EM114026大数据计量经济分析2025年10月9日第6周星期四第7-8节-pptVideo\_原文

2026年01月04日 13:04





发言人 00:12

产品。



发言人 00:49

跟你说话。



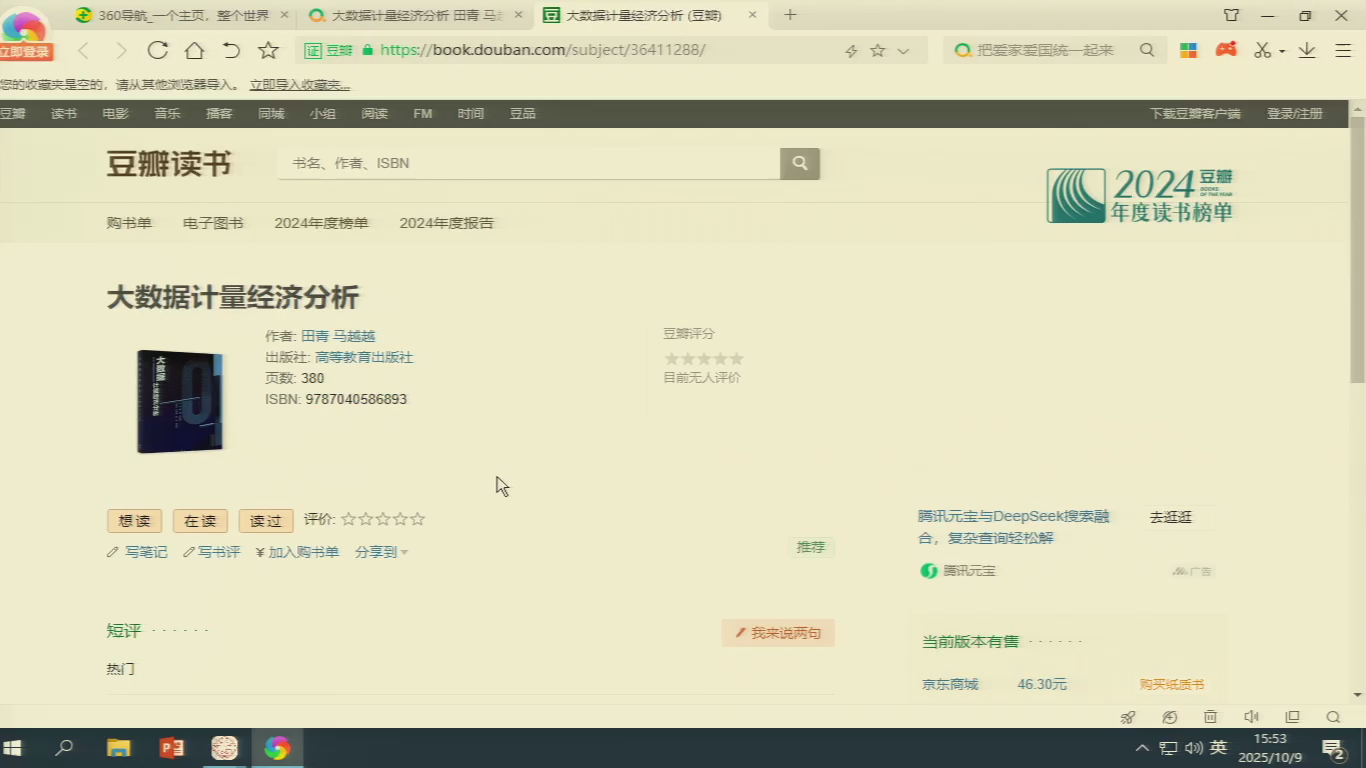






发言人 03:39

这个学期我们上的这个大数据计量经济分析，用的是叫蓝色封面的这一本田青马月圆边。大家都可以把这本书这样子购买。你拍一下这个应该京东上面都有。



发言人 05:52

好的，那我们继续我们这个大数据计量经济分析导论课的学习。在上一节课，我们通过一些例子给大家大概讲解了为什么要学习这样一门课程。然后也让大家带入了一些实际的这样一些经济的问题。比如说什么你作为一个校长，要不要去扩增老师，缩小班级规模。或者说是啊对于这样一个家用汽车市场的这种理论进业界工作的，就是可以做实际的这样一些问题的分析，包括也有股票经济问题的这样一些，我们怎么样用我们这个什么实证研究的这样一种思维的方式去解决你可能碰到的一些问题。

发言人 06:50

那这些问题有一些什么实际的，或者说是啊我们一般性的这种共性的问题。比如说我们可以先提出所研究经济问题以及这个问题的度量方式，然后分析它的主要因素。然后接着我们看不同因素之间的互相关系，对吧？然后再探究这些关系的数量形式或者是数量规律，然后我们再分析和检验这样一种数量规律的可靠性。最后我们用这样一个可靠的数量形式，或者说我们测量出来的这样一种规律去对未来做预测。这个就是学习我们这门课程这样一些例子，我们可以有这样一些共性的思考的方式。

发言人 07:48

我们在上一节课也给大家讲了一些什么，我们计量经济学基本的定义、分类、性质，还有它和其他学科之间的一些关系。比如说计量经济，我们肯定还是在经济学的范畴，那这样一门学科它也不是数学，它也不是统计学，它是什么？这三者融合统一的部分，是我们这门课程涉猎的这样一个范畴。包括我们上节课给大家介绍了它与其他学科存在的一些区别。

发言人 08:24

最重要的或者说我们延续下来，大家学习经济学的一些传统的经济学的科目里面，大家可以也学到一些什么经济学的这样一些基本的规律或者定理。比如说收入假说，这样是我们把这个收入的假说怎么样放到一个量化的考虑范畴的时候。这个例子就是一个非常典例子。我们怎么样把一个经济理论过渡到一个实证研究的这样一个什么一个方程，或者说一个模型，对吧？那上节课我们也让大家看了这两个式子存在一个最大的区别是什么？这就是我们上节课结尾很明显可以看到，等式的左边是一样的，对吧？

发言人 09:13

也就我们关注的这个Y这个Y可能是什么？比如说西电学生的消费情况，或者说西电学生的什么各种学习成绩、绩效表现，或者说其他一些大家想要关注的问题对吧？或者说西电学生对这个新能源汽车的购买意愿？我们提点学生在外租房意愿，各种问题都刻画在我们本次的左边。

发言人 09:38

大家可以看到它跟我们经济学的这样一个理论它是一样的。区别在于什么？区别在于我们这个计量经济学的方程后面有一个尾巴，他加了一个尾缀。这个尾缀是什么？这个就是这个mut对吧？这个mut是什么意思呢？就相当于我们这门学科我们要关注的，我们关注现实的问题是想，然后我们把这是现实发生的事情，我们把它抽象成一个模型，或者说是一个方程。

发言人 10:16

那你把现实的问题抽象出来的时候，我们会发现西电学生购买新能源汽车的意愿或西电学生的消费倾向，他受很多很多因素的影响，对吧？他不一定受不一定只受大家这个生活费的影响，他也可能受什么。比如说天气下了一个礼拜的雨，大家也没办法出去消费，或者是什么大家新交的一些朋友参加一些社团，或者说还有一些什么地铁开通，还有一些我们这个什么教学课程排课。如果大家时间很晚，你也没有时间去消费。就类似这样一些很多杂七杂八的不同种的因素，或者说现实中影响这个Y的是左边的这个现象的问题。或者说这些因素我们全部要把它刻画到我们这个模型里面。因为我们是用这样一种实证的思维去解决现实的问题。

