EM114026大数据计量经济分析2025年10月21日第8周星期二第7-8节-pptVideo\_原文

2026年01月04日 12:54



发言人 00:40

对，取消不是。这什么天？

发言人 00:56

这边好像没有提示。还是在这上面会显示。

发言人 01:08

这个应该可以，但是不服的可以，因为我好我好像没有处理过，我每次来都没签到。













发言人 05:21

没签到的同学买到一些。

发言人 06:00

我们继续我们这个大数据计量这样一个课程的。在上一节课我们已经进展到我们这个第三讲，也就是我们这个计量经济学里面最核心的内容。我们从最简单的这样一个经典线性回归模型入手，然后去看怎么样去开展我们之前课程里讲的这样给你分析的几个步骤？我们怎么样去提出问题，构建收集数据，构建模型，估计模型的参数，然后检验模型，最后把模型拿来应用。那我们看这个什么经典线性回归模型的时候，我们上一节课给大家讲了一些什么叫做回归分析对吧？以及回归分析它的定义和还有我们回归分析虽然说的很通俗，但很多时候我们会跟什么相关，会存在一个什么一个混淆。比如说我们一起来回忆一下回归和相关有什么区别呢？

发言人 07:26

艾丽孜热亚。回归和相关你觉得有什么区别，或者他俩一样吗？不一样对吧？那哪个跟哪个相关，包括回归吗？

发言人 08:00

相关分析里面我们会包括回归分析吗？还是回归分析里面蕴含着相关的。

发言人 08:19

就是相你的意思是相关的分析范畴更大是吧？是这个意思吗？

发言人 08:33

是不是我们去研究两个变量或者两个指标之间关系的时候，比如说是什么X和Y，那X和Y是相关的这样一个结论，它会包含X和Y它俩是回归的这样一种关系吗？

发言人 09:09

或者说X和Y是回归分析的范畴的话，那他俩肯定是什么是相关的对吗？这样一句话。

发言人 09:27

或者举一个实例来讲，就是收入会影响消费。就收入这样一个变量或者这样一个因素，它会影响大家的消费，对吧？比如说今这个月大概收入了多少？收入5000块，有可能消费4000。比如说下一个月你涨工资了，你的什么收入增加到了8000元，那你的消费还会是4000块钱，你可能会也会涨对吧？

发言人 09:55

就是他会有一个什么互相的什么影响，而且还有一个类似这样箭头导致的关系，是我消费提升是什么？是因为我的收入上涨了，所以我的消费支出提高了。那他俩是相关的吗？那他俩是相关的关系吗？是的，对吧？好，请坐。也就是说如果说两个变量是啊存在这样一种因果关系，或者说是互相影响关系的话，那他肯定是什么是相关的。但是有的变量他们是相关的，但是不一定是什么，不一定是存在这种因果关系，或者说是开展我们这个回归分析的。

发言人 10:44

就比如说什么我们学校小卖部卖冰淇淋的销量和秦岭里面什么水库的灭亡人数，他们可能是相关的。比如说我冰激凌销量上涨的同时，肯定会伴随着我什么这个水库溺亡人数的上升。但是他俩是不可以进行这样一个回归分析的。

发言人 11:10

就举了举了这个例子，它很显然。但是可能大家后续在开展你们自己感兴趣研究问题的时候，很有可能会碰到一些这样的问题，他会隐藏的比较深。就是你可能还是需要去在我们最开始的时候，对这样一个回归分析和相关分析有一个很清楚的界定和理解，然后才能有啊有有有能应用到我们实践的问题当中。

发言人 11:44

比如说还有一些常见的误区，比如说可能很多人会觉得吸烟导致肺癌和吸烟的这样吸烟年限和你肺癌发生的概率，就可能很多人会来构建一个回归模型。但是实际上它在医学里面，它也不是一个什么因果导致的。至少他们在生物学研究上面没有发现有这样的很强的因果关系。所以他俩可能也是一个相关关系。包括什么one night day，什么keep down away，它都是一些什么一些可能是相关的影响，但是它不是一个导致的导致或者说因果这样的关系。他可能很多时候我们在构建回归模型的时候，你就比如说你再把这个天气夏天天气的这个地方反复去琢磨，或者套用到你的这个实际的问题。

发言人 12:43

在上一节课我们给大家讲了什么总体的回归函数，还有相应的例子。我们从一个什么一个图里面我们就可以很清晰的去看到。我们怎么样来开展这样一个计量分析，或者说这样一个量化的分析。我们实际上是什么把现实的数据捕捉出来，捕捉出来之后，我们去把它抽象成什么，一个模型简化的模型。比如说外语A加贝塔X这样的形式去抽象的表达出来。

发言人 13:29

因为我们不管是实战分析也好，或者说大家去解决任何问题，也我们想要得到的想要得到的普遍性的规律，或者说是针对总体而言的这样一个规律。就是可能比如说我想了解的欺骗学生的消费收入关系，或者我想了解的是啊中国居民，比如说像什么新能源汽车购买意愿，类似这样的问题，就我的总体，大家一定要去把它很好的去界定。就在开展这个机长更新或者不实在更新之前，一定要明确你的问题的总体是谁。比如说你明确好你的问题总体，那我们想得到的是什么？我们想得到的是针对我总体而言是不是存在一种这样的规律。但实际上我们把这个什么数据抽象成这一条线，我那我最理想的方式是什么？

