概念化中，我们就可以给社会地位三个维度：财富，声望和权力，这三者是衡量社会地位的三个维度。当维度不能被直接观察时，这个时候就要用到指标来观察。指标是很具体的，是可以直接被观察到的，是概念具体的量化标准。比如为了衡量权力，我们可以用这样一些指标。一个是管辖的人数有多少，管辖的人数越多说明你的权力越大。另外一个管辖范围有多大，管辖的范围越大，你的权力越大，比如系主任管辖的范围就太小，而军官管辖的范围就很大。这就说明系主任的权力

较小，而军官的权力很大。衡量权力的指标还包括每年的预算的制定，你每年的预算也是对你的权力的衡量，你的预算越多，你的权力越大，尤其是在计划经济时代，预算越多，要的越多，下一次预算就更多，这是计划经济体制的弊端。另外一个衡量权力的标准是你所控制的设施的多少，你掌握的设施越多，你的权力就越大。

下面我们来比较一下概念、维度和指标。这三者在某种情况下是一样的，可以相互转换，比如一个研究对象只有一个概念，一个维度，一个指标时他们三者就是一样的，而他们的区别就在于

概括的层次不同。概念是高度的概括，维度是一般性的概括，而指标是很具体的。所以这三者的关系就在于具体的程度。有的概念比较简单，比如性别，概念，维度，指标都是性别，只有两种可能。一个概念可以存在多个维度，一个维度可以有多个指标。当一个概念只有一个维度，而且这个维度只有一个指标时，这就是一个变量。变量是一个统计术语，是指可以取不同的数值的量，维度和指标都可以是变量。比如权力，可以大可以小，这就是一个变量。性别可以男也可以女，这也是一个变量。在复杂的情况下，一个概念可以有几个维度，

一个维度可以有不同的指标。指标是一个操作性的东西，是具体的，概念和维度是比较抽象的东西，是思想性的。下面我们讲测量（Measurement）。当我们谈到测量的时候，就必然涉及到变量和指标。因为变量和指标是很具体的，我们在做测量时必然要引入变量，而变量的测量要通过指标来表示。测量就是给分析单位赋以一个具体的数值，以表示变量的类别和差异。这个数值在不同的情况下表示不同的意义，很多人在这方面会犯错误。测量就是要抓住事物之间的差异，通过一定的指标来衡量这种差异，把研究对象区分开来。我刚才讲了

数字在不同的情况下含义是不一样的，为什么呢，因为测量的时候有不同的尺度，用不同的尺度测量出来的数值表示不同的意义。

因此我们接下来要讲测度的类型。测度的类型有四种，名义测度，次序测度，间距测度和比率测度。名义测度是根据某些变量，将案例区分成两个或者多个类别。比如你把所有的人分成不同的类，好，中，坏等。名义是可以任意赋值的，这些数字是代表任何数学意义的。比如白人 = 1，亚裔 = 2，黑人 = 3，因此这些数字本身并不代表任何数学意义，它不能用来进行运算，这里的数字

只是起到一个区分的作用。这里的分类有两个原则，一个是穷尽性，就是每个个体都必须归到其中一类，第二个是排斥性，就是一个个体不能归到两个或多个类。测度的第二种类型是次序测度，在次序测度中，赋值只是表示一个维度上的排序位置，它只能说明排序的高低，比如最典型的立克特量表。事实上次序测度并没有内在的测度标准，你仅仅知道他们之间的相对排序，不能用来做简单运算。假如你想知道不同测度之间的差距，你就要用到间距测度。间距测度假设数字间有相同的距离或间隔，间距测度的数据不仅表示排序位置，而且也

表示数值。如果一个间距测度有一个绝对零值时，那么这个间距变量就是一个比率测度变量。比如重量，它有一个绝对零值，它还有单位，因此它是一个比率测度变量。下面，我们来比较一下这四种测度，名义测度只告诉你分类，次序尺度不仅告诉你分类，还告诉你各个类型之间的排序，间距测度不仅告诉你分类，告诉你次序，还告诉你大小，而比率测度则既告诉你分类，告诉你排序，告诉你大小，还告诉你绝对零值。所以这四种测度是由一般向特殊变化的，最一般的是名义测度，所有的研究都是从这里开始，因为最开始我们要根据研究

对象的差异分出不同的类别，然后随着研究的深入，测度也要不断地特殊化，具体化。

下面，我们讲估计的基础知识（Basics Of Estimation）。统计的知识很多，书本上讲的不是很难，但是要问为什么就比较难。所以我想给大家重温一下估计的基础知识。我们做统计，首先要知道我们做统计是不得以的，如果我们可以直接研究总体的话，我们就不需要样本，就不需要统计，因为我们做统计是用样本来推论总体，如果我们能够直接对总体进行研究分析的话，我们就可以不需要统计。什么是总体，总体就

是在一个封闭的系统中，所有元素的完全排列。样本就是总体中的一个子群，总体中的任何一个子群都可以叫样本。我给你们举一个例子，我住的那个地方 1990 年普查的时候有 109592 个人，我们抽了一个 1096 个人的样本进行研究。我下一次会讲到具体抽样是怎么抽的，在这里先不讲。然后是参数与估计。具体而言，我们研究总体就是要研究总体的参数，参数是总体的某个特征，比如中心趋势的测度和离散趋势的测度等。我们用 θ 来表示总体参数 (population parameter)，这个总体参数是我们永远得不到的，我们只能通过估计来得到一个

尽量接近总体参数的估计值。估计是根据样本来推算真实参数的估计值，这就是样本估计值，用 T 表示。样本估计值也被称作样本统计量 (sample statistic)。这里有两个名词我希望大家能记住，一个是总体参数，一个是样本统计量。我们想要总体参数，但是我们得不到，我们只能通过样本统计量对它进行估计。比如我们想知道以下总体的参数，如总体的均值，总体的方差，总体的相关性，总体的交互表等。我们就可以通过样本统计量来对总体参数进行估计，比如用样本的均值来估计总体的均值，用样本的方差来估计总体的方差，用样本的相

关性来估计总体的相关性，用样本的交互表来估计总体的交互表。样本统计量是我们可以计算出来的，通过样本统计量来估计总体参数就叫做统计推论（statistical inference）。统计推论是用来评价样本统计量的可靠性的一种方法。对于任何总体参数都会有许多样本统计值，为了评价样本统计值的好坏，我们就要进行统计推论，而要进行统计推论，我们首先必须对总体进行抽样。最简单的抽样方法就是回置的简单随机抽样。它有几个特点，一个是简单，简单是指总体中每一个元素被抽到的概率是一样的，随机是指任意抽取样本，回置就是

指抽中的元素要重新放回总体。回置还要看总体和样本的大小，假如样本相对于总体而言很小的话，回置与不回置是没有多大区别的。我在这里举一个例子，在一个封闭的区域内，科学家要怎样来估计熊猫的数量。这里使用的方法是一种我们很常用的方法，叫做“捕捉—再捕捉”的方法，这里有一个假设，就是假设捕捉是完全任意的。我们先捕捉一批熊猫，做上记号后放回去，过一段时间后我们再捕捉一批熊猫，看第二次捕捉到的熊猫中有多少是第一次捕捉到的，如果第二次被抓住的熊猫中很多是第一次被抓住的，就说明总数很小，而如果第

二次被抓住的熊猫中很少是第一次被抓住的，就说明熊猫的总数很大。假如通过抽样我们得到这样一个数据。

第二次捕捉 总数

捕捉到 未捕捉到

第一次捕捉 捕捉到 10 90 100

未捕捉到 90

总数 100

第一次抓了 100 只，第二次又抓了 100 只，第二次抓的 100 只熊猫中有 10 个是第一次抓到的。我们假设这两次捕捉是独立的，是不相关的，那么两次被抓住的可能性是第一次被抓住的可能性乘上第二次被抓住的可能性，我们假设