发言人 11:16

但是你又不可能，比如说我把什么天气情况，下雨不下雨，然后大家这个礼拜的课程是多少节课，然后大家这什么有没有新加入什么社团，大家有没有新交朋友，有没有谈对象，有没有什么你的性格对吧？或者说你家里有几个姊妹，就是各种你不可能说把所有的因素，我这一个黑板可能都列不完，对吧？包括上上一节课的同学还提到，我可能头发长了也会影响我的消费，我还要去剪头发，理发的这个也在我的支出考虑范围内。

发言人 11:55

如果说我们实际的这样一种问题的刻画，我们用方程或者模型的形式全刻画出来，那就可能是Y等于X就是大家关注的这样一个比较重要的因素。那后面还有什么X1、X2、X3、XXX可能有X200对吧？可能就是黑板都起步，那我们怎么样去把它抽象但是又很重点的去表示出来呢？我们就可以用我们这个PPT上展示这种形式，后面加以后缀mut这个mute是表示什么？你不关注的，或者说是对这个Y影响比较细微的其他一些因素，我把它们打包全放在这个mut里面。它作为一个作为一个随机的录像。这个时候我把它打包汇总起来，大家就可以看到它相当于什么X2加X200所有的影响，我把它放在一起对吧？那这些影响里面大家可以想象到有的是正的，有的是负的对吧？

发言人 13:05

那这些影响因素加合在一起，可能有的影响因素或者说是正负抵消，它对于我们整个市值的影响可能就是一个产一个随机的状态。这个就是最理想的我们刻画的这样一种因素。这个是我们这个数理经济学方程，大家从前学的一些经济学理论，它表达出来都是这样。数学数理经济这样一个方程过渡到我们这个计量经济学模型里面，它最大的区别我们把所有的因素都考虑在内。那些什么微小或者说是啊不确定的，或者说是啊影响比较细微，你又不怎么关注的，这样一些因素全打包放在mut UT，就是作为我们这个什么随机脑动向，放在我们什么计量经济学模型里面看这样子所有其他危险因素打包起来对这个马乖的影响。那实际上，如果我们考虑周全，或者说一个理想状况的假设，大家就可以想象到这些所有其他微小的打包起来，正负抵消它最理想的状态，是不是BUT就对我们这个什么，但是左边的影像是微乎其微的对吧？那这个时候我们就可以很很好的去刻画贝塔和阿尔法，他俩的估值就可以代表我们什么关注的这样一个影响的量化的这样一个分析。是上一节课讲完了什么，我们降低一些定义，有一些例子，下面我们来看一下，降低一下它的起源和发展。

发言人 14:50

所以它起源于什么很古老的什么亚当斯密拉在国富论里面，他他的发表是我们这个社会科学的这样一个诞生。什么弗里西，它仿造生物计量的悬疑词，从而提出了计量经济学。这样一个提出，它就根据我们这样一个计量经济学学科的这样一个诞生。在1930年才出版，用完全回归体系的统计河流分析，就进一步进一步沉淀了这样的这样经济学，它的这样一个量化的分析的这样有技术。

发言人 15:34

然后在世界经济大萧条之后我们也出现了一种什么凯恩斯革命。对新古典经济学它是存在一种否定的，然后也使经济学家更加的去重视什么，我们一些微观或者说是经济现实的这样一些问题，那就产生了什么以问题为导向的这样一种研究范式，然后开辟了这样经济学作为实证研究主要方法论的这样一个发展和应用的空间。它产生的例子大家就可以看到它是什么，对一些经济问题定量研究。所以在我们这门课程的开始，我给大家讲了很多不同的例子，对吧？大家扮演不同的角色，它实际上是一个非常实用性的，或者说是啊以问题为导向的这样一种科学或者说研究的范畴。那他在什么通过这样一种生物计量的演变的词，从而形成我们什么计量经济学的这样一种这样这样一种这样一种名称。它的特点最主要的是什么？它并没有什么固定的经济理论。

发言人 16:53

它其实是大家可以看到它三个学科的交叉，是我们这个学科的主要关注的内容。它产生的这样一种意义，也是从定性的研究到定量研究的这样一种发展。它也是什么？我们经济学更精密更科学的这样一种表现，也是我们现在经济学这样一个基础。

发言人 17:19

随着这个计量经济学的发展，现在也演变了有有一些不同的这样一种方向和趋势，从前这样拎进去，因为可能大家收集到的或者说掌握的什么信息比较少，那很多都是手算或者是手动推的那现在随着计算机的应用，很多结合到一起的。比如说这样力学后面再发展出来，它也是从统计学衍生出来，比如说像什么一些神经网络，一些什么大模型，它也跟这个计量经济学有一些交叉结合的这样一种方式。模型的变量和方程也趋于从少到多，又趋向于较少有这样一个趋势。

发言人 18:14

就从前可能就是很简单的一些模型来刻画，比如说我们这个收入的问题。那你要进步，或者说你要去在前人的什么基础上再改进的话，你很多时候就考虑一些别人没有考虑的因素，不断的加到你这个模型。比如说刚开始人家只考虑X1，那你在前人的肩膀上，站在巨人肩膀上，你要再往前走一步的话，你可能再考虑X2、X3、X4、X5，对吧？那它就存在这个健康经济学发展，它就存在于什么？由少到多。那你多到大家可以想到刚刚老师给大家讲过对吧？那一个模型如果你真的要去把所有因素刻画出来的话，比如说200个因素你做一个黑板都写不下去，那它实际上违背了我们这个计量经济学的初衷，对吧？那它由最后他又由多向少有这样一个趋势是吗？

发言人 19:11

你还是要把它什么大道至简，你要把这个最精髓最精华的这样的规律性的特征提取分，所以它最后也是趋向于较少，或者说是多个模型归并为整体模型的这样一种研究范式。这个应用领域的拓展，它也是在方方面面的。它不仅是在宏观或微观经济领域进行应用，它也由这个预测为主转向了更多对经济理论假设和政策假设这样一种检验。理论方法它也有一些新的突破。比如说简单的可能开始的一些线性的模型里面。之后它有一些非线性的非参数的半参数模型和动态模型，时间序列模型，写真理论，还有panel data这样一些数据模型，以及空间计量模型，这样一些新的研究领域有一些突破。

发言人 20:19

计量经济学的发展，它核心的内容，特别我们这门课程，也不会给大家讲的特别深。这个可能在之后，比如说中级计量、高级计量，可能最后老师会给大家来讲，比如说包括一些非线性的，或者说因果，感恩节，或者说是其他的一些更深层的计量的内容。在我们这个初级的这样一个计量定义课程里面，我们只给大家讲比较经典，就是只考虑线性的这样一些回归内容的拓展性的或者说是非线性的内容，会在高阶的课程里面会给大家讲。

发言人 20:57

在经典的这样一个线性里面去考虑学会授权这个线性，我们会有一些在这个模型里面会有一些基本的奖励，这个也会在后续的课程里面详细的去给大家介绍。这些奖励大家可以先做一个了解，那些奖励包括什么？首先这个模型是线性的，也就它不是指数型的，不是不是什么对数型的，不是其他一些形式的这样一些方程。那他也不存在多重贡献性，多重贡献性之后我们会给大家做一些详细的介绍。还有自变量和误差项不相关。这个误差项就是我们刚刚讲的，我们把所有其他微小的，或者说是不是我们重点考虑的因素打包的那一项，随它也它叫随机扰动项，也叫误差项。有的时候如果你打包的不好，你把一些重要的因素也扔在里面了。

发言人 21:56

它比如说有一些正向影响，有一些负向的影响，它最后加总起来，打包起来，它的影响是不能被忽略的，也就是它是区别于零的那这个时候他就有可能什么跟你重点关注的X是相关的那这样相关的大家就可以看到，你就不能把这个什么X的影响完全提取出来。因为它有一部分还受它它跟mut他们之间的相关的影响。第四个是误差项平方差，并且误差项之间不相关。第五项是误差项要服从正态分布。