发言人 14:30

我比如说我的问题是我想了解新建学生收入消费的情况，然后来看到底要不要给大家发放这个生活书，很多我通过什么样的政策来刺激大学生去消费，那我比如说即使是这样一个西电学生的总体，它的规模也不小，比如说几万人，我把每一个学生都发放一个问卷去问的话，就你刚才想说的这个成本是很大的对吧？那你更不用说我们通常研究的问题是一个省或者说一个国家这样的层面上。也就是说我们往往不可能对总体的数据全部收集起来。更多的时候是什么？我们去尽量的获取一些可以代表总体的样本，然后对这个样本的数据进行什么进行建模，然后在服务器分析。这样大家可以想象是不是就有一个问题，我估计出来的这个样本的这样一个规律，它是不是就是总体的这样一个很好的代表。

发言人 15:40

在这个图里面我们上节课是不是给大家讲了，我们这条蓝线是什么？是不是我们的这个样本回归线？或者说我们的样本回归模型，也就是它从数实际的数据当中估计出来的，还是存在这样一个关系。这条蓝线。但是实际上对于我们总体而言，或者说我们这个大规模群体而言，真实的规律它可能并不并不是这条蓝线，它是什么这条红线所代表的。大家可想而知是这两条线会存在差异是什么？也可能存在抽样的误差，对吧？

发言人 16:23

你从数据，你在这个叫什么参数估计到一条线，抽象到一条线来表示的，它也会损失很多的信息，对吧？那这样我们这个估计的好坏，或者说我们决定这个实证分析的好坏，就从就简化成了一个最简单的问题。就是什么我怎么样来判断我的什么样本回归线离我的这个总体的这条回归线是最近的，或者说他俩从如何最好，对吧？如果说我能严格遵循法抽样的规则，然后也很好的收集到了所有人的数据，然后汇大家汇报数据的时候，也不存在一些什么小心思，或者说错填什么之类的这些问题。那我就可以很理想的得到一条样本回归线，它能很好的去跟什么红线，一个总体的这根线去重合。

发言人 17:27

但是更多实际的情况是什么？你得到这个样本回归线，他跟什么总体的，就是你想要得到这项规律，它是存在一定差距的。就比如说你想要得到的结论是什么，欺骗学生收入上缴上涨一个单位，你的这个消费支出会增加0.7个单位。但是这个是比如说针对我们班同学，我收集到的这样的数据是这样的表现。那可想而知，比如说我们都是什么经管背景的学生，可能本身就比较有经济头脑，你可能就不会大手大脚。我只是举一个这样的比方，因为比如说其他可能那些学其他工科专业的，他们可能比如说他对钱没有太多的。我只是举一个这样的例子，可能不一定是这样，那他可能他就会有不同的表现，所以它会存在一个这样的误差。上节课我们给大家讲到了这个我们这两条线之间的关系，或者这两条线的距离，我们是不是现在就可以看出来，我们从一个什么实际的值YI距离，我们找一个同学们起来，我看一下这个图。

发言人 18:57

郭勇。这个大Y的点是什么呢？那我刚这个是实际值，真实值对吧？就比如说我们西店消费收收入消费情况就是你的收入和消费情况对吧？

发言人 19:19

真实值那这条蓝线跟它之间的差距是不是有个小的EI行，那这个EI表示的是什么？就是绿色这个叉。对它是一个残差值对吧？它是什么？你跟我们班同学平均估计出来的是不是也有差距，吧？就是比如说我们班，因为大家可以想象到，在开始的时候，老师就给大家强调了，我们构建模型得到都是一般意义的结论，对吧？

发言人 19:56

就假设我对我们班所有同学都调查数据之后，构建这条缆线的模型，最后是发现什么？大家收入上涨，一个单位消费支出增加0.6个单位。但是因为这位同学他可能平常花钱就比较大手大脚，他在这个YI的点，它跟预测到的平均值就有差异。比如说我预测出来我们班平均上是什么0.6，对吧？但因为这个同学他花钱带给大家，他跟什么我们估计出来这个样本，这样它它也会存在一个差距。比如说他可能上涨一个单位，它可能就消费支出就多一个单位，它可能就是今朝有酒今朝醉的这种，它可能就会存在一个残差值，也就是EI。那你可以看到大YI跟这个什么红线之间是不是也有一个差距？这个缪I它表示的是什么呢？

发言人 21:03

是不是我们最开始说的我们这个计量模型和理论模型差别出来的这样一个随机扰动项，对吧？随机因素或者随机扰动项，这个值它的意义是什么？就是这位今朝有酒今朝醉的我们班这个同学的这个样本，他跟什么他跟我们西电学生平均得到的我们这个结论。就可能因为我们是经济类的，都很有经济头脑，你可能不会全花完。但是比如说放在我们这个工科院校的背景，假设说我们把所有学院所有人都统计完了，那可能得到的结果是什么？我们训练学生收入上涨，一个单位消费支出多0.8个单位。那这个时候，这位同学距离这个总体或者说是最理想情况下，我们估计出来这样的总体的这样一个函数的距离，就表示在什么又安利了对吧？就这个差距你就可以看到长颈鹿了。