熊猫总数为 N ，第一次被抓住的概率是 $100/N$ ，第二次被抓住的概率也是 $100/N$ ，而两次都被抓住的概率则是 $10/N$ ，就有 $100/N$ 乘上 $100/N$ 等于 $10/N$ ，算出来的 N 就是 1000。然后我们要讲估计值的抽样分布，我认为统计中很大的一个障碍就是你们不了解抽样分布。当我们做统计时，我们并没有观察总体，而仅仅是观察了一个样本，假设我们从一个总体中重复的抽样，所抽取的各个样本可能是不同的，根据许多次重复的抽样，同一估计值可以形成一个分布，这就是抽样分布。在我们实际的操作当中，我们总是把自己通过统计得到的样本统计值

放在抽样分布的中间，认为即便这种分布会发生变化也只是在这个样本统计值左右变化。这种分布是假设性的，因为我们只取了一个样本，只得到了一个样本估计值，而其它的估计值是在假定的重复抽样条件下设想出来的，因此这种抽样分布是假设性的。今天我们就讲到这里，谢谢大家！

马戎主任：由于时间关系，我们就不提问了，今天的讲座就到这里。

发表时间：2004-12-9 12:55:00

主题：忽略变量偏差 和生态学分析

Omitted-variable bias and ecological
analysis

主讲：谢宇教授（美国密西根大学
社会学系教授）

主持：马戎教授（北京大学社会学
系主任，博士生导师）

时间：2002年10月28日晚7:00

- 9:00

地点：五四体育馆体教报告厅

马戎主任：今天谢宇教授给我们讲方法论的第四讲，希望大家在讲座期间尽量少走动，以保持安静。好，下面我们用掌声欢迎谢宇教授。

谢宇教授：谢谢大家！这是方法论的第四讲，上一次我把社会学方法的基本概念介绍了一下，今天我先把上次没有讲完的抽样理论讲完，然后我会讲到忽略变量偏差与生态学分析。今天的内容比较多，可能提问的时间会比较少。首先我们讲抽样设计，这些你们都已经

学过了，我在这里只是给你们做一个简单的复习。第一个是简单随机抽样。它实际上是一个很简单的操作，先给总体中的每一个元素一个数值，然后通过随机抽取一部分作为样本，这个过程经常是由计算机来操作的。这种方法与抽奖是一个道理，简单是指每个元素被抽到的概率是一样的。第二种抽样方法是系统抽样或叫等距抽样，是先根据总体的规模和样本的规模计算抽样差距，然后对总体的所有元素进行排序，根据抽样差距把总体分成等距的多个部分，在第一个部分中随机抽取一个元素，然后按照抽样差距依次抽样。这种方法比简单

随机抽样还简单，但是使用这种方法时要注意一点，就是在对总体进行排列时，排列顺序不能与自然排列有关，否则抽出来的样本就不具有代表性，可能会忽略很多重要的信息。要强调的一种抽样方法是分层抽样，分层抽样要比简单随机抽样和系统抽样更复杂。但是任何方法都是有得有失的，付出的代价越多得到的回报也越多。分层抽样虽然复杂，但是他有他的好处，他抽出来的样本保证了某些重要特征的代表性。比如我们要抽取一个样本容量为 100 的样本，假如我们用简单随机抽样的话，我们无法保证样本的性别比。但是如果我们的研

究需要保证男女性别比例相同的话，我们就要用分层抽样的方法。我们先把总体分成两个子总体，在这个例子中我们先把男性和女性分成两个不同的子总体，然后在每个子总体中随机抽取 50 个人，这样就保证了男女性别比为 1 : 1。这种抽样方法用的很多，在研究不同种族，不同地区，贫富差距等问题时，我们都可以用这个方法来进行抽样。有了分层抽样的基础就可以过度抽样。过度抽样是指当子总体比例不一致时，在一个子总体中抽样多，在另一个子总体中抽样少，目的是使样本比例比较平衡，因为在总体中他们的比例是不平衡的。

比如研究贫困问题，贫困的人在总人口中是少数，如果我们要保证样本中有相对平衡的贫困人和非贫困人，相对于贫困人的总体而言，我们抽取的贫困人的概率就要比在富人中抽取相应数量的样本的概率更大。这种方法还可以用于做民族之间的比较研究。比如汉族人口占中国人口的绝大多数，我们在抽样时可以用过度抽样的方法，增加少数民族的人数，使各个组的比例比较平衡，这样就可以增加样本中任何一组的代表性。分层抽样的应用很多，在美国做研究时要分成南方和北方，在中国分成城市和农村，这就是为了保证样本的代表性。

整群抽样与分层抽样是相反的，它是从经济的角度考虑的，是为了省钱。但是要省钱就要付出代价，整群抽样的代价就是抽取的数据质量不高。比如我们对北大学生的经历很感兴趣，我们在抽样时，非常经济的一种方法就是我去随机抽取一个班级，把这个班级的所有个体作为一个样本，然后对他们进行研究。因此整群抽样是把总体分成不同的组，然后以组为单位进行随机抽取，抽到的组中的所有元素都进入样本。它跟分层抽样不同，分层抽样是先把总体分成不同的子总体，然后在每一个子总体抽取样本，这样就保证了样本的代表性，

而整群抽样得到的数据的质量就不高。但是整群抽样比较实用，比如你要做一个乡村研究，你不可能在全国各地都抽取一个村庄进行研究，而只能进行整群抽样，虽然样本质量不高，但比较实用，也比较经济。另外一个整群抽样获得的样本的有效性较低，因为整群抽样抽得的样本的内部重复的概率很大，很多样本无效，致使可用的样本数量大大降低。为了使整群抽样做得比较好，我们要假设各个群体之间没有很大的差别，假如这个假设不成立，抽取的样本比简单随机抽样具有更大的差异，从而降低样本有效性。比如就刚才那个北大的例

子，我们要使整群抽样做得比较好，就假设北大学生在班级与班级之间没有很大的差别。他们的知识面，家庭经济条件，朋友关系程度等都比较相似。如果这个假设成立的话，一个班与另一个班没有很大的差别，这时整群抽样不会造成很大的有效性的降低，假如班级与班级之间差别很大，那么整群抽样对有效性的丧失就会很大，这个问题我在后面还会讲到。一般来说，整群抽样只会降低有效性，而不会增加有效性，降低多少，要看具体的问题。我现在用稍微系统一些的方法来讲这个事情，我们来探讨一下样本统计量中方差的来源。一个

来源是总体差异，一个总体中的所有元素之间是有差异的，总体差异是元素差异的总和。我第一讲讲的就是社会现象之间有差异性，而且这种差异性是一必然存在的，我们社会学研究也就是对这种差异的研究。我们之所以要抽样，就是因为总体内部各元素之间有差异，如果社会现象之间没有差异的话我们就没必要抽样了。我第一讲讲过社会科学与自然科学的区别就在于社会科学说研究的社会现象有差异，而自然科学可以通过一个典型现象得到真理性的规律，从而举一反三。社会科学就不行，他必须关注社会现实的总体，而总体内部是有差

异的，这就要求我们在做研究的时候要
进行抽样，从样本来推测总体。

既然总体内部有差异，那么哪一个
元素被选择，哪一个元素不被选中就会
影响到样本，从而影响呢最后得到的样
本参数，具体的说是因为总体中的各元
素的某一变量值不同，导致了元素之间
的差异。因此我们要进行抽样，而抽样
是一种随机选择，这种随机选择必然会
造成另一种差异。因为你对不同元素的
选择会直接影响到样本统计值。