发言人 22:37

只有满足这样一些经典渐渐回归模型的基本假设，我们估计出来的什么主要参数的，我们这个叫贝塔，就是X对Y影响的这个量化的规律，它才有一系列优良的统计特征。你看它什么，我们估计出来这个参数才是可靠的。但是大家可以想象到现实当中大多数的我们这些经济的数据，都不满足这些假设，或者说都不能够完全满足这些假设。我们这门学科讲起来，其实都是围绕这五个相应的假设展开的当不满足假设一的时候，我们要怎么样去处理的，用什么样的办法能够缓解，当不满足假设2的时候，我们应该采取什么样的方法或什么样的检验，尤其是我们要怎么样去识别出来，我们收集到的数据是不是还存在不满足假设一的情况，不满足假设二的情况。这个是我们这门课程比较核心，围绕着这样一些内容。

发言人 23:57

随着什么我们现在可以得到的这个数据规模越来越庞大，经济观测的数据它其实是什么以指数增长快速增加，那怎么样我们来面对这样一些什么大数据的这样一个挑战？大家可以看到传统的计量分析，它主要面对的是数字型结构化的数据。那大数据还包括大量的什么非结构化和半结构化的数据。比如说什么一些多模态的数据，比如说文本图形、音频视频，还有什么定位的数据，还有卫星灯光图片，以及比如说我们这个什么时空行程轨迹，比如说大家手机的这个码可能没有注意到。像我们从前也有同学就拿，从前读博的也有同学拿的是什么？大家这个联通或者移动，它能达到大家这个轨迹的数据。比如说像什么港中文一个老师，他当时疫情的时候，就是跟移动合作，他拿到了大家什么手机移动的轨迹的数据。他从而来预测什么新冠疫情的走向，还有包括做一些预测。

发言人 25:26

这个是买我们在这个大数据时代会包含或者说迎接的这样一种挑战。就是把我们这个数据形式非常丰富，就面临挑战不一样，像我读硕士的时候，从前可能基本上大家如果做这个计量分析或者说实证分析的话，拿到的是什么？比较少量的一些宏观的数据分析。比如说统计年鉴，公开的这样一些政府报告的数据，数量也比较少。当时我记得我还问我老师一个问题。

发言人 26:01

我说这个课本上写了我们要有一定量的数据规模，我们得到的数据什么估计结果才有可靠性，对吧？那个时候定书本里定义了什么什么叫大规模数据，或者说不够什么大样本数据。那个时候是30个，这30个数据大家可以放到现在来一下，就是很不可思议的对吧？因为现在大家可以拿到分析的数据规模都非常的庞大。那包括这样就包括还有不同的这样一种数据形式，比如说区间数据，什么函数数据，还有什么符号数据。

发言人 26:44

从前第一届的这个区间数据的这样一个计量分析，是在中科院开的这样一个国际会议，我当时也去参加了。他们就相当于是在计量数据的分析基础上，大家可以看到，比如说它这个区间数据是什么意思呢？就比如说大家想分析欺骗学生的这个消费情况，大家可以想象到我一般用的什么，你每个月支出的节点？那区间数据它包括什么？它有可能把这个数据分成你每个月单笔最高的消费，单笔最低的消费。那这个时候这个数据就不再是啊你关注的重点，就不再是你每个月的什么总消费的，他关注的是单笔消费的，或者说你消费的这个区间变动的范畴。那用这个区间来进行这样一个数据分析。这些都是如果大家感兴趣，在后续的研究过程中可以去做一些探索性的这样一些工作。

发言人 27:55

这样经济学，它其实也是蛮起源于国外。在国外，它首先是应用于什么微观经济的那在宏观经济理论出现之后，它也发展的比较快。在国内，它是从什么上世纪50年代末才有人开始研究过。很快这个旅游团大家都知道，直到70年代末我们才恢复有关的什么研究和学习。从80后80年代后才开始快速的发展。

发言人 28:36

在国内我们这个课程的定位，它就是我们教育部高等学校经济学学科教职委规定的核心课程之一是计量经济学。它这个是什么？在国外一些和他的什么发展当中，还有国内都是越来越重要的。在我们国家意识到这个经济自然经济学的重要性开始之后，大家可以看到，我们这个财经类的学校或者高等学校这个专业开放课程的比例，基本上到06年开始就全面覆盖了这样一门课程的教学。

发言人 29:23

那现在经济学研究或者我们这个经管大类的研究现状，如果说我们文章没有什么复杂的数理模型或者计量模型支撑的话，那基本上稍微好一点的这个杂志，基本上都是拒绝发表，或者说不能甚至不能进入审稿环节。那这个就是什么现在大家都说计量为王的，或者实证为王的这样一个。虽然现在因为这个时政文章太过量了，也有慢慢往查一些定性研究的去发展，但是还是微乎其微，就是非常庞大的这样一些研究范式，都是这个计量范畴和实践范畴的。

发言人 30:15

在国外的这样一些顶级刊物中它有一些比如说什么A12 review，这个就是最顶尖的经济学的这样一些刊物，还有什么APE，还有这个QGE，这些都是国外的非常顶的这样一些堪。那国内的刊物，比如说这个经济研究，这个金融研究，中国工业经济，它在它这些期刊上面，没有数据模型，或者没有实证的文章，它也是什么凤毛麟角的那在这个我以国内的这种期刊为例，比如说经济研究，大家可以看到08年以后，这个实证类型的文章，它比例基本上是高达90%左右。你说它是一个什么非常主流，非常绝对主流的这样一种研究方法。同样这些也是老师过去发表的一些文章。这些文章基本上也都是采用的这种实证分析的方法。

发言人 31:33

比如说第一个这样一个文章它是什么？这样一种数值型的数据和文本型的数据结合在一起，构建我们这个实证研究的这样一种范畴。除了可能就是第二个，它是一种纯理论推导或者纯博弈构建的这样有理论性的文章。其他基本上都是用实际的数据构建相应的实证分析模型，去进行的这样一些研究发表。那这种发表工作或者实证研究工作，他其实相较于理论性的工作，他发表的什么难度就要低一些，获得可能性要高一些。大家学习这么好计量经济学课程，也是什么我们将来走向社会实际工作的需要，对吧？

发言人 32:36

不管你进什么政界，你进业界，或者说你即使啥也不干当个自由职业者就想在家里炒股，对吧？你也要学习这样一门这些这样一门预测，或者说是去量化考虑问题的这样思维方式，对吧？不管是你用来分析这样的宏观经济数据，公司的销售额，经济的整体发展趋势，以及股票的价格走势。更现实的需要的是什么？我们学生的课程有助于什么？将来的读研，包括参加一些竞赛，急着什么出国深造。在国外，对于这门课程的定位，它实际上已经成为经济学课程里面最权威的一部分。

发言人 33:33

那一个统计数据来看，它也是最近这个落奖。虽然在落奖，大家可以看到我们从前确实是这样的，经济学获奖者里面大多都是什么？单从这数据就可以看到大多是什么？因为对计量经济学发展贡献获奖。当然现在很多不管是什么，甚至包括物理化学生物，他们都在买。因为贡献了A对AI贡献计算机的人买了是吧？

发言人 34:05

那近年来与计量有关的获奖情况，也是2000年很多。2003年格兰杰在你们的什么金融市场里面，全新的这样的经济时间序列分析。那也就是说什么？我们学习这门课程也有助于什么？有助于大家以后拿诺奖？也可以往这个方向去去去努力，甚至少你比学其他的经济课程这个可能性要高，对吧？