发言人 22:12

大家可以看到这里面老师一直在给大家强调这个总体回归和样本回归这两个概念。下面我们一起来看，我们在收集完数据之后，我们也有什么样本数据信息了。而我们怎么样来先得到，就是我们可能得先估计出来这样一条蓝线，然后我们才来看这条蓝线估计的好坏，对吧？那我们先来看我们怎么样来估计这条蓝线。在估计这条蓝线之前，我们对我们构建的这样一个模型，它就有一些预先的假设。首先我们上节课也给大家讲了什么叫做什么一元回归，什么叫多元回归。

发言人 23:00

听名字就知道一元回归就是说什么我考虑的影响因素只有一个。多元回归就是我考虑很多的因素都放在这个模型里一起建模，一起估计，对吧？那这个模型比如说一元线性管规模，它也称双变量模型，多元线性管规模。现在它就什么X？它就都放在这个模型里面的。

发言人 23:28

对模型的设定的假设是我们这个线性回归模型基本的假设。首先我们估计这个模型的时候，它我们先假设它是一个线性的模型。其次是这个回归模型是正确设定的。这个正确设定它有几个不同的含义。我们找同学来回忆一下。

发言人 24:02

李东基，就是我们通常讲我设定了一个对的模型，这个对的模型是一个什么样？它包括哪些方面？

发言人 24:20

对，就是从我们都已经设定好了，估计完了，然后这样来看对吧？就他最好是这样。比如说我们只在第一步，就是我在设定这个模式。比如说我们研究收入消费的这样一个问题，我现在列出了Y等于A加FBX加X2。如果怎么样来说，我设定的这个模型是一个好的模型，就还没有开始这个参数估计的这样一个过程。

发言人 24:55

首先就看它它的变量的数量，比如说收入少。面料做不过去。对，就是可能理论上或者实践经验里面大家能想到的，跟这个收入消费，或者说影响消费的这样一些因素，至少都要包含在这个模型里面，对吧？然后可能你根据你自己的研究问题，比如说你可能考虑性别是不是会影响收入消费的问题，或者说你的头发长短也会你再逐渐加增加你自己的变量进去。那除此之外？

发言人 25:41

比如说我设一个什么Y等于阿尔法，加上这个E的X就是我把那个收入消费写成指数的形式或者对数的形式，这样是不是也合适呢？或者这个问题上他可以称作一个好的公司。

发言人 26:20

有什么好处好不好？我不知道是他其实是什么，就是我们上节课讲过，就是我们在设定这个模型的时候，我们比如说收集好数据，在设定这个模拟是不是也要看这个数据之间的关系大概是一个什么样的形式。比如说线性的关系，我们肯定就不能给它设定成指数之类这种模型的形式对吧？他的这个模型的形式也是这个模型好坏的一个方面对吧？好几个。这个就是我们在什么对模型奖励的时候，我们就假设前面都已经搞定了，就是模型是线性的。然后这个回归模型也是正确设定的。我们根据我们涉及到的数据去给它拟合了一个最好最简单的这样一个形式的去表达它。

发言人 27:15

对自变量我们也有一个奖励，是就是在我们回归建模里面。因为大家可以想象到我们的X自变量Y是因变量，或者说是。在这样一个由自变量去影响因变量的过程当中，我们的这个什么因变量往往认为它是一个什么，它是一个随机的，它是受什么东西影响导致的，就是它是一个不确定的东西。因为比如说大家的这个消费的问题，你肯定要把它假设成随机的这样一个变量。因为这样的话它可能受很多因素的影响，它一影响它可能就变了。所以它是一个什么不确定的这样的值。但是对于我们这个模型里的所有X来讲，我们就都把它假定成把一些确定的值。就是它不具备一些随机的属性，我们都认为它是给定的。比如说我给定你这个月的这个收入是多少，就类似这样的问题，我给定你的性别是什么，我给另一个头发长短，就类似这样的情况。

发言人 28:29

其次我们也假定自变量之间它不存在什么多重贡献性。或者说是如果说是两个变量完全相关的话，那这两个变量是完全共线性这个参数，这个时候它就是没办法识别的那举一个最形象的例子，比如说有一个人在那个经理甲站在这里，乙和丙两个人手拉手握着拳把他推了一把。那大家可以想到这两个拳头开始，贡献的对吧？他就是我，我这个手往前，另一个手肯定也往前的那这个时候你要来判断让这个甲让甲这个人往前动，他这个什么乙贡献了多少，丙贡献了多少，你可以想象到它分不出来，对吧？那没办法，你没办法分出来。比如两个人手抓着手把这个人往前推一下，你不知道往前推一下你和丙分别贡献了多少，或者说要估计出来，他们的影响其实是很难。所以在我们这个自变量假设的时候，我们也需要去假定你考虑的这些因素，它不存在一个法完全相关的这样一个关系，或者不存在这样一个多重贡献性。

发言人 29:56

就类似我们刚刚讲的这个收入消费的问题。如果你想做一个性别类的考虑，比如说可能很多人会觉得，都是因为可能有一些传统观念就觉得女儿就要富养，男生就要穷养，那可能消费习惯就会存在差别。那你把这个性别这样一个变量考虑进来之后，你可能又考虑到是不是理发也会影响到我这个消费，那是不是头发长短可能也会影响我这个消费支出。当你把头发长短和性别这两个变量都放在这里的话，尤其我们又是西部的高校，那可能长头发的男生比较少，然后短头发的女生也比较少。