可见，样本统计量中方差的来源主
要有两个，一个是总体差异，这种差异
是客观存在的，另一个是抽样差异，就

是在对样本进行选择时造成的差异，这种差异其实是由总体差异决定的，如果总体之间不存在差异的话，我们就不需要抽样，也就不存在抽样差异。因此我们可以对总方差进行分解，把总体方差分解成两部分，组间差异和组内差异。组间差异是指各组平均值之间的差异，一个班级和另一个班级差在什么地方，就差在平均值。组内差异是同一组中各元素相对于组平均值的差异。方差分析就是组间差异加上组内差异。组间差异大的时候组内差异就比较小，而组内差异大的话，组间差异就会比较小。也就是内部异质性越大，外部同质性越强；

内部同质性越强，外部异质性越大。因此分层在这个地方就有很大的作用，我们在做分层抽样的时候，通过分层就可以减少抽样方差，因为我们在对总体进行分层之后再抽样，这就控制了组间方差，抽样方差就只是组内方差，而由于分层之后组内元素同质性较强，因此组内方差也会比较小。各层之间的异质性越强，层内同质性越强，分层得到的精度越大。因此，假如层与层之间差异很大，或者层内部各元素较均匀的时候，我们使用分层抽样得到的效果就会很好，相反假如层与层之间没有什么差异，我们做分层抽样就没有什么意义。比如

我们做城乡差别的研究，就必须进行分层抽样，因为城乡之间存在很大差异。另外美国的南方和北方也要分层，他们在经济上，文化上都存在很大差异，如果你不分层的话就会损失很大的精度。过分层抽样，我们就引出了设计效应。设计效应就是复杂样本设计的抽样方差与同样规模的简单随机抽样的抽样方差的比值。设计效应是一个比值，是复杂抽样方差和简单随机抽样方差的比值。那么分层抽样的设计效应会大于 1 还是小于 1 呢？因为我们做分层抽样的时候控制了组间差异，这就使得分层抽样的方差会小于简单随机抽样的方差，因此

分层抽样的设计效应会小于 1，仅当各层之间无差异时分层抽样的设计效应会等于 1。而整群抽样的设计效应呢，因为整群抽样的样本有效性较差，内部差异较大，整群抽样的抽样方差会大于简单随机抽样的方差。因此，整群抽样的设计效应会大于 1，仅当各群之间无差异时，设计效应会等于 1。我们比较一下分层抽样和整群抽样就能看到，假如层间异质性很强，层内同质性很强时，用分层抽样的效果会比较好，精度会比较大。如果这种情况下用整群抽样的话，就会大大降低抽样的有效性。以上就是抽样的部分，这部分我们就讲到这里。

下面我们转入另一章，分析因果推理的潜在危险。从一定程度上来讲，我要讲的这一部分是第二个讲座有关因果推理的一个延伸，只是我们在这里加入了生态学分析。真正的因果关系是很难找到的，它有几个潜在的危险，一个是伪相关，一个是表面上的时间顺序，你不要以为知道了事情发生的前后顺序就能找到因果关系，发生在前面的不一定是原因，发生在后面的也不一定是结果。另外一个忽略变量偏差，这个我们在上一讲中也讲到过，在这里我还要强调一下。第四个是样本选择性偏差，我刚刚讲完了样本选择的问题，因此我们要

到现在来讲样本选择性的偏差。

我们先讲伪相关。A 和 B 之间有关系，但是这种相关是通过 C 来实现的，假如 A 和 B 的关系是又 C 造成的，那么 A 和 B 在表面上的因果关系就是伪相关。比如有人认为数学成绩取决于语言能力，语言能力越强，数学成绩越好，这里就忽略了一个智力因素，智力强的人语言能力强，数学成绩也好，因此语言能力与数学成绩的相关是通过智力因素来实现的。语言能力与数学成绩之间的因果关系就是伪相关。第二个就是表面上的时间顺序。我曾有幸和一位社会学大师进行过一次座谈，他对我讲，社

会科学要找到真正的因果关系是很难的。当时我就提出来是不是能够根据事情发展的前后顺序来寻找因果关系，在前的是原因，在后的是结果，因此我们通过纵向研究就可以找到这种因果关系。他立刻提出了反对意见，他举了一个例子，圣诞节前夕会有一个购物高潮，那么是圣诞节造成了这种购物高潮，还是购物高潮造成了圣诞节？我一想，这确实很有道理。我们人的行为是有目的性的，这种目的性促使我们去预测事物的发展方向，从而提前做好某些准备，这就导致了有些事情结果在前，而原因在后。有很多人做很多事情是为了将来，

而不是为了现在，因此我们要侧重从理论上对它进行分析，找出真正的因果关系。第三个威胁到因果推论的因素是忽略变量偏差，这一点我们前面讲过了，我这里再强调一下。忽略变量偏差的产生要有两个条件，一个是忽略的变量要与自变量相关，另一个是忽略的变量要与重要的因变量相关。这两个条件都要成立，缺少其中一个都不会造成忽略变量偏差，我等下讲到生态学分析的时候会再讲到这两个条件。

问题：谢教授，我想问一下忽略变量偏差和前面的伪相关有什么区别吗？他们都是因为忽略了一个中间变量而造

成的。

谢教授：这个问题问得很好，从统计上来讲，它们之间是没有区别的。这是一个概念性的东西，这四个因素在本质上都是一样的，我只是根据比较传统的方法把他们区分出来，但是在统计学意义上，他们是一样的。我这里把忽略变量偏差强调出来，其实其他几个都是忽略变量偏差的一种表现形式，而忽略变量偏差也是其他几个的表现形势之一，其实从统计学意义上而言，他们在本质上是一样的。

我再举一个例子，假如北京有两家医院，一家非常好，有先进的设备，一

流的医生，管理和服务也特别好，另一家是很差的医院，设备陈旧，技术落后，医生医德不好等等。假如你以哪个医院死的人少来衡量医院的好坏的话，就会出现忽略变量偏差，因为肯定是好的医院死人多。死亡是与病情有关的，病情越严重，死亡率越高，另外病情越严重越容易被送到好医院，因为差的医院看不了，所以好医院死的人会比差医院要多，如果你忽略了病情这个变量的话，就会造成忽略变量偏差，因为病情严重与否与死亡有直接关系，另外病情的严重程度也会影响到对医院的选择，因此忽略了病情因素而得出死人少的医院是

好医院的结论造成了忽略变量偏差。

第四个是抽样选择性偏差。假如你所得到的样本对总体没有很好的代表性，就会造成抽样选择性偏差，比如有一个面对物理学研究生的奖学金，这个奖学金是根据研究生入学考试 GRE 的三个部分成绩来评定的，一个部分是语言，一个部分是数学，一个部分是分析。物理学家不知道该把奖学金给哪些人，就请了一个社会学家来评定。最后社会学家认为这个奖学金应该给语言能力好的人，为什么物理学系的奖学金要由语言能力来决定？因为申请这个奖学金的人都是数学和分析能力很强的人，他们

的差异关键就在于语言能力。这个问题的关键在于样本的选择性，假如他这个奖学金是面向所有同学的话，就不会出现这样的结果，而他的样本是来自于读物理学的研究生，所以才出现这样一个抽样选择性偏差的问题。

接下来，我举另外一个例子，这是一个模型，叫 Willis and Rosen 模型。他们这个模型讲的是高等教育的经济回报的问题，他们认为有些人通过接受高等教育能够增加他们的生产能力，这些人如果不让他们去读大学而是让他们去从事体力劳动的话，他们创造不了多大的社会价值。另外，有些人你让他去接受