发言人 34:42

你从数据上来看是这样的那后面我们就给大家介绍，就和我们这个计算机去给大家初步了解了它这个发展、起源，包括它的定位。下面我们就实际来看，就在导论这门课程里面，我们给大家看一下这个大象到底怎么样，长什么样子？他的手脚，它的鼻子是什么样的那我们看一下这个计量分析或者计量研究还是存在哪样的一些步骤。其实如果简单概括来讲，它就包括。几个问题对吧？

发言人 35:23

你说我们计量经济学怎样去研究经济关系它的这样一种数量表达或数量关系，那他的研究方法是什么？前面我们是不是举了一些例子，也归纳出来的一些共性的问题。这几条我们来先看这个什么前面这个这样一种模型设定来说，你要选择变量和数学关系是你说你要确定好你想研究的X和Y你的变量是什么？这些变量之间它是线性的还是非线性的，还是指数的，还是什么，还是对数的这样一些不同形式。你说开始我们先进行这个模型的设定，我们确定变量间的数量关系。如果要估计X对Y的影响是哪，我估计贝塔，估计相应的参数估计出来参数，我们就检验我们得到结论的这样一个靠性，对吧？

发言人 36:32

就说对这个模型进行检验，检验完得到发现我们这个数量关系它是可靠的那这个时候我们是不是就用可以用这个可靠的这样一量化的关系去做经济分析和经济预测，就比如说大家都说黄金上涨跟什么美元下跌，还有它是强相关关系，吧？假设说你们构建一个模型去预测黄金价格的走势，那这个时候，你分析出来得到的一些关键的预测因素，比如说X1、X2、X3，甚至还有什么一些美联储降息，一些政策性的因素，怎么样对黄金价格进行影响。那你得到这样一个模型，并且检验出来GX对Y的影响大小的时候，是不是尤其是你还发现这个影响的系数它是什么？你通过检验之后发现它是可靠的那是不是还需要把这个模型拿来应用？就用来做经济分析和经济预测，那这个时候你可能就能大捞一笔了，对吧？因为如果说你得到这个模型是可靠，或者说你真的是抽取出来了它这个规律性的话，你可以拿来进行相应的预测的分析。

发言人 38:06

首先来看什么叫做模型的设定。我们对经济模型这样一个设定，它这个模型的概念，它其实是什么？对经济现象或者过程一种数学模拟。所以我们把什么实践或现实当中发生的事情，如果你用语言去表达的话，你可能就会说，今天下雨，今天我去商场购物，对吧？这个是现实发生的练习，如果你用文字表达是这样的，但是如果你想要去提取出一些规律来，你可能比如说大规模西店有多少学生，在什么样的规模下出去购物了，对吧？然后多大规模在学校里面，你要把它更抽象或更提取出一种规律？最好的方式是不是你用最简单的方式，我把它形象，或者说用一些简单的什么符号来代表你的一些行为，或者你的最后一些表现对吧？啊，所以这个模型它的含义对经济现象或过程的一种数学模拟。

发言人 39:26

那模型它只能抓什么主要的因素和主要的特征，不得不舍弃某些因素。这个也是刚才给大家讲的，有一些因素你也能捕捉到，你甚至你也有数据。但是我们去刻画模型的时候，你不能事无巨细的。比如说老师我真的很用功，对吧？我也我也对这个问题做了非常深入的探讨。我对什么提炼学生消费情况，做了一个底朝天的摸底调查。然后我也发现了200个因素能影响气垫学生的什么消费支出的情况。你咔咔咔你不能把什么模型里面200因素全部堆起来，对吧？

发言人 40:14

那这个时候为什么我们不能这样做呢？就是我们计量经济还是要什么要抽取一些规律，如果说你把200个同时放在一起，它的什么是不是就事无巨细了，你就没办法突出一些重点刻画的内容。就是说这200个因素也会影响Y第190度也会影响Y就没有重点了。所以在我们刻画这样一些模型的时候，我们只抓主要因素的主要特征，那对所研究经济变量之间的关系也只能选用适当的数学关系，数学关系式近似的简化的表达出来。

发言人 41:04

假设说大家做了底朝天的什么对西电学生消费支出情况做了这样一个调研。大家又拿到了很多很强势的这样一些实际数据，然后从数据去刻画对他们之间关系的。你发现老师这个X对Y的影响，这个画出来的图它是一个什么？它是一个对数的形式，或对数的形式再加3次方的形式。X2对Y的影响它可能是一个什么指数的形式，再加上一些倾斜变种什么之类的那大家可以想象到，假设说我只研究一个什么问题，大家都觉得什么生活费，大家讲的都是生活费。生活费对你这个消费支出的影响的话，如果你刻画出来是一个什么指数加3次方或者加4次方，加5次方的这样一种形式。虽然你们非常完美的确定和大家收集到什么西电学生消费的情况，大家但是你可以发可以发现什么，它的规律性是不是就被你复杂化了，你没办法去做。

发言人 42:22

68当爸妈给你多涨100块钱的生活费的话，你的这个消费支出会增长一个什么指数级加上30号的形式对吧？那在我们这个模型里面，很多时候你不需要很完美完全的去拟合它，更多的还是你要抽象出一种简化的规律。假设说你这个采用的线性的，就是最简单的形式，它也能7788的或者说是比较合理的去刻画出现实的情况。

发言人 43:04

那这个时候这种形式它就要优于你采用一个什么指数加对数，加3次方、加5次方的这样一个表达形式对吧？因为指数加现在加3次方加上5次方，你没办法表示出它的规律性来，对吧？或者是规律性太那个了。

发言人 43:27

那模型的设计和形式它的取舍实际上也存在一些主观性。假设来说在一些模型设计里面，比方说如果说我们这激烈消费情况用一个线性的能表达，那有的人他可能觉得我用分段的陈述也能更好的去表达。那这个时候它就存在一些主观性了。因为你线性其实能很简化的去把这个问题表达出来。有时候你把这个分段的这样一种方程，它虽然复杂一点，但是可能你觉得它能更好的去体现。比如说可能什么生活费超过5000，他的这些人的消费支出行为和和什么生活费不超过5000，他可能是两种不同的什么消费表达，对吧？比如说如果生活费超过5000，他可能妈妈再给他加生活费，他也没有什么可以消费的了。所以他的那个函数的表达式，就可能不会随着生活费的上涨而支出它的这个款，而增加它的这个消费支出。

发言人 44:39

它构成这个经济模型它的基本要素，大家可以看到，就是从我们刚刚那个例子里，其实大家可以看到它包括了几个部分的内容。首先是什么经济变量，这个经济变量它在不同时间，不同空间它的表现也是不同的，有取值不同。它是什么可以观测的因素，也是模型的这样一种研究对象或者说影响因素。有的时候可能你在考虑一个什么现实的研究问题的时候，你突发奇想觉得什么我这个因素从前没人考虑过，那这个时候我把它放到模型里是不是很很好。实际上我们还要考虑的一个重点，你加到模型里的或者说你考虑的什么因素，他需要是什么可以观测的，或者可以有数据来体现这个变量。有的有一些影响因素他没有办法被观测，那这个时候我们就很难把它作为一个什么惊喜变量。

发言人 45:56

那经济参数是什么呢？这个参数大家一看就知道它是什么，我们要估计的这样一种一种形式。它是用来表现经济变量相互依存度决定经济结构和特征这样一种相对稳定的因素。通常它就不能直接被观测，也就是说我可以观测到的量是吗？