发言人 30:39

这个时候他会有一个什么？有一个可能两个变量就是在一些情况下它很有可能是分不开的对吧？那你把这两个变量一起加到你的模型里，你就可以想象到它跟老师刚刚举的这个两个人手拉手把另一个人往前推一下的这个感觉是很类似的。你没办法分清楚你往前推是由乙造成的还是由丙造成的，或者说他们俩分别造成了多少。所以你又再科学，再厉害的统计大牛，它其实也分段分分分段不出来，对吧？所以我们在进行假设的时候，我们起初就假设他们你在筛选的时候，你的头发长短和性别这两个因素，你就可以看到他俩是非常强相关的，可能是完全相关的那这个时候你就应该剔除掉其中一个变量，就是让我们构建的模型里面这些自变量，它不存在这样一个构成构建性的问题。

发言人 31:43

除此之外我们对这个什么随机误差项目也有相应的价值。首先就是这个随机误差项的平均值价格。平均值的假定很好理解，就是我们在构建模型的时候体现。刚刚那个李东彬同学也说了，就是我们在构建这个模型的时候，已经把什么所有核心的关键的变量都考虑在内了。也就是说对我们的Y存在大比较大影响的这些因素都在X里面了。其他一些微小的对影响因素就都被放在什么随机误差项BUI里面。

发言人 32:23

这个缪案例大家可以想到，我没有考虑到的这些因素，或者说也有可能是什么数据收集的误差，或者说是啊大家在这个汇报的时候也存在一些心理的问题，可能会隐私的问题考虑，可能会把这个收入往低了报，就怕露富对吧？就怕大家知道我有钱，万一被抢了100公里，就会有这样安定位，就会有这样一些各种方面的误差。最理想的情况是什么？我们把这些所有其他没有考虑到的也好，我这个甚至可能填报数据，或者说是收集数据，或者说抽样形成的误差。所有所有我我所有没有考虑到的，或者说我可能考虑到我也没办法去解决的这些因素，我全把它打包放在了这个什么模型最后的这个尾巴的缪安里面。

发言人 33:23

这些大家可以想象到，有的可能是正向逆向对吧？比如说有的人觉得我怕露富，我就往高了。反正有的人觉得那我往高的话可能还好一点，对吧？那就有这样一些正向的负向的影响在里面。

发言人 33:38

那最理想的情况是什么？这些我们打包在一起的没有考虑到的细微的因素，或者说统称为误差随机残差项、随机误差项打包在一起的话，它的这个什么加合起来是零，那这样就是最好的。也就是说正负相抵了，抵完了之后剩下的我们估计出来的就是我们已经放在自变量里了。那这些就我们可以用来估计，就其他的反正我也不不不细究你对我这个Y的影响是什么，反正你的影响都是0.1，但是另一个又是-0.1，我把你们全放在一起，你们加合起来是零，这样我们这个模型就是最好的。

发言人 34:24

除此之外，还有什么同方差的奖金。同方差的奖金是什么意思呢？就是我们在给定我们这些自变量的时候，大家可以想到有一些情况下我这个X分组存在差异的话，有可能我在不同的什么情况下，我们这个影响因素也好，它可能会存在一些差别。在这里类似是啊我们假定所有的情况都是一样的，都存在相同的方差，可以把它表示成一个set方的这样一个形式。

发言人 35:05

同样我们对这个随机误差项也存在什么无质相关的这样一个假。比如说在我们给定相应的我们无法确定的这些自变量的情况下，我在探究不同的外部。比如说大家可以想象到我们什么不同的模型，它存在就也不存在一个相关这样的关系。这个时候可能我们得到的结论才是准确的。

发言人 35:39

除此之外，我们也做一个是震撼性的这样一个假定。因为我们假定这个随机误差项的刚刚讲了，它最好的情况是正负相抵，均值为零，对吧？然后同方差假定我们也讲前面也讲了。所以最后这个正态性的假定，其实就是对前面均值和分开假定的这样一个归总。就是会发现就我们假定最后的这个冷冻项，它是服从一个正态分布的，均值是0，然后这个什么方差系的方这样一个产品。那如果说我们构建的这个模型，比如说Y等于2派加贝塔一X加贝塔X加BY这样一个模型，它满足了上面所有的讲义，那它就是一个什么古典线性回归模型。这个是我们最经典的这样一个回归模型。那后面，我们就看，如果说满足了前面的这样一些假币，然后我们这个经典的这样或者说古典的这样一个经典的这样一个线性回归模型，我们该怎么样来对它进行一个参数估计。

发言人 36:54

同样也是以我们第一次上课的这样，就现在校长你面临的问题就是要怎么样去选择一个合适的这样一个班级规模，因为你的什么发现你的什么预算是有限的，你肯定也不能把班级给我缩减到1。每个老师教一个学生，那成本太高了。但你也不可能让一个班200个人让一个老师教，那可能最后这个上课的效果也很差。