高等教育反而不如让他去接受生活技能的训练，读大学对他们没有什么收益。这就是两组不同的人，一组人不读大学要比读大学要好，一组人是读大学要比不读大学好，如果你让这两种不同的人去做他们不擅长的事情，他们就不会有很大的发展空间。这就是一种自我选择，高等教育的经济回报有的人高，有的人低，回报高的人就去读大学，而那些没有去读大学的人正是那些回报低的人。人和人是不一样的，去读大学的人正是因为回报高才去读，不选择读大学的人，他们选择了另外一种适于他们工作，并且回报高的职业。因此高等教育的回报

率是因人而异的，如果你在抽样的时候不考虑到这种差异，就会出现抽样选择性偏差。

下面我们要切入今天的正题，就是生态学谬误。生态学谬误是这样一个错误，对于汇总层次上的信息判断到组成这个汇总层次的低一级分析单位上，这就可能犯了生态学谬误。就是你把高层次的信息、经验、发现应用到低层次的分析单位上，你就可能犯了一个错误，我要强调这只是一种可能的错误，这种错误就叫生态学谬误。也就是说假如你利用汇总层次的证据或事实，你把这种证据或事实作为对个体的推理，你的理

论是建立在个人层次上，你就可能犯了生态学谬误。我给你们举一个例子，有两个班级，A 班和 B 班，A 班学生学一门课的平均时间是 10 小时，平均成绩是 90 分；B 班学生学这门课的平均时间是 15 小时，而平均成绩是 80，由此我们得出结论：虽然 A 班学生比 B 班学生花的时间少，但成绩比 B 班学生好，因此学习时间和成绩是负相关的。这个结论对不对，那个地方有问题？问题就出在这里，你的理论假设是在个人层面上，花多少时间学习是个人的事情，而你的结论是以班级为单位的，统计数据也是在班级层面上，这两个不吻合，这就有可

能造成生态学谬误，不能由此得出学习时间会造成成绩负效的结论。这里可能的现象是在一个班级里面，就个人而言，学习时间对成绩是正相关的，但是由于两个班级学习起点不同，智力水平不同，一个班级学生花的时间少，但是成绩提高较快，另一个班级学生花的时间较多，但是成绩提高较慢。因此对于个人而言，花的时间的多少和学习成绩是成正向关系的，但你平均了以后，在班级的单位上就是负面关系，这就是一个很经典的生态学谬误。

我们讲生态学谬误就必须讲到 Mr Robinson，他是一个很有才华的社会学

家。他在 1950 年发表了一篇论文，这篇文章已经成为一篇很经典的文章，我现在都常教导学生去读这篇文章。在这篇文章里就提出了生态学谬误的问题。这篇文章产生了一个危机，对他自己也产生了一个危机，在他写完这篇文章后就决定歇笔了，为什么呢？因为 1950 年以前，所有的数据都是汇总的数据，都是以地区为单位的，Robinson 就说假如我们社会科学的目的是要了解人类的行为的话，这种汇总数据是不能用来推论人的行为的。人们为什么结婚，为什么生育，为什么上学等等，假如我们要研究这些的话，用汇总的材料是办不到。

因此 Robinson 在写了这篇文章后，就决定不再做研究了。

这是对社会学的一个很大的挑战，这个危机也给下一代人带来了机会，这就是“调查革命”。这个“调查革命”的发起就是为了应付 Robinson 提出的挑战。在研究中，我们不用政府提供的汇总材料，而去做个体和家庭的研究，收集个体层面上的资料。我们密西根大学社会学系在那个时候就做出了很大的贡献，不仅从理论上，也在实践上做了很多事情，这些我会在星期五的关于美国社会和美国社会学的讲座中详细讲到。调查革命就提出来要收集以个人为单位的，这种

单位是要与你的理论层次相一致的数据。你们没有经历过那段只有汇总数据的日子，当时所有社会学研究用的都是政府提供的汇总数据，正是因为有了调查革命，才有了密西根大学社会学系的兴起。我再讲的另外一个人物叫 Gary King，他是哈佛大学的一个政治学家，他最近出了一本关于生态学谬误的热门书。他这本书里讲的是怎样解决生态学谬误，你们如果对这个感兴趣的话，可以去看一下他这本书。下面我们要讲一下为什么会产生生态学谬误，是因为未观察到差异性偏差或忽略变量偏差。这就回到了刚才那位同学问的问题，其实

这些原因在本质上是一致的，忽略变量偏差只是其中的一种表现形式。这里我们就要回顾一下造成忽略变量偏差的两个条件，一个是忽略的变量要和因变量有关系，另一个是忽略的变量要与自变量有关系。在那个学生成绩和花的时间的关系的例子中，忽略的变量是什么？就是学生的个人学习能力，学习能力强的学生花的时间少，学习成绩也好，学习能力较差的学生花的时间多，成绩还相对较差。也就是说学生的学习能力既影响到学生的个人学习时间，也影响到学习成绩的好坏，因此当你忽略了学习能力这个变量时就会造成忽略变量偏

差。另外一个是这样差异是以班级为单位的，假如班级与班级之间不存在能力的差异的话，也不会出现这种忽略变量偏差，因此你在寻找差异的时候是应该找以班级为单位的原因，而不是找以个人为单位的。我先讲抽样，再讲生态学谬误，然后才讲忽略变量偏差，其实当我们在利用抽样获得的数据进行分析的时候，可能出现的错误都是因为忽略了某些变量造成的。下面我们再来举一个例子，这是一个忽略变量偏差的例子，不过这里也有生态学谬误的因素。这是一个真实的例子，佛罗里达州是还保留有判死刑的一个州，有一个社会学家花

了 20 年的时间做了一个调查，对 300 多个凶杀案进行了分析。凶杀案是很容易被判死刑的，在美国种族问题是很敏感的问题，在判刑的问题上是不是有种族歧视，是不是黑人更容易被判死刑？我们来看下面的数据。

佛罗里达州 20 个县的 326 位被告凶杀案的有关信息

被告人的种族 被判死刑 总计 比率%

是 否

白人 19 141 162 12

黑人 17 149 166 10

在 326 位被告人中，白人有 162 位，

其中 19 位被判了死刑，所占比率为 12%。黑人有 149 位被告犯了凶杀案，被判死刑的有 17 位，占 10%。这个数据和我们的最初的想法刚好是相反的，白人被判死刑的概率比黑人还要高。请问为什么？

学生一：可能白人和黑人犯的罪行不一样，白人犯的罪行更严重。

谢教授：你们想一下造成忽略变量偏差的两个条件，被忽略的这个变量是什么，在这里它既要和种族有关系，又要和被判死刑有关系。如果这个变量是罪行的话，那么罪行和种族有什么关系，是白人更容易犯严重的罪行吗？

学生二：这可能和法官的种族有关系，一般法官会偏袒自己种族的人。

谢教授：显然，美国社会的白人法官要大大多于黑人法官，按照你的说法白人法官会偏袒白人的话，白人中被判死刑的人应该更少才对，怎么会更多呢？要注意忽略的变量应该与因变量有关系。

学生三：可能佛罗里达州的白人更多，黑人更少。

谢教授：我们的案例中被告凶杀案的白人和黑人的数量是差不多的。

学生四：是不是与被害者的种族有关系，如果被害者是白人就更容易被判

死刑。

谢教授：对，讲的很好，原因就在这里。为什么呢，这里有两个条件，首先，被害者的种族和犯罪者的种族是相关的，第二是杀了白人更容易判死刑。通过下面这个图表就可以看出来。被告人的种族 被害者的种族 被判死刑 比例%