发言人 46:26

大家比如说我现在发放一份问卷，给西边所有的群众填一下，你可以观测到数据是什么？你每个月的正常码消费支出是可以观测的对吧？你每个月的生活费是可以观测到？或者说你每个月出门的次数，这个都是可以观测的对吧？但是比如说我现在能不能发放一份问卷，让大家填你的这个生活费增加100块，你的这个消费出会增加多少，你就没办法去？这个值叫做什么？这个值就叫做经济变量之间相互依存的程度，对吧？它就变成了一个什么经济参数。

发言人 47:15

这个参数我们需要什么？通过对经济变量部件模型去估计出来，然后统计方法去把它估计出来。这个是你不能直接观测到，对吧？

发言人 47:31

随机扰动项也是构成我们这样经济完全基本，所以说我们刚刚讲的他是什么模型当中没有包含所有因素的代表。比如说我们模型里面等号左边是Y对吧？然后右边有时候有一个截距，这样有时候没有然后就是XX1X2、X3。随机导动向一般用缪或者you来表示，那这个谬它就是除了X1、X2、X3之外，所有其他的因素就都加在这个表里面。

发言人 48:17

那个设定我们这样一个计量经济学模型，它也有一些基本的要求。你说怎么样来设这个模型是Y等于阿尔法加贝塔X1加6呢？那这个夕阳经济学模型的设定，它是需要有科学的理论依据。这理论依据往往来源于这个经济理论。比如说收入上涨会影响消费支出，吧？这个是你要有相应的理论依据，或者说你什么GDP增长受什么贸易量的影响，比如说WTO或者说有一些正常，这些都是需要有相应的理论依据，你才能适应设定相应的模型。

发言人 49:09

其次要选择适当的数学形式。有一些类型，比如说单一的方程，联立的方程，你需要去确定，有些是线性的形式，有些是非线性的形式。此外我们这个什么模型它也要有所取舍，也就是说我们需要这个什么模型要兼顾真实性和实用性。

发言人 49:42

这里我们提出来有两种都是不太好的这种模型。一种就是你构建的那个模型太真实了，对吧？太真实了也不好。你比如说你就变成了一个什么类似于高像素的照相机，这样照进去的结果，他这个模型显然就会什么过分的复杂。你这个数据很多时候也没办法收集，对吧？那你这个计量的方法肯定也会非常的复杂。甚至你得到这种结果，它的分布性质你也没办法判断。但实际上这种模型是很不好。也就是说在我们这个模型构建的时候，不是说越复杂的模型就越好。

发言人 50:35

那另外一种不好的模型就是什么过分的简单了，你说简单的话，你们不要太复杂，那我就搞最简单的对吧？就像我们刚刚那个收入跟消费里面，那就是最简单的模型。比如说大家要研究新电商学生消费支出的问题，你就只收集了一个什么大家这个收入的这样一个因素。因为你太简单了，或者说你省略掉了一些，或者舍弃掉了一些重要的因影响因素，那你这个模型是不是也说明不了什么问题了。就比如说你最后收集了西电学生的这个收入情况和消费支出的情况。最后大家可以可想而知，你只能得到的结论是吧？得到的结论是大家都早都知道的这样一个结论，收入会影响消费支出？

发言人 51:30

那你构建这个模型太简单了，是不是就也不行，第三，我们这个模型尤其是我们一直在强调的，就是我们刻画现实，或者说我们这个实证研究里最重要的一项，你一定不要把这个什么随机误差项给漏了。这个随机误差项，是我们这什么经济模型和计量模型，它最最重要的这样一种区别。最后我们再构建这个什么。这样一个模型，它里面的变量必须一定要是可以观测的。或者我们这个变量代表了现实这样的一个经济运行的过程，它需要能够获取数据，这样你才能用到模型里面来。如果你那个变量就不能观测，也不能获取数据的话，那他放到这个模型来是不是你也没有用能起啥作用，对吧？这个是也是需要注意的这样一个问题。

发言人 52:43

大家根据研究里面第二个步骤就是估计参数。那我们为什么要对这个参数进行估计呢？这个因为我们不知道，对吧？所以而一般来说参数它都是事先不知道是未知的那它又不能直接被观测，所以经济关系又有一定的这种随机性。

发言人 53:13

比如说在我们这个模型里面，刚刚说有一个缪？这样一个随机扰动项。那这样就随机扰动项，它就使得我们的这个参数，你看前面那个模型因为存在着这样一个缪，如果没有谬，我们解这个方程是不是就可以直接把这个阿尔法贝塔求解出来，但是因为有这个随机变量缪缪它是随机的结果，这个贝塔它又是什么？用样本的这样一个观测值对它进行估计的那就使得这个什么贝塔它得到的估计结果，也就成为了一个随机的变量。也就是说这个参数它是一个随机变量。那这个时候它就不能精确的去加以计算了，就只能通过样本去估计。

发言人 54:11

通过样本去估计，大家就会想到了这种估计，他就又涉及到估计方法的问题。那它参数只能估计，不能去计算，或者不能直接去确定它。既然经济学这门课程，它的核心内容就是我们要去把千方百计的去得到一个可靠的这样一种估计的结果。它这个估计结果或者说这个参数它其实就是把我们想要探寻一种量化规律，对吧？那怎么样通过我们样本观测值去估计总体模型的参数，这个方法怎么样才是科学的，然后怎么样才是一个好的合理的估计。这个就是整个计量经济课程最核心的这样一个内容。回头看我们就如果说这个课程整个结束，大家回头看就会发现，我们整个课程其实都围绕这个参数的估计去展开的。

发言人 55:18

说你什么收入消费的，这样问怎么样才能得到一个合理的，比如说我们得到一个合理的可靠的结论，能说明比如说西电学生上涨这个什么消费，收入上涨100，你的这个消费支出会上涨多少。这个假如说是贝塔，那这个贝塔就比如说你这个什么收入上涨100块，你的这个消费支出增加75块。那这个七十5就是我们这门课程最核心的这样一个内容。怎么样确保这个75是合理的、可靠的，你你你咋就知道我这个多给100块会不会消费掉75块呢？

发言人 56:11

那两个概念需要做出一些区别，比如说我们后面会经常要用的，要明确的一个称为参数的估计值。英文就叫它是指的是什么？它指的是估计出来的这样一个具体的参数值。就比如说我们这个编辑的消费倾向是0.75这样一个值，它就是参数的估计值。那参数的估计式是什么呢？估计式然后怎么样得到的这样一个0.75的这样一个值，所以你这个0.75这个值是怎么计算出来的？他的英文也叫express，所以它俩的区别英文单词都是一样。但是区别就是参数的估计式是我怎么样去求得这样一个0.75这样一个估计值。

发言人 57:18

参数估计我们最常用的方法就下面这样一些，比如说普通最小二乘法，它的英文缩写就是OLS，英文它是个ordinary liner regression square，它就是OLS。普通最小毫升，广义最小高层就是GOLS。这两个是用的比较多的那在我们这门课程里面只要大家掌握这款普通最小二层就可以。其还有一些其他的估计方法，比如说极大自然估计两阶段最小二乘，三阶段最小二乘及其他。比如说据估计，在后面会给大家做介绍，但是掌握只需要大家掌握这个OLS。

发言人 58:21

第三个步骤是我们什么对模型要进行检验。就是说有的时候我们这个建模，它的理论依据可能不一定充分。那理论他可能如果说本来就不是合理的，很有可能我们获取的这个统计数据或者是其他那些信息可能就并不可靠，也有这个数据本来就是假的或者是说伪造的，然后得到了你得到了这个什么模型可能就不一定准确。