发言人 37:29

所以这个是大家作为校长去就去自己下面的班级进行了一些调研，或者说你收集到这样一个数据。大家可以看到这个数据有什么observation，就是我们的这个什么样本量，你说过比如说观察一组数据它就标记为01，总共是有多少组？总共是有420组这个观测值或者说是观测变量对吧？

发言人 38:05

那第二列大家可以看到这个数据，就是我的这个五年级的他的这个平均的考试成绩，他把它做了一个记录对吧？第一个是690，然后第2 661 10。其次他也收集了每一个学校他的students和老师数量的这样一个他叫student t恤的两个video。把这个比值它肯定作为什么X可能比如说有的学校老师可能多一点，有的学校的老师少。所以他把这个变量的数据也收集了。

发言人 38:51

第三列还是什么？他是一个每个学生他的这样一个经费也给他进行了一个收集。还有第四列就是这个学生学习英语的这样一个比例的这样一个信息。大家就可以看到我是不是相当于至少如果我关注什么师生和学生成绩这样两个关键变量的数据，它其实是不是也都收集了。

发言人 39:25

收集了，我们后面就看我们构建一个最简单的模型，我的这个Y就是我的，我作为校长，我肯定想提升我的这个升学率跟我的这个学生的成绩，那我这个Y就是什么？我们这个test score，就是它可能期中考、期末考试和期中考试的这样一个平均成绩。贝塔零就是它的截距项，那贝塔一乘以STR60码，我们需要估计的参数，贝塔1STR内部自变量对吧？那大家可以想象到我们之前讲过我们分析的步骤，对吧？我现在已经把数据收集起来了，所以起来了我是不是就可以很简单的去把我这些不同的什么样本，或者说我们的观测值observation，把它都画在图里面。大家可以看到的这个图，横轴就是我们师生比，重轴就是我的学生成绩。原来我把这420组数据，或者说就形成了420个点对吧？它对应的就是每一个学校，它的什么学生师生比和什么学生的成绩。

发言人 40:45

大家可以看到我的贝塔一，如果我要去估计它的话，大家可以想象到我这个什么甚至就是用一个差值，就相当于是我们高中还是初中学过的内容，我怎么样去估计一个斜率，如果我要有一条线去拟合这些点的话，是不是贝塔一就相当于什么什么斜率项斜率项，那我们在计算里面是不是就是一个差值的比值，所以，是这个学生学习成绩的变化除以这个师生比的变化，吧？也这个也是我们这个估计参数这样一个解释的意义。就我们这个贝塔一含义，它就是什么？师生比变化的一个单位，会引起学生分数我变化的多少个单位，对吧？就是贝塔一的这样一个含义。

发言人 41:51

如果我们把这些点都标注在上面之后，我们构建这个模型也构建好了，就贝塔0加贝塔EX？那我是不是就要找到一条线？就我不能你说我作为校长，可能我也没有啥文化知识，因为看你给我这些点干啥，你就直接告诉我，我到底这个学校到底要招多少老师，你看我一个班定多少个学生合适，你现在给我一点点，我也看不出啥头头荡荡来，那我们想要得到的我们增加一个老师能提高多少成绩，那这个时候我们是不是就要把这些点抽象成为我们这个回归模型，或抽象成一条线出来表示它。

发言人 42:45

抽象成一条线我们就不把所有点画出来，就我们这个牙齿摘取一些点表示的是什么？我们这个什么样本值获得真实的这样一个数据值的点。我们什么回归出来的这条线呢蓝线就是我们刚刚这个同学讲的，他是一个什么样本回归线。那我这个样本回归线距离什么，我的这个什么真实值，大家可以看到它是不是存在着一些差距，对吧？那这个差距大家可以看到，如果说老师没往后讲，大家可以想象一下，我要有一条线能最好的去代表这个图里面出现的所有点。大家可以想到在这条线是具备一个什么样的性质？发散思维一下，看一下大家有没有成为统计深圳跟统过去统计大佬的想法。

发言人 43:52

陈小时。就比如说我现在就就就就有这些点让你来画一个曲线，你怎么样能确保这条线是最好能就能代表这些品最好的那条线呢？它是线性的，你不能把它全连起来。

发言人 44:21

什么叫分享？

发言人 44:33

你的是尽可能我穿过更多的点对吧？那有没有同学有别的想法？白雪，你同意刚刚这位同学说的。同意。好，那请坐。如果说是按刚刚陈小池同学讲的意思，那我是不是就是比如说我在所有的这个是吗？所有的这个嘛我们的这个阿法相当于是拮据项目。我给定所有的阿玛的值，给定所有贝塔的值，它相当于是你在一个圆里面，你找所有的线，你把每一条线都试一下，然后你看它是不是，然后你再计数它每一条线穿过的点越多，那就说明我这条线是最好的，就如果按刚刚陈老师这样的这个行业，你可以这样做来找到这条线？

发言人 45:38

但实际上，我们也有更好的办法去处理这样一个。大家可以看到我是不是因为你想你比如说你要在一条圆里面，相当于要是无数的线，对不对？你这条线是这条线也要试，这条线也要试，然后每条线还要数它经过多，而且真的是我经过的点越多，我的代表性就越好吗？比如说我们得到的点是有可能也是这样的一些情，这些点都在这里，但是我也有些点是在这里的那我这样的线和这样的线也有点在这里，那可以想象到这条线肯定经过的点更多，对吧？但是这条线也有可能它是一个最优的对吧？因为这条线它可能都把后下面的线给落了。所以。