是 否

白人 白人 19 132 12.6

黑人 0 9 0

黑人 白人 11 52 17.5

黑人 6 97 5.8

可见，白人杀白人被判死刑的概率很高，黑人杀白人被判死刑的概率更高，而白人杀黑人就几乎没有被判死刑的，黑人杀黑人被判死刑的概率也比较小。因此这里忽略的变量是被害者的种族，当被害者是白人时，不管被告人是白人还是黑人，被判死刑的概率都较高，当被害者是黑人时，被判死刑的概率就较低。可见，在美国，种族歧视还是很严重的。我们这里犯错误的原因是认为罪行是与罪犯的种族有关，而实际上被判的罪行是与被害者的种族有关的。因此，我们在考虑问题的时候，要把造成忽略变量的偏差的两个条件都考虑进

去，忽略的变量是否与因变量有关系，同时这个变量是否与自变量有关系，只有当这个变量满足了这两个条件时，忽略了这个变量才会造成忽略变量偏差。

上面这个案例其实也是一个生态学谬误的问题，因为我们从第一个表中得到的结论是被判的罪行与罪犯的种族有关系，而实际上被判的罪行是和他所犯的罪行有关的，如果被害人是白人的话，被告人就更容易被判死刑。因此从这个角度来看，这也是一个生态学谬误。我举这个例子的目的有两个，一个是要让你们明确忽略变量偏差所需要的两个条件，另外一个是为了让你们明白忽略变量偏

差和生态学谬误在本质上是一样的。

接下来我要讲多层次数据的处理手段，这部分是这一讲中比较新的东西，也比较难理解。多层次数据就是说你的数据的来源是多层次的，既有团体层次的数据，也有个体层次的数据。比如你的学习能力是个体的，家庭背景是个体的，但是教师水平，设备等都是以班级为单位的。假如你的研究单位是个体，你就不能用班级层次的数据来汇总；假如你的研究单位是团体，你也可以用个体层面的数据来汇总。我这里讲两个模型，一个是随机效应模型，一个是固定效应模型，然后再讲层次分析。我在讲

到这两个模型的时候，会讲到组内差异和组间差异的问题。我先讲第一个模型，随机效应模型，在这个模型里，要引入组的概念，我们分析得到的数据时要分析两部分差异，组与组之间的差别和组内的差别。组与组之间的差异是随机的，我们对两个组各取一个平均值，两个平均值相减就得到了组内方差。假如各组之间没有选择性，那么既可以用组间方差，也可以用组内方差。在这种情况下，如果你仅有一组这一层的变量的话，就不会出现生态学谬误。在这里最好的方法是两个都用，这样得到的数据会更加科学。但是如果你只有一个层次的变量，

你就只用这个也行，只是效度不如前一个高。由于随机变量模型无法完全解决这个问题，这就引入了固定效应模型，这一手段是用来控制“非观察的异质性”，组间差异不一定符合统计模型。此手段放弃组间差异，仅关注组内方差，它之所以被称为固定效益模型，是因为总体的异质性所产生的共同特征已被各组的阿尔法所控制。这种方法在方法论上不是很牢靠的，但是在方法论上很严格很站得住脚的方法，会牺牲很多统计的信息。最后一个就是多层次分析，第一个随机效应模型讲的是组与组的差异，这种差异是随机的，第二个固定效应模型

就是把这种差异作为一个固定不变的东西，这种差异是不可以用统计模型来进行分析的。前一种是不对差异进行考虑，只把它作为一个随机变量，第二种是考虑这种差异，但把这种差异固定下来，不对其进行分析。这两者代表两个极端，他们居中的一种方法就是多层分析方法。这种方法有一个很强的理论来源，就是人与人是有差异的，而且人的表现，行为方式是随着环境的变化而变化的。那么我们就想知道人在不同环境下的表现是怎样的，他的行为是怎样的，这是社会学上一个很重要的东西。涂尔干提出了“社会事实”的概念，说人到了一个

团体之中，就要受到环境的影响，行为方式也会发生变化。当我们想知道人是怎样受到环境影响时，就要用到多层次分析方法，这就是多层次分析方法的理论来源。在使用这个方法时，研究的因变量是个人层次的，但自变量（或解释变量）即可以是个人层次上的，也可以是更高水平层次上的，比如社会环境，学校，班级等。当你的数据具备这样的要求时，你就可以做多层次分析。多层次分析的关键在于宏观层次和微观层次的自变量的交互作用。自变量有两个层次，一个是个人的，一个是环境的。比如个人的行为随着环境的变化而变化，

这就是说个人作为自变量时，其参数是随着环境的改变而改变的，这就是一种交互作用。因此，如果个人水平上的变量随着环境的变化而变化的话，你就可以做多层次分析。我可以给你们一个例子，这是我和我的一个学生做的一个研究。我们做这个研究实际上是为了批评一种市场交换模型。他这种模型认为教育的回报率是与市场发展程度成正比的，市场经济越发达的地方，教育回报率会高。为了反对他的观点，我们就做了一个地区差异比较，发现在市场经济发达的地方受教育水平较高，但是教育的回报率并不高。而在市场经济不发达

的地方受教育水平较低，但是教育的回报率却比较高，而且市场经济发达的地方的教育回报率反而比市场经济相对不发达的地方要低。这就反映出很多光靠市场经济不能解释道问题。这种现象在单位中也存在，收益好的单位，大家得到的钱都比较多，但是差异并不是很大，也就是回报率比较小，反而那些收益差的单位，其领导得到的报酬比一般员工得到的要多的多，这是一个体制的问题。好，我今天要讲的东西讲完了，大家有没有什么问题？

问题一：谢教授，您讲教育回报率随着市场发展而降低，这里您有没有考

考虑到教育对政治资本的影响，是不是教育水平越高，政治资本越大，因此回报率也更高？

谢教授：我们考虑到了政治资本的因素，政治资本是不随经济发展而变化的，它确实会受到教育水平的影响，但我们这里讲的回报率是指经济回报率，不考虑政治因素。

马戎主任：由于时间的关系，我们今天的讲座就到这里。

发表时间：2004-12-9 12:57:00

主题：选择性、实验 法和评估研究

Selectivity, experimentation, and
evaluation research

主讲：谢宇教授（美国密西根大学
社会学系教授）

主持：马戎教授（北京大学社会学
系主任，博士生导师）

时间：2002年10月30日晚7：00

- 9：00

地点：五四体育馆体教电教厅

马戎主任：今天由谢宇教授给我们做方法论的第五个讲座，大家欢迎！

谢宇教授：谢谢大家！今天是第五讲，也是方法论的最后一讲。我想把以前讲的东西归纳一下，再把一些新的思想介绍一下。感兴趣的同学可以根据我后面讲的东西去参考具体的文献，去掌握这些方法，我在这里只是简单介绍一下。今天的题目是选择性，实验法和评估研究。我会先把这个题目解释一下，

下一个定义，然后再讲他们之间的关系。听过我的讲座的人可能会得到这样一个共识，就是社会学研究中最根本的问题就在于因为选择性的存在，观察数据所提供的变量关系可能是虚假的。因为社会现象是多变的，是有差异的，因为这些差异性，我们在取样时就会产生差异，如果你不注意这些差异的话，你得到的变量关系就有可能是表面上的虚假关系。因此社会学方法论当中最根本的问题就是由于选择性的存在，使我们观察到的变量关系可能是虚假的。这是对我们最大的一个挑战。那怎么来解决这个问题呢？有两种方法可以解决这个问题。

题，一种是用观察数据建立一个统计模型，也称为“结构法”。组与组之间的差别可能是不可比较的，那我们有一个办法，就是把各个组之间的差异全部测量出来，然后用统计的方法建立一个统计模型，把所有因素都考虑进去。比如我的一个老师，现在也在威斯康星大学做教授，他就讲过传统的社会分层的模型没有考虑到人与人之间的权力关系，而马克思就考虑到了阶级是由权力不平等造成的，因此他的模型能够解释更多的现象。这是一种方法，是用数据和模型来完善一些关系，弥补一些漏洞。但是这种方法有很大的缺陷，因为它不知道自