发言人 58:58

第三个，可能说我们用来估计参数的这个样本，它可能比较小，只能通过小样本得出这个结论。它可能只是抽样的一个偶然结果，它并不能真实的反映总体的客观实际。这个时候，我就需要对这什么模型加以检验。其次他也有可能什么违反经济学方法的一些基本的假定。对模型检验我们我们要检验什么呢？主要是对模型和模型设定需要检验，模型估计出来的参数也需要进行检验来判定在理论上是否有利，在统计上是否可靠。

发言人 59:50

这样一种检验，举个最简单的例子，比如说我们西电的学生，现在比如说本科生一年多少？一年6000人，然后我们就抽60个人来作为代表了。大家就可以想象到那人数是不是就有点少，对吧？但是我们还是抽了60个人出来作为代表，来进行这样一个建模的这样一个分析。最后我们通过这样一种收入消费的这样一个调查，得到一个结论，可能收入上涨，结果它的消费支出反而下降，做收入对消费的影响是负向的影响。

发言人 01:00:32

这个时候大家可以想象到模型检验是不是就很有必要了。也就是说我们不能只是我设定一个模型，我学会了统计估计参数的方法，我把它一估计好，这就是我们这个计量实证研究得到的结果了。那这样肯定不行，对吧？

发言人 01:00:53

就像刚刚我举的例子一样，我们在6600千个人里面抽取了60个人，那可能你抽样方法也有问题，对吧？就跟刚刚讲的，他可能这个收入消费的这个模型是一个分层或者分阶段的模型。比如说刚好我抽到这60个人里面都是什么都是什么五千这个什么收假设收入超过5000的，他可能在学校里面就顶多去综合楼消费一下。他可能你的这个收入再给他他可能还会增加他的什么投资需求了，对吧？他可能还觉得我还应该更多存一点钱。所以可能得到的结果是我这个收入上涨反倒使我的消费支出下降了，对吧？

发言人 01:01:43

得到这样一个结论就可以想得到这样一个结论。你下一步就是对他进行检验，对吧？你发现你看这跟我学的理论完全不相符，我的收入上涨怎么会使消费支出下降呢？那这个时候可能我们就要去review一遍你的这个分析的过程。就有可能是什么你估计的时候算错了，估计方法出了问题。

发言人 01:02:11

也有可能你在最开始数据收集就有问题了，对吧？因为你的这个什么你选取的样本，你看那个，这个收货的情况都是5000以上的。我的这个收入在涨，可能会增加我的投资需求，而不是我的消费需求了。那这个时候，你就再过一遍，看一下问题出在哪里。

发言人 01:02:37

这个是我们在计量分析里面也是非常主要的一个部分。比如说你需要去检验你估计出来的结果设定的模型。那对这个计量经验模型检验的方式，它也是有一些步骤或者有一些统一的形式的。比如说最开始你可以对你得到的这个估计的模型结果做经济意义的检验。你说你检验这个，你估计出来的模型跟经济理论是否相符，对吧？那有可能因为你刚刚得到这个什么收入对消费的影响是负向的影响，就收入增加会减少消费。那这个时候你觉得好像跟什么经济理论不相符。

发言人 01:03:29

但是你又过了一遍，就是review了一遍你的这个分析过程。你发现抽样也没问题，数据也没问题，估计方法也没问题。那实际上是不是你也可以去思考它可能是什么？可能是啊他把这些什么上涨的收入，他觉得他自己已经够到了一个投资的起步了。他可能就会抑制这样一个需求的形式，这个消费的形式增加了他的投资形式，或者说也有一些你没有考虑到的因素，你没有加在模型里面，这个可能就是模型设定的问题了。

发言人 01:04:07

比如说像现在一些你的悲观的预期是不是，或者说你这个大众的情绪是不是也会影响这个收入和消费的问题，就包括现在可能表现也比较正常的这种负向的关系也是比较常见的是吧？那你有这种悲观的负向的预期的时候，你可能收入再怎么涨，你可能还是更倾向于储蓄，对吧？因为你觉得可能将来就没有那么多的收入了，可能会下降。比如说你可能会遗漏一些经济的变量，这个是我们开始这个经济意义的检验，去看所估计的模型和经济理论是否相符。如果不相符，是不是可以找到其他的一些可能问题出在哪里。

发言人 01:05:03

除了这个经济意义的检验，还有什么统计推断的检验。给我们检验参数估计值它是否抽样的这样一个偶然的结果。我们做一些统计检验，来看这样一种参数估计值，它是不是满足我们这个统计的这样一些检验标准的，他是不是落在一些假设检验里面，是落在一个拒绝域，或者说是不是一个小概率的事件。如果说你估计出来的这个什么负向的收入和消费的关系，你通过这个什么统计的检查发现，这样的负向的关系最后落在了5%的区间里面，有小概率事件的区间里面。那可能这个结果就不可靠，对吧？

发言人 01:05:56

或者说是不显著的那第三步就是我们计量经济学的检验，去判断你设定这个模型是否符合降经济学方法基本假定。比如我们刚刚讲的5种假定，你的这个方程是线性的，你的这个就经经典的计量机学理计量机线性混合模型里面，你需要去判断他们的关系是不是线性的，变量之间是不是存在多重共线性，你这个自变量和随机扰动项是不是存在一个相关性，这个随机扰动项是不是服从正态分布，然后你的这个随机脑动向是不是存在相关性。做完这样力学的检验，我们也要做一个预测的检验。也就说我们可以把这个模型预测的结果与经济运行的实际进行对比。就比如说大家这个收入消费的问题，你可以你得到这样一种负向的关系之后，你可以把实际的数据，实际的情况也检验一下，来看一下。比如说大家可能大一的生活费，你可能到了大二爸妈会给你涨生活费。你就看是不是当这个生活率上涨了之后，什么消费支出反倒要下降了，对吧？看一下它跟你的实际的运行或者实际的经济的表现，是不是存在一些差别。

发言人 01:07:28

图做完了这个什么模型的检验之后，第四步我们就是对这个的模型开展应用，对吧？那这个呢啊将经济学这样一个模型估计出它这样一个模型它有什么用呢？它主要的用处有四个方面。

发言人 01:07:50

第一个是做什么经济结构的。也就是说我们要分析变量之间的这样一种相互的数量比例的关系。实际上它是不是就体现在那些参数和贝塔里面，对吧？或者说也叫做边际分析、弹性分析或者说程度分析。这些是不是都是希望经济学里面学过很多的这样的选择，收入上涨多少，或者说什么价格的变化怎么样，会影响你的购买意愿。就类似这样一些估计出来的参数与他进行。

发言人 01:08:31

第二个方面就是我们可以拿来做经济预测。这个我就不再说了。因为这个预测注意可以注意到有两种都叫预测。一种叫动态的预测。动态的预测是我用时间序列模型，就用今年的模型去预测明年的情况，这个就叫动态的预测。

发言人 01:08:56

预测他也不只能或者不止理解为动态，它在空间上也可以做预测。比如说我可以用什么我们尽管学院的这样一个学生的消费情况，建立一个模型来预测。什么算机学院，他们这样一种学生他的什么经济行为对这种空间的预测。或者说我们把西电的学生消费的情况，支出的情况也可以拿来预测。比如说交大的学生，他们的什么收入消费这样一些情况，比如说这个经济的预测，它不只是做一个时间动态的这样的预测，做黄金价格的预测、股票的预测，经济走向的预测。它也可以是什么空间上的这样一种预测的范畴。

发言人 01:09:55

第31个是政策评价。所谓政策评价就是我们可以把计量经济模型作为一个什么实验室来看待。那你既然故意出这样一个参数了，你当做一个实验室是不是就是假如我们某项经济政策发生了改变，说我们影响因素发生了变化，那我们研究对象这样一个变化会有多大，有什么反应？是不是我们就可以分析出来这样一种政策它是好还是不好，我们这个政策应该采取多大的力度才好。