发言人 46:37

我们在考虑这个问题的时候，它其实可以简化成一个什么。大家可以看到如果我在这个图上画一条线的话，是不是每一个点跟我这条线都有一个距离，对吧？那如果说我这条线是最好的点，那就说明什么？我这条线跟每一个点的距离加合起来，它肯定是最小的对吧？那加合起来，大家从这样从现在这个图里面可以还有一个？有个值我减去我的这个图画的线，预测出来的这个值它是正的。但是有的结果减去我预算出来它是负的对吧？

发言人 47:19

那如果说我要算出那个距离所有真实点距离所有真实点这个距离最短最短之和最短的这条线，那我是不是还要把它加绝对值，对吧？大家绝对值的运算，大家可以想要来了很快的对吧？你大家计算机都知道，你很少会有什么绝对值的运算的公式在里面。

发言人 47:49

所以我们更理想的一个方式？我要去掉那个副值？让他去表示这个距离。然后我最简单的方式是不是我就给它取一个平方项，取个平方项，你这样负值和正值的距离差它就不会相抵了，它就变成什么X1Y1距离的这一段的平方，再加上X2和Y2距离这个线段的这个平方的值，对吧？那这个时候，大家可以看到，我就变成了一个什么最小凹型估计的这样一个最什么最底层的这样一个逻辑了。

发言人 48:30

这样一听是不是就很简单？那我们在进行估计的时候，就可以通过这样一组。已知的这样一个样本观测值，然后利用它去构建这个什么样本回归函数，去尽可能接近的去估计我们的这个总体回归函数。什么叫做最好的估计？它的含义就是我们刚刚讲怎么样找到一条最好的线去代表这些点对吧？那你找到这条最好的线，我们刚刚也讲过，那最好的方式是吗？就是我们找到那个距离所有点最短的有这样一条线，那它就是最好的去代表时间点。

发言人 49:22

因为距离你化验的绝对值的运算是很复杂的那它就变成了一个什么？你把那个差值求一个平方，把那个正负号的符号给去掉。那这个时候它就不存在相互抵消的问题了。所以就简化成了这样一个优化的这样一个问题的函数。

发言人 49:42

用什么我们的真实值减去我们样本回归函数估计出来或者说预测出来的这样一个值外貌，还是我们用什么样本函数，就是这个蓝线估计出来的这样一个外貌。它俩的差值的平方，然后所有的点我们给它加全起来if I等于一到N我们总共有算法N个样本的点对吧？那这样一一一求和我们就计算它的什么最小值，从而让我们找到一个什么这样一个最优的这样一条线，或者最好的这样一条线。

发言人 50:25

在这样一个优化的这样一个目标函数里面，然后在这样一个优化的问题里面，它里面可以变的这样一个控制变量？我可以调整的是吗？这样是不是就是我的这个截距项和我们这个什么和我的这个什么斜率对吧？也就是在我们这我们把这样一个要么一个函数预测的外贸的值带进去。大家可以看到是不是就相当于我在这个图上不停的试这个什么，你往上挪就相当于是什么贝塔取一个稍微大一点的值，然后再调整一下它的角度，是不是就是那个贝塔一，你给它变一个稍微小一点的值或者稍微大一点的值。然后这样每一个人去求它的这个距离，算出来，找到一个最合适的那个码，贝塔0和贝塔一来让他这个距离最小。所以这个什么一元信任回归模型，我们的叫什么最小二乘估计？

发言人 51:34

我们求解的这样一个形式，大家仔细想一下，是不是就是把我们的这个目标函数求什么一阶导，然后让这个一阶导等于0，是不是就可以求到两个正规方程，就说我们这个一阶最优的条件就是我们这个？求解的这个就是我们高数里面学过这个相应的内容，对吧？比如说我要求这个什么最小值，那我是不是就是对他求偏导。对那个什么其中的两个参数，一个是贝塔0，一个是贝塔一对吧？我把这个贝塔零求偏导让它等于0，然后对贝塔一求偏导让它等于0。然后求偏导的只有我们得到两个正规方程。正规方程估计完了之后，我们可以看到，我们这个是不是就可以从这两个等式我们就可以求解出来这个威海零貌威海一貌它的这样一个表达式。

发言人 52:48

为了衣帽，大家可以看到我们这个式子，它就可以表示成什么？我们这个XI减去X均值的平方求和分之什么？XI减去X均值乘以YY减去Y的冠军值，它俩相乘求和。在这里面它的含义是什么？

发言人 53:18

X大家可以想象XI减去均值乘以Y减去Y减？我们变成样本协方差它那下面是什么？下面是就是X的这样一个方差的形式，对吧？如果每一个增值减去平均值平方求和，那现在就是方差的这一个。贝塔零它就是什么？Y的均值减去贝塔一帽固定出来的乘以，这个就是这样一个代进去它就得到了。这个证明过程，大家可以回头下去，自己感兴趣的话，我们下去可以自己去推导一下这个过程。这个在这里老师就不再赘述了。