己漏掉了什么，这就需要另一种方法，叫实验法，也可以称为“简化法”，是通过实验来解决选择性的问题。所有的方法都是这两种方法的演变和具体实现。

下面我们来讲评估研究，评估研究也称“项目评估”，是指试图评估社会项目的有效性的一种应用性研究。评估研究实际上不是一种研究方法，只是一种应用性的研究，其目的只是在于评估研究项目的有效性。我刚才讲了选择性，选择性大致可以分为两种。一种叫可观察到的选择性，这种选择性是指如果接受与未接受社会干预的两组研究对象在观察特征上是有差异的，那么这种选择性

就称为可观察的选择性。在多元统计分析中，可以通过统计分析使这两种研究对象具有可比性，这就解决了可观察的选择性问题。另一种选择性是未观察到的选择性，这种选择性是无法测量到的特征的选择性。在社会学研究中，更为困难的问题就是处理未测量特征的选择性。这个问题是社会科学中最大的问题，你解决了这个问题就解决了社会科学中最大的难题。这个问题也被称为“内生问题”，原因是项目参与者对于被评价的结果变量而言是有内生选择性的。人是有目的性的，他是根据他的目的性在参与研究项目的，这种目的性就是一个内生

问题，而不是受外来因素控制和影响的。这个问题比较难以解决，统计模型需要很强的不切实际的假设。有一位经济学家通过建立这方面的一个模型而得了诺贝尔奖。但是他这个模型需要的假设太多，难以操作化，因此现在对这个模型持批评态度的人越来越多了。因为他的假设太多，我们根本无法知道未观察到的选择性的存在性。这一潜在的威胁对所有根据观察数据进行研究的人都是存在的。我们做社会科学研究的人都是用的观察数据。只要我们用的是观察数据，我们在写文章，做报告时，我们的结论就有可能是错误的，因为有可能有未观

察到的选择性存在。极端一点来说，所有的社会科学的观察资料都是不科学的，都是可以提出疑问的。我在星期五讲美国社会和美国社会学的时候，会讲到经验研究在美国社会学是一个很大的主流，但在最近 20 年，我们已经意识到经验研究和统计分析是有局限性的，并不是万能的，我们的数据和方法都是有缺陷的，我们的结论也有可能不成立。因为我们在做研究的时候要做出种种假设，不做假设就得不到这种结果，社会科学的统计分析中不做假设是寸步难行的。

那么这种观察性偏差是怎么产生的

呢？我们就要探讨一下这种偏差的来源。一个来源是影响结果的未观察到的前提条件在组与组之间存在差异，就是说我们还没有做实验之前，组与组之间已经存在差异，即控制组与实验组之间存在差异。如上大学的人和不上大学的人之间就有差异，上大学的人能力比较强，工作比较努力，不上大学的人工作能力较弱，工作也不是很认真。假如你要测量这两组人，一组上了大学，一组没有上大学之后的生产能力的话，其实在他们都还没有上大学时，他们的能力差异就已经存在了，这就是产生观察性偏差的第一个来源。另一个来源是研究

中未观察到的反映在组与组之间存在差异，有的人反映大，有的人反映小，有的人得益多，有的人得益少。也就是结果变量 Y 与未观察到的变量 Z 之间有相互关系。在教育的例子中，有一部分人通过接受大学教育后提高得快，一部分提高得慢。也就是你提供的机会和条件是一样的，但是不同的人的反映不一样，同样的输入，可能有不同的输出，这是另外一个来源。这两个来源都会导致观察性偏差。一个是没有观察之前就已经存在的差异，另一个是对同一过程的反映存在差异。我们回到第二讲关于简单比较那部分。我讲过简单比较要有两个

假设，就是实验组对象与控制组对象大体上是相当的。一个假设是得到干预的实验组假如没有得到干预的话就和控制组没有得到干预的情况大体相当；另一个假设是没有得到干预的控制组假如得到了干预就会和得到干预的实验组的情况大体相当。第一个假设就是说实验组和控制组在起点上是相同的，在没有受到干预之前，实验组和控制组是大体相当的。有了这个假设，就没有了可能出现观察性偏差的第一个来源。而第二个假设是说如果控制组和实验组一样都受到干预后，实验组和控制组也是大体相当的，也就是在接受干预的过程中不会

产生不同的反映，这就没有了可能出现观察性偏差的第二个来源。这两个来源都没有了，就不会出现观察性偏差。接下来我们要讲实验，实验是指操纵某些变量来观察其它变量的作用的一种研究。实验是一种研究，这种研究的特征是我们可以通过人为的方法来操纵一些变量，后面我会讲到怎样操纵变量。我们先来回顾一下产生忽略变量偏差的两个条件，一个是有关条件，就是忽略的变量要与主要的因变量有关，第二个是相关条件，就是忽略的变量要与主要的自变量相关，只有这两个条件同时存在，才可能产生忽略变量偏差。而实验的关

键就是破坏产生忽略变量偏差的第二个条件——相关条件，就是使忽略的变量与自变量是相互独立的，这样我们就可以忽略这个变量。而在这里，第一个条件是无法破坏的，它是客观存在的。比如在我们讲的受高等教育与工作能力的例子中，你没有接受高等教育之前的能力水平，智力状况等等都是无法改变的，我们能够通过实验破坏的是第二个条件，就是使所有因素都和主要的自变量都成正交。讲实验的时候都要讲到内部有效性，内部有效性是指对于参与实验者而言实验所得的结论是有效的。假如存在这个有效性，那我们就说这个实验

有内部有效性，合适的实验设计都可以建立内部有效性。最简单的实验设计法就是“前测-后测”控制组设计。我们先不考虑其他因素，只是通过随机抽样的方法选取样本。对抽取的样本，我们先对其进行检测，就是前测，看他们在接受干预之前的状况。在他们接受了干预后，我们再对他们进行后测，看他们的变化。在这个设计中，前测并不是必需的，其设计主要是为了提高准确性，而并不是为了避免偏差。另外，对于实验组合控制组的协变量的测度和考虑也不是必须的，其目的都是为了提高准确性，在样本很大的时候就不需要了。

很多人可能会不理解，为什么我们在统计的时候要考虑多方面的因素。一个原因是如果我们忽略了一些变量，就有可能造成忽略变量偏差，我们得到的结论就有可能是错误的；另一种考虑是我们要提高精确度，在你做实验的时候，如果你考虑了其他的变量，你的实验结果的精确度就会提高。但是一般来讲，你只能包括实验前的变量，而不能包括实验以后产生的变量，因为包括了这些变量，就会包括进很多假设，你得出的结论就是有前提假设的，这种结论就有可能是不可靠的。因此如果包括其他变量与不包括其他变量得到的结果不一样

的话，你应该相信那个更简单的模型，因为它的假设条件少，结论更可靠。

讲完了内部有效性我们就要讲外部有效性。研究有可能是内部有效的，却是外部无效的。外部有效性要求在对结论一般化的时候，提出这样一个问题，就是你的实验结果在多大程度上是普遍适用的，它有多大的普遍意义。你不知道你的实验结果在实验对象范围之外是否适用。比如你在大学生群体中做的实验，你就不知道你的实验结果在没有上过大学的人中间适用不适用。许多因素都会威胁到外部有效性或代表性。一个因素是你所做实验的对象不能代表实验