发言人 01:10:31

比如说我们这个还是这个收入消费的问题。比如说像校长说要给大家多发奖学金，都给大家上涨收入。那这个时候大家就可以用我们这个模型来刻画这样一种政策的好坏。就比如说我们刚刚其实可能初衷给大家这个普适性的可能不是校长决定，或者说是我们现在刺激消费的这样的政策。他不是可能在香港或者很多地方，他之前都采取过这样一种微信的现金货币的形式。就是直接给每个人发几千块钱对吧？他来刺激这个经济，让他赶紧运作起来干。

发言人 01:11:21

那这样一种情况，比如说他就可以比如说我采取这样一种政策的时候，然后我们又得到了收入和消费这样一种关系的刻画，或者这样的模型的刻画，你是不是就可以评判出这个政策的好坏。尤其是比如说我们在西电这样一个预测范围内的话，我们发现给这个学生增加收入的话，他们不会去消费，对吧？那这个时候这样一种政策，它实际上是不是就起不到这样一个刺激消费的作用。那这个时候可能你就可以对这样一个政策进行评价。

发言人 01:12:02

另一个是最后一个，是我们可以验证理论或者说是什么发展理论。比如说有的时候我们这个经济理论甚至是不完备，甚至是错误的。但是实践是检验真理的唯一标准。就比如说老师有一项研究是做什么？就比如说大家可以看到现在大家都在网购，对吧？有这样一种常识，或者说是在他有的时候也被称作一种理论，就是什么？

发言人 01:12:36

因为大家现在都是在网上去购买，就和很多经济活动都发生在线上的这样一种环境里面。那是不是你所处的位置就不再重要。就比如说现在大家想创业的话，你想甚至也是卖这样一种标准的，你想卖这样一个iphone手机。那你可以想象到如果你在网上开一个店，你开在深圳，开在西安，或者说你开在某个小地方的话，是不是感觉好像都一样，对吧？你都是在网上去发生的，就不像从前，你可能必须得到什么大城市人流量多的地方开这样一个店，才会有人去去找到你的店。

发言人 01:13:23

但现在到在这样一个在线的市场当中，有这样一个理论是说这个物理的空间不太重要，或者说是物理的区域都不太重要了。比如说有有的理论叫什么地球村理论。就你在哪里都可以用很便宜的物流成本去发货。然后大家也都点击下鼠标就可以访问到你开在某个小地方的这样一个店铺了，对吧？那这个是大家广泛接受这样一个理论。

发言人 01:13:55

但实际上我们通过把淘宝所有后台合作的一些数据我拿出来看，就发现实际的经济表现不是这样。它是什么？比如说我就卖这一台iphone手机还是存在一起。比如说我在一些地方，我就是能有更大的一个什么成功的可能，或者销售的可能，说有一些可以我们去构建相应的模型，去验证一些理论它是不是正确的。从而如果说是这个理论和你得到这个实践的观察结果是不一样的，那这个时候可能就有一些可以开展空间。那我们先休息一下。

发言人 01:15:48

我做的这个一切都很享受。

发言人 01:16:15

裙子。

发言人 01:20:34

好的，那我们继续。所以说我们这门计量经济学这样一种步四个步骤，如果说归纳起来的话，大家可以用这样一个什么框图我们来说明。你就说什么我们依据经济的理论去设定，我们相应的这个。这个模型的设定除了理论作为依据，我们也需要去把根据我们实际的这样一种经济活动。比如说有两头，一头是这个理论，另一头就是我们事实根据我们的认识经验和理论来设定相应的计量模型，这个是第一步。

发言人 01:21:26

如果说设定错的模型，根据我们收集的统计数据来固定参数估计出参数以后，这样一个模型需要加以检验，检验什么呢？那就检验我们刚刚说的四个方面的内容，看它符不符合我们的要求。如果都不符合怎么办呢？那我们就是不是需要修订模式。就比如说我们得到了什么收入上涨会抑制消费支出，对吧？如果说你得到出了这样一个这样一个模型的解模型的结果，它不符合你们或者说没有通过我们刚刚4种检验当中的一种。那这个时候你就返回去修订相应的模型。

发言人 01:22:17

那这个模型可能设定错了？也有可能数据收集错了。那修改完了之后也有可能遗漏了重要的变量，对吧？就比如说我们刚刚讲的这什么经济的预期，对吧？有可能是因为现在大家普遍预期不好，所以收入上的他的消费还是会下降。那位要去修订相应的模型，然后收集相应的数据。修改完了之后再次的去估计，重新估计参数，然后再检验，直到检验通过。

发言人 01:22:55

比如说我们几个方面是合理的，这个时候模型就可以来加以应用。应用就这四个方面，结构分析、经济预测、政策评价以及验证一个。这个是我们计量经济分析的。如果说通俗来讲的话，就是这样几个步骤要明确任务。也就是说我们需要首先确定我们的研究问题，你你有你有可能？

发言人 01:23:30

我们开展这个什么计量经济分析之前，你需要明确你到底要针对的问题是什么。设定相应的模型。也就是说运用经济学的理论对你明确的这些问题进行思考，或对某一现象进行研究。分析方案之间的关系选择并建建立计量模型。那之后你设定好了模型，是不是就可以收集或者获取样本。数据获取样本数据之后，就是估计模型的参数，然后再检验，然后在应用。

发言人 01:24:08

那那明确任务也是比较重要的一个问题。很多很多时候我们说的是提出一个好问题可能比你去解决它还要更重要，对吧？那明确任务就是你要确定你研究的问题是什么。这个问题有可能来自于很长时间对某一现象的关注，也有可能来自于你对某一问题的突发奇想。

发言人 01:24:46

比如说要研究中国家庭消费支出受家庭收入和财产的影响有多大，比如说你也想要分析我们新上市的，比如说这个小米，小米十7它的定价和这个销量之间的关系，你想知道，有的时候你也想知道。你比如说你进了一家商场，那你发现这家这个商场的冷气开的很足，你是不是也想研究一下，是不是这个商场如果冷气开的越足，是不是我这个商场的销售表现会越好。或者说你去了一个商场一开这个商场跟打的很亮，是不是也会突发奇想，是不是我这个灯的颜色，灯的亮度也会影响。而我们什么销售的表现，或者商场的营收。那你明确这个任务之后，就会明确这样一个问题。

发言人 01:25:50

之后，你还要考虑这个是模型的这样一种适用的环境。在模型设定里面，它有比较重要的几个部分。首先确定变量，变量它也有一些划分。比如说什么我们计量经济学中的变量，它可以分为什么？

发言人 01:26:12

解释变量和被解释变量，也可以称为什么自变量和因变量。自变量、因变量可能好理解一点，变量什么变量是X你说我变化会引起别人的变化，因变量是我因为你变了，所以我才带来了我的变化。那解释变量和被解释变量也是这种。解释变量就是被解释变量就是我的变化是因为可以由你来解释的，所以他被称为被解释辩。那被解释变量其实就是什么？我们的Y，也就是我们要研究的问题。那解释变量或者说自变量，它其实就是我们所要研究的问题，它所涉及的这些所有的影响因素，那建立计量经济模型，它最关键的，你可以想象到它是确定的确定我们的什么解释变量。

发言人 01:27:18

那一般我们确定我们这个模型有哪样一些X？它是根据我们什么经济理论或者说实践经验来进行判断的。也有可能根据我们现实的这样一种研究工作进行具体的确定。就比如说我们现在探讨的一个，那会不会是什么消费和收入之间的关系，受什么受大家对未来的预期的影响，或者说有可能受地缘政治的影响，有可能受关税政策的影响，对吧。这个可能是你突发奇想看到新闻觉得现在老美对我们采取的什么贸易保护这样一些网吧单子的政策之后，是不是会影响到我们大家这样一种消费行为，或者说这样的消费习惯，那你突然奇想想出这个问题，你可能就可以把这个问题开展延伸。