发言人 54:12

我们是不是就可以通过通过给大家讲解这样一个逻辑，怎么样找到一条最好的线，或者说找到一个让我的什么目标函数，就是YI真实值距离预测值之和最小的那一点，就是我们最好的什么贝塔0和贝塔一的这样一个参数多界值。那这个时候就要这样一讲，你也就理解了什么叫做为什么它叫做什么普通最小二乘法。它这个二乘的含义，它为了去解决什么正负号距离相抵的问题，那它乘以那个平方，所以它就是我们要得到一个这个距离平方最小值，就它叫什么普通最小二去估计。这样我们就得到了我们模型里面的这个什么贝塔0和贝塔一，模型的参数是不是就估计出来。这样估计出来的值就是最好的那个值，就是说最能代表我这个样本点这样一条线。这里面贝塔零它其实就是什么我们的截距项，贝塔一其实就是我们频率项。然后通过N组的或者N个样本值我们求解得到，然后以它们来作为什么总体的截距和总体斜率的这样的一个估计。

发言人 55:49

大家可以看到，如果说什么X和Y它是什么正相关的话，这一项关键的表达公式大家里面都学过对吧？还有协方差，如果说X和Y它的协方差就XI减去Y均值乘以Y减去Y均值，它俩相乘。它的这个什么正相关的话，那大家可以看到这个，贝塔一它也是正相，它也是正的。如果说X和Y是不相关的话，就是大家可以看到相关性的那个表达式，COVXY他是负的，他给他了，他给他一，他肯定也是负的对吧？那也就是说如果我们在回归模型里面，它存在一个什么这样一个正负向的影响关系的话，那它俩肯定都是正相关或者负相关的。然后求证的过程和推导的过程，大家可以下去自己感兴趣的时候就好推一下。

发言人 56:57

因为课堂我们的时间比较有限，我们就继续往下。刚刚给大家讲解的是啊一元线性回归模型，刚刚是不是也讲了多元线性回归？多元线性回归模型它它其实就是我们在这样一个多重回归的研究里面，我们把其他的变量也考虑在里面。这个时候根据线性代数的内容，我们这样一个Y等于贝塔0加，贝塔一X1加贝塔32的这样一个多元线性网络模型。那它就可以写成这样一个线性代数，这样一个表达的形式。大家可以想象到它中间是不是会有一个举证的内容，就是我们要估计的参数就不是一个单个的参数了，对吧？

发言人 57:53

我们要估计的是一组参数，那这一组参数就是我们的贝塔表示出来它是有截距项，贝塔0，然后有X一的这样一个影响参数贝塔一，然后一直到贝塔K这些，这组参数都是我们估计的那同样Y和X作为我们的自变量和应变量的，它也可以写成这样一种矩阵的形式，然后表达式就是Y等于XK。多元线性回归模型，它最小二乘，它也是这样一种呼吸方式，能写成矩阵的方式。矩阵的形式你就可以把它来进行这样一个相应的求解，也就是我们这个什么残差项平方和残差项相乘，然后他们我们来求这样一个最小值，对吧？通过推导我们就可以得到相应的这样一个多元线性回归模型这样一个系数贝塔这样一个向量值。下面我们来看我们这个最小二乘回归得到这样一个估计的结果。也就是说我们得到最好的这样一个结果，它存在哪样一些性质？我们先休息一下后面。

发言人 01:07:27

好的，那我们继续看，就是我们通过这样一个什么求得距离平方和最小的这样一种估计方法，也就叫最小二乘法估计出来的这样一些这样最好的参数，贝塔0和贝塔一它有哪些一些性质。首先我们根据最佳邀请法得到的这样一个样本回归线，他肯定是什么穿过这个样本的均值的。也就说比如说我们拿我们班的这个同学这样一个数据拿来分析的话，这种班平均的收入和平均的消费的这个点，它肯定是在我们估计出来的这条线上，就估计出来的这样一条样本的线上。那么这个是可以通过什么？我们这个最小二乘法，关于这个贝塔零的这样一个一阶最优条件去得到这样一个求导出来，它就得到了这样一个Y均值等于贝塔0加倍的12X均值，对吧？

发言人 01:08:50

除此之外，我们这个最小层估计出来的这个残差，它的均值也是0。也就是我们刚刚那个图里面是不是我们这个真实值距离我们这个样本归线它有一些距离值，对吧？我买这个绝对值不求不求和，也不做什么绝对值处理，就把他们全加在一起。

发言人 01:09:16

那这个时候他的这个残差均值它就是0，那意味着什么呢？就是我们最好去代表的这条线呢，有一些值是高于我们这个样本库，有一些点还是低于样品文件。所以它你每一个真实值Y的点，Y的真实值减去外贸这个预测值，我们不求平方的话，它的均值就是0。这个也很好理解，它是由什么我们这个最小的关于贝塔零的一个一阶最优的条件也是可以得到的那对于这样一个结论，大家得思考一下。比如说如果说对于我们这个什么不带常数项的这样一个回归，那我们这个残差均值为零，它的这个性质还存在吗？大家可以看看。