对象以外的总体，这就会产生样本选择性偏差。第二个原因是你做的实验得到的结果在小范围是适用的，在大范围就不适用，我在后面会给你们一个例子。

那么区别内部有效性和外部有效性是不是有价值呢？这种区分是一种传统方法上的区分，但这种区分是不是有价值呢？我们一直是认可这种区分的，但是绝大多数经济学家认为这是没有意义的。因为评估研究的最终目的是为了对实际项目的有效性作因果推理，如果研究是外部无效的，那么它就是毫无意义的，因此假如一个研究没有外部有效性的话，这个研究本身就没有意义。所以

你的研究出发点就应该是建立在这个项目是不是有效，而不是说这个实验是不是有效。那为什么这么多年来这两者的区分会一直存在呢？这来源于心理学。为了证明小样本实验的有效性，心理学家普遍接受这两者的区别。心理学家的样本基本上都是小样本，虽然他们不敢保证他们的研究结果在整个研究总体中都是成立的，但是他们想证实他们的结论在他们所抽取的小样本中是成立的。因此他们就把内部有效性和外部有效性区分开来。但是很多经济学家对此是持批评态度的。我希望你们也能培养一种批评的态度。

下面我们转入下一个问题，讲一下准实验设计，准实验设计使用的范围很广，没有固定的模式，也没有固定的方法，有一些比较接近于实验法，有一些是用于特殊解释的多元回归，这只是一个标签。准实验设计是自然发生的，是在自然环境下发生的，不是人为强加一些变量进行控制的，研究者可以将类似于实验设计的方法应用到他们收集数据的过程中来。因为准实验设计是来自于自然环境，来自于观察数据，他的内部有效性就可能受损，严格来讲，准实验设计的内部有效性都不是完全的，都有可能不成立，你只能知道他的结果的可

信性，你不知道他完全的可靠性。但是因为准实验设计的数据来自于自然环境，来自于实际社会，因此他的外部有效性得到加强。准实验方法其实是通过统计控制或结构方法来实现的。用观察数据时，我们不能通过随机指定来控制误差，这种误差就会影响到准实验设计的内部有效性。那么我们就通过多元分析来控制误差。我们把可能忽略的变量考虑到统计模型中，对它们进行测量。比如美国社会存在这样一个现象，就是老师的工资比较低，而且女老师的工资要比男老师的工资低。很多人就认为这是一种男女不平等。有一个研究就对以

下变量进行了多元分析，一个是性别，一个是个人能力，另一个是课程。这个研究就发现，男老师因为个人能力比较强，他们教的课程多是理工类的，而女老师多教一些语言艺术类的课，而教理工类课程的老师的工资是要比教语言类课程的老师高的，因此男老师的工资比女老师高。这里就是能力和课程造成了男女老师工资的差异，但是造成这种差异的原因中可能还有其他我们没有考虑到的因素，比如性别歧视，但是性别歧视造成的差异不会大于现有的差异状况，而只会小于这个量。

我们再来考虑一下实验法，实验法

的创始人就认为任何研究结论都要通过实验来验证。他当时还不承认吸烟对人的健康有害，他认为已有的研究结论都是建立在观察数据上，而观察得到的数据是可能有选择性偏差的，是不可靠的。但是我们现在还是肯定吸烟和导致肺癌是有直接的因果关系的，因为假设这其中有选择性偏差，那这种选择性偏差会极大的偏离其实际可能的数值，而这么大的偏差是不可能由人的行为造成的。因此我们现在还是认为，虽然吸烟和得肺癌之间的因果关系是由观察数据得出来的，但我们现在还是相信这种因果关系。因为我们无法通过实验法来解

决这个问题，我们只能依靠观察数据。实验法解决了选择性问题，这种方法又称为简单模式。下面我给大家举一个例子，我所住的城市旁边有一个城市叫 Ypsilanti，这是一个比较贫困的城市。研究者在这里用实验的方法进行了一个学前教育的研究。一组人接受了学前教育，一组人没有接受这种学前教育。通过追踪调查发现，27年后，这两组人之间产生了很大的差异。接受过学前教育的人，他们的行为、收入等各方面表现都要比没有接受学前教育的人要好。这是比较著名的一个实验，这个实验现在还在继续。

我下面给你们介绍一篇文章，作者是 Manski and Garfinkel。其中 Manski 是我的老师，他是一位经济学家，是经济统计学方面的专家。这本书的名字叫“evaluating welfare and training program”。在这本书的绪言中，他们提出实验设计有一个经常被忽视的缺陷，就是通常研究者不能将实验背景下取得的结果推广到自然环境中，你的结果在实验环境下可能是好的，离开了这个环境就有可能不适用，因此 Manski and Garfinkel 对实验法提出了公开的批评。事实上，简化模式的实验评估需要特殊的带有推测的结构性假设。这个假设就是个体与组织

对项目做出反应回答时，实验的答案与真实的答案要是相同的。如果你不做这个假设，你的实验是徒劳的，实验结果根本没有实际作用。而这个假设在社会科学中是很难成立的，在实验中有效的结论，在实际生活中未必有效。我可以给你们举一个例子，在底特律进行过一个调查，调查的是问卷的回答率。研究者想知道怎样才能提高回答率。他们操作时采用了这样的方法，当被访者不愿意回答时，他们提出给一定的报酬，而且这个报酬的数量还随着被访者不愿回答的次数而增加。的确，收到钱的人的回答率要高于未收到钱的人。但是这个

实验只是局限于一定群体，如果每个人都知道了这件事情后，你们再去做这个调查的话，就实施不了了。所以，很多情况下，你的理论可能是很好的，在真正实行的时候就会有一个反效应，尤其是在研究政策的时候更要注意这个情况。一个政策可能在小地区之内是很好的，但是当它被推广到全国时会是一个什么情况，会不会产生动荡，这都是很大的问题。所以你要考虑到整体，当你把结论从实验的环境推广到实际的环境时，情况就有可能发生变化。因此 Manski and Garfinkel 提出来的问题就是一个缺乏外部性的问题，是外部性受损的问题。

关于前面这部分大家有没有什么问题？

问题一：经济学家经常用到思想实验的方法，我想问一下思想实验的方法有什么缺陷性？

谢教授：我们在做研究的时候，先要建立一个理论框架，而这种理论框架的建立就需要思想实验。另外当实验结果出来以后，你无法预测你的实验结果推广以后会出现什么情况，这时你也需要思想实验。但是这种思想实验还不是一个经验研究，还只是一个逻辑层面的思维。

问题二：谢教授，我想问一下是不是在一种极端情况下，内部有效性就意

意味着外部无效性？

谢教授：不一定，过分的强调内部有效性确实会损坏外部有效性，但并不是说内部有效性会完全否定外部无效性。好我接着讲，Manski and Garfinkel 就建议以结构法作为起点，结构法是通过观察数据对因果关系建立模型的一种统计方法。这种方法是通过收集数据和理论上的因果关系来建立模型的。我们还是举启蒙教育这个例子，在这个例子中，我们就控制家庭的社会地位，父母对孩子的参与，家庭关系是否和睦，社区的关系，学校的好坏等等。我把能够考虑到的因素尽量考虑进去，建立一个

比较完整的统计模型，这就是结构法。

下面我再给你们几个定义，一组是外生变量和内生变量。外生变量是指那些只能在方程中作为自变量的那些变量；内生变量是指那些在某些方程中作因变量，可能在其他方程中作自变量的变量。另外一组概念是结构方程和简化方程，结构方程是将内生变量作为自变量得到的理论性方程；简化方程是所有自变量都是外生变量的方程，即在简化方程中忽略了那些中间变量。我们来比较一下这两种方法，结构式方法的优点是：

第一，它是在自然条件下得到的，

一些结果直接用于总体，相反，实验法得到的结果必须要推广。第二，它是建立在理论之上，但同时又可以检验理论，相反，简化法只是对简单问题的简单回答，它不需要理论，它不回答有效性的问题，这种方法在科学上是有不足的，但是很实用。