发言人 01:28:22

你的Y是什么？你的X是什么？比如说根据理论来讲，你的消费anyway都是受你收入的影响，那这些就是非常确定的，你可以根据理论加到模型里的X那也有一些是根据你的什么，比如说他根据你的什么实际的研究工作和你的兴趣。比如说它的关税，可能对，不是说关税，就可能这样一些贸易政策也会影响到你的消费习惯，或者你的投资行为，或者说现在A股的表现，可能也会影响到你的什么消费支出，对吧？因为大家都被套牢了，那我收入再高我可能也不行，对吧，也没钱消费了。

发言人 01:29:13

模型形式的设定，我们根据经济理论去设定相应的这样一种形式。然后根据经验和事实分析相应的研究问题，确定这样一些主要的影响因素。然后根据样本数据我们来绘制相应的散点图，通过散点图也是通过实践数据表现出来的变量之间的关系，我们来确定它们的函数形式是什么样的。比如说很明显你可以看到我横轴是X纵轴是Y我把X的数据和Y的数据都画一张图，可以很明显的看到他们是这样一种线性的关系的话，那我就可以设定线性的这样一种函数关系。那有可能X和Y我这个图画出来它是一个抛物线的关系，或者说是一个其他形式的关系，我们就可以设定相应的这样一种不同的函数形式。

发言人 01:30:22

很多时候我们会遇到这个模型数学形式难以设定的这样一个情况。这个时候我们就可以好吧接近的几种形式。比如说有的它那个图形你不一定就是用什么X方，比如说抛物线，或者说用log曲线就可以刻画出来，有可能是什么log加3次方，类似这样的形式才能很好的去把这个变量之间的关系或者这个曲线表达出来的时候，这个时候我们可以选择几种不同的形式去进行模拟计算。然后再根据我们之前的原则，不能太复杂，也不能太简单，你最好能用一个简化的形式，比如说7788的，你也能把这个变量之间的关系表达出来。那这个时候就是比较理想的一种数学心理，说它会存在有选择的情况。

发言人 01:31:24

模型设定里面我们也需要事先分析参数的符号和变化的范围。那你这个客户道格拉斯生产函数来讲，比如说Y表示产出，A就表示技术进步。估计值是不是应该大于阿法和贝塔的部分？分别表示资本和劳动力的产出弹性。这个估计只能根据理论来讲，它应该是0到1之间的。如果说模型的参数符号合理范围不相符，那之后我们就有什么检验的部分，你就可以判断它是不合理的，可能就需要进行相应的这样一些改动。

发言人 01:32:14

那在获取数据方面我们通常来讲获取数据可以从传比如诸如比如说政府统计部门，比如说国家统计局、国际能源署、世界银行，还有一些什么学术机构，比如说我们学校也有购买一些数据库，比如说wings，这样一些金融的数据，它通它是一些商业机构收集整理出来。那我们可以从一些这样这样这样一些二手数据，比如说经过加工的非原始数据来获取。那另外，我们也可以，根据我们自己的研究需要，比如说你很难去从什么这样一些二手去二手数据里面去找到西电学生的收入和消费情况，对吧？那这个时候我们就需要自己去设计问卷，进行访谈，去收集一些数据，这个时候就称为一手数据。

发言人 01:33:33

常用的如果说大家对一些做一些研究感兴趣的话，常用的这样一些国内的数据可以从这样一些数据库获得。比如说中经网数据库，中金网它是涵盖了宏观行业、区域以及世界经济各个领域。那这个时候你可能就不用再去世界银行，比如说能源署这样一行业和其他官网去收集，他可能直接给你整理好了。还有问的万德数据库，这个数据库它与宏观和金融的数据为主，涵盖了一些股票、基金、债券、外汇、保险、期货这样一些数据。

发言人 01:34:24

也有比如说这个中国家庭动态跟踪调查，CFPS是北大他们构建的这样一个数据库，就长期的追踪调查，访谈中国各个区域不同的家庭这样一些情况。他的那个问卷里面收集的信息就比较详细，有一些你的家里有几口人，你的收入是多少，支出是多少，你的资产是多少，它里面的一些微观的数据非常多。比如你的这个教育水平，健康状况是什么样的那除此之外，还有中国健康与养老追踪调查，中国健康与营养调查、中国家庭收入调查，这样都是一些二手数据。

发言人 01:35:26

常用的国际数据也有一些数据库，比如说当前人口调查CPS，还有一些长期追踪调查，医疗支出面板调查，资产价格研究数据库，还有一些比如说宾夕法尼亚大学世界表。上面是我们去收集数据的一些渠道和来源。数据收集我们有一些不同的类型。常用的统计数据主要有什么时间序列的数据、截面数据和面板数据。这个分起来就很简单，时间序列数据它是以什么时间为维度的。比如说从2000年到现在我们西电的学生人数，西电的或者说从2000年到现在我们中国的这个GDP特征。或者说也有一些比较高维的信息。比如说你的什么天气的变化信息，比如每一天的记录，甚至每个小时他都记录了一些时间的，以时间为维度的情截面数据是什么呢？

发言人 01:36:53

截面数据就是以不同个体进行区分的。比如说2000年中国的各省的GDP情况，它就是一个什么横切面截面的那面板数据它是什么一个交叉混合的数据，就是它可能包含了中国每个省份2000年到现在每一年的什么GDP的数据。那这个就是不同种的这样一些数据的类型。

发言人 01:37:29

时间序列，比如说GDP，失业、就业、股票价格这样一些时间间隔，它也可以是不同的，有年的、季度的、月的、日的、时的、分的、秒的。前面数据它的什么收集的时间成本比较低，它在经济分析当中也是一种比较常见的数据类型。比如说我们要收集现在学生什么经济，收入和消费之间的关系的时候，是不是我们就可以收集一个检验数据的信息。我们通过对不同学院不同年级进行分层抽样，抽象出来的这些人，我们全部给他进行一个信息的收集，让他记录你的收入情况，消费渠道。面板数据可能就是什么也包含了时间，也包含了截面。就比如说我们在这个厂十年的时间内，持续的每一年都收集一些这样的消费指数的数据。

发言人 01:38:40

在大数据时代，我们去收集这样的一些数据，也有一些需要注意的事项。首先是什么？我们去选取数据的时候，应该选择具有公信力的这样一个权威机构的数据。如果说是自己收集的一手数据，也需要注意数据的可行性和科学性。收集到数据还需要进行数据预处理，才能用来建立模型，满足可比性和一致性的这样一些要求。比如说大家收集什么消费支出的情况，很多时候你登记的这个单位有可能不一样。对，而且登记的这个维度也有可能不一样。这个时候你就需要对数据进行一个预处理。

发言人 01:39:32

在大数据时代，80%以上的数据都是非结构化的数据。非结构化的数据它可以通过数据处理后变成结构化的数据，作为变量引入计量模型当中。比如说如果我们想要把文本的数据、图片的数据、视频的数据、音频的数据也放到我们这个计量模型里进行分析的时候。大家可以看到，比如说现在做一些音频的数据，他就会把这个音调、语气给他用数值去刻画出来，加到模型里面对吧？那大数据它并不代表全样本，无论多大，它也是一种样本信息，需要建立在抽样理论基础上去进行统计分析和计量建好，那我们这节课就先上。

发言人 01:41:02

我一天没下。

发言人 01:41:54

我明天下午没有这个盒子。

发言人 01:42:03

没有的。