简化法的优点在于：第一，在简化法中，内生偏差由于随机指定而被抵消，可以建立很好的内在有效性，而实验法在这里是有缺陷的。举一个例子，我们指定一部分人去学计算机，另一部分人不学计算机，但结果往往和我们希望的不一样，那些被指定去学计算机的人有

一些对计算机不感兴趣，他们会想方设法去干别的事情。相反，那些被指定不能学计算机的人反而会通过种种途径去学习计算机。可见在这里，内生偏差不可能通过随机指定来消除，因为人的行为是有目的性的，为了达到这种目的，他们会充分发挥自己的主观能动性，内生偏差在这里会产生很大的影响，这样实验法就无效了。简化法的第二个优点是它只需要很少的假设，得到的结论更加简单；第三，它并不需要复杂的统计模型，因此公众与政府官员很容易理解。

我下面要讲到的东西技术性比较强，可能花的时间会比较多，前面讲的

有没有什么问题？

问题一：在生物学里，有一种小规模实验的方法，您认为这种方法能否在社会学中得到应用？

谢教授：我现在所知道的社会学的实验方法有两种，一种是小范围的实验，通过控制很多外部因素来实现，但这种方法不容易得到推广。还一种方法是大范围的，分层次，长时段的实验方法，这种实验花费很大，也比较难控制。这些方法采用生物学的模式是不大可行的，因为你无法保证调查对象的稳定性。

问题二：谢教授，我想问一下结构法是不是不存在内部有效性和外部有效

性之分？

谢教授：结构法不谈内部有效性和外部有效性的问题，他的结果要么不成立，成立的话就是有外部有效性，因此他不区分内部有效性和外部有效性。

接下来我就讲几个研究设计方法。我先讲准实验设计方法，我会给你们举一些例子。一种是利用空间差异进行准实验设计，一种是利用时间差异进行准实验设计，还有一个是聚类设计。讲完这些之后，我再讲一下固定效应模型，然后再讲工具变量估计，这是结构法的一种特殊情况。

第一个是利用空间差异的准实验设

计，在美国有这样—个实验，某—特定的政策在 A 州中实施，但没有在 B 州中实施。比如在一个州中把最低工资提高了，在另—个州没有提高，而且在其他方面，A 州和 B 州是可比的，然后观察结果变量在 A 州与 B 州之间的差异。有些经济学家以为通过提供贫困群体的最低工资可以改善他们的生活状况。另外有些经济学家则认为提高最低工资水平反而会害了那些较贫困的人，因为最低工资提高后，物价也跟着上涨，物价上涨就导致需求下降，需求下降又导致失业增多，这样造成了—个循环，结果还是贫困人群遭殃。但是研究结果表明提

高了最低工资后他们的生活状况并没有发生什么变化，物价也没有怎么上涨，失业率也没有提高，这和理论上是不太吻合的。还有一个例子是一位经济学家提出一个市场转型模型，这一模型认为随着经济改革的深化，市场会出现转型，而在这个转型过程中，教育回报率是随着市场经济的发达程度的提高而增大的。我和我以前的一个学生就做了一个研究，按照前面的模型，在中国经济改革的步伐在地区之间存在差异，教育回报率在地区之间的差异与经济改革深度在地区之间的差异就会是相联系的，经济越发达，教育回报率就越高。但是我

们的研究结果表明情况恰恰是相反的，市场经济越发达的地区，教育回报率越低，而市场经济相对不发达的地区，教育回报率反而越高。这是我们利用空间差异做的实验设计。

下面我们讲利用时间差异做的准实验设计，比如我们想研究种族差异的持续缩小是否存在显著性，特别是民权运动以来，我们希望能够看到种族差异的缩小，我们就可以检验不同种族的社会政治经济状况在时间上的变化，是不是种族差异减小了。另外在检验教育成果的时候，我们也可以用时间差异的准实验设计，比如我们要检验新的教学方

法是否有效，我们就可以收集不同时点的教育水平，教育状况的数据，然后做一个纵向分析，就可以得出结论。虽然你不可能肯定你的因果关系的正确性，但是你可以增加它的可信度。

我们下面讲固定效应模型的例子，这个是一个血缘模型，一个人的成长，成功都和他的家庭背景，社会环境有关，这里就有两种处理方法，一种是把所有相关的因素都考虑进去。第二种方法是把所有因素中共享的部分固定，解除对这部分因素的控制，这就是固定效应模型。固定效应模型有一个假设，就是你观察的层次要比你考虑的层次要低。比

如你考虑家庭层次，你要观察的就是家庭中的个人。如果你的理论是在家庭层面上，而你的观察的层面也在家庭的层面上的话，你就没有自由度。固定效应模型是给每一个分析单位一个参数，现在假设有 1000 个家庭，每个家庭有两个兄弟，我们的观察点一共有 2000 个。而实际上我们用的参数只有 1000 个，所以这里我们浪费了 1000 个观察资料，但是如果你只从家庭层面上抽取 1000 个样本进行研究的话，你就没有自由度了。固定效应方法的性质是它控制了所有共同的特征。这个性质的好处就是，在固定效应层次上，没有观察到的异质性也

被控制住了。但是因为它控制了所有共同特征，这也使它浪费了许多信息。在上面这个例子中，我们就浪费了一半的信息。这里有没有什么问题？

问题一：请问一下，您说在固定效应层次上，没有观察到的异质性也被控制住了，这一点是怎么得出来的？

谢教授：用这个模型可以保证把所有观察到的共同特征和没有被观察到的不同特征都控制住了，这就是他的特点，其主要目的就是把观察不到的特征控制住。比如有些东西，像父母对孩子的感情等是无法观察的，我们就通过固定效应模型把他们都控制住。

问题二：您为什么说数据会损失一半？

谢教授：因为我们原来在个人层面上有 2000 个数据点，但我们考察的是家庭层面的，所以我们用到的只是 1000 个参数，这就浪费了 1000 个数据点，因此信息损失一半。但我们之所以要做 2000 个数据点然后再损失这一半的信息，这是为了保证研究的自由度。

最后一个方法是工具性变量方法，现在很多领域都考虑到这种方法，但是各个领域对它的使用不一样。这种方法的条件是工具性变量 Z 除非通过 X ，否则不会作用于 Y ， Z 这个信息就是我们

强加的。统计就是处理信息，当我们没有信息时，我们有两种办法，一种是去收集信息，另一种是去制造信息。什么是假设，假设就是制造信息。有了这个 Z 后，就意味着 Z 与 Y 相关，但不直接作用于 Y ，如果你找的变量与 Y 不相关的话，这个变量就是无效的，另外 Z 与 X 也要是相关的，但这种相关关系不能太强，关系太强了的话，这个变量就也不是工具性的变量。所以这个变量 Z 既要与 Y 相关，又不能直接作用于 Y ，与 X 也要相关，但相关关系又不能太强。只有这样 Z 在这里起到的作用才是工具的作用。因此这里最关键的问题就在于

找到一个好的 Z 非常困难。我给你们举一个例子，比如说你哥哥上大学的愿望对你上大学的愿望会有影响，而你哥哥上大学的愿望又是受到他的同学的上大学的愿望的影响，因此你上大学的愿望是受到你哥哥的同学上大学的愿望影响的，只是这种影响是通过你哥哥的上大学的愿望来实现的，在这里，你哥哥上大学的愿望就是起到了工具性变量的作用。大家明白了吧，好，我今天就讲到这里。

马戎主任：今天的讲座就到这里，星期五谢宇教授会给我们带来他的最后一个讲座，主题是美国社会和美国社会

学，地点在理教 207，欢迎大家参加。

发表时间：2004-12-9 12:58:00