发言人 01:10:35

韩某。你觉得他这个长度像回平，那就不再尝试这样这个回归它是不是也存在这样的一个性质？残差这样一个均值等于。

发言人 01:11:07

就不再尝试这样的回归是什么意思？是不是我们就是Y等于什么？本来是什么？Y等于阿尔法加贝塔X对吧？我不在常数上就变成了什么形式。Y等于贝塔X对吧？那这样的形式它还存在这个性质吗？就是我这些点距离我的样本不会线，那个距离我全给他加在一起求个均值，他还是理由吗？是对吧？

发言人 01:11:49

好，请坐。还有没有同学有别的意见？等于慢。同意刚刚那个同学的想法吗？同意，就是他那个点之后跟那个线求和起来，它也是零，对吧？好这个。郭子旭，同意刚刚他俩的他们。

发言人 01:12:23

也是同意是吧？好，行。于博陈。你同意跟那三位同学的看法？也好，你们不会是什么搞什么心理学，什么羊群效应，对吧？一直跟着徐浩正。徐浩杰也同意。周林芊芊。

发言人 01:13:12

小木也。

发言人 01:13:21

付晨阳。

发言人 01:13:28

也同意是吧？好，谢谢王建芬。

发言人 01:13:38

夜宵了。

发言人 01:13:45

能不能无脑跟随吗？刘文博。写作业。都到点名了，黄诗怡。

发言人 01:14:10

杨启超。同意吗？你同意还是不同意？

发言人 01:14:25

赵伟杰。我这不能当社会体验，就是人越多，反正我就无脑跟车管，你们都不知道，老师点名就是为了得到一个不同，就没有人我这过。

发言人 01:14:47

也同意。张子贺。也同意行，那个我不点了，这条件都给大家写出来了。就是可以看到我们得到这个性质，通过什么条件得到，是不是通过对截距项求一阶偏导得到的那如果我拮据上都没有了。我点到你了，那可能就是可能全班得轮了，还得轮到一个不同意，这条件不是写在这里吗？大家可以思考一下，就是如果说我贝塔零都不存在了，因为他没有尝试像这个回归模型，它对它灵都不存在了，我就更别说这个偏导得到这个值，就得到这个性质。所以如果说我们这个回归模型不存在常数项的话，那我们每一个点，你对什么二层回归的估计出来的这个码，那么回归线它的这个残差你求和起来，它就不存在这个等于零的性质了，那就有可能是不等于。

发言人 01:16:08

这个是大家在以后构建模型的时候需要注意的。因为很多时候通常都是默认一个函数或者说一个模型计量模型，它都是有这个长度项的。就比如说我们这个什么收入的消费问题里面，拮据项的含义是什么？或者常数项的含义是什么？

发言人 01:16:36

过生活消费的这个问题里。刘思雨收入消费这个问题里，我们这个拮据项的含义是，我说我能不能不要这个减去项目做常数项，为什么？那相当于是一个启示是吧？对，就是我即使收入为零我也得活着对吧？对，就所以大家在构建模型的时候，虽然我们通常都是默认这个函数或者这个模拟它是有截距项的。

发言人 01:17:19

但也有也有时候你在一些问题的时候，你可以不设这个场。比如说刚刚刘书记讲的这个收入的问题里面。因为如果你没有你，你没有收入，收入为零，你是不是也得活着？就是你也得保证你的基本生存，对吧？你也有没有你也必须得要有消费支出。你说假如说你每天一天三顿吃馒头，我买到了有五毛钱，一天一块钱，那就你基本就保证维持。还有什么衣服什么的，可能比如说也要有一个150、160对吧？你得要保证你基本活下来的这样一个需求其实没有错误。

发言人 01:18:01

所以大家关注到这样一个情绪上的这样有很多时候你设定这个有截距项和设定没有截距项，它这个模型也存在比较大的区别。很多时候因为我们我们这个什么计量不能实在朋友们去关注，而变量之间的关系因果关系，它很多时候就会把这截距上或者长度上忽略掉。很多时候大家也需要注意，这个截距上，它在特定的问题上它也是有含义的。有的时候你可以不用，就是我们没有这种相应的问题的话，你可以把这个常数项或者定项去掉。很多时候它有含义的话，你就不能把它去掉。

发言人 01:18:52

下面我们看这样二层回归，它只在第三个性质，就是我们这个回归变量与ONS残差之间，它的这个样本协方差是为零的。也就是说什么我们这个回归变量它跟这个残差之间它什么是不相关的对吧？那这个我们就可以推导，也是通过这个最小二解解法里面，我们求最优目标函数的时候，应该进行一阶的求解，得到我们这个XI乘以缪I的这样一个估计值，它两个之间的协方差是在零的，也就意味着什么？我们的这个，如果它是一个最小二乘，估计它是一个好的最好的这样一条线。它也有一个性质，就是我们这个模型里的X跟你没有考虑到的其他因素加和其它都是不相关的对吧？它是一个协方差的，以协方差为，也就意味着比如说我们在考虑收入消费的问题里面，比如说你的这个性别，或者说是这样的问题，这样的变量你就不考虑了。它有可能也是一个关键的变量，被你放在了这个码随机误差项里面，也就在缪安里面。

发言人 01:20:20

那如果我们这个治疗能估计出来，它就有一个性质是吗？我们这个XI也就你的收入跟你没有考虑其他因素，他俩之间他们之间就打包起来的这个因素的几何，他们之间的关系是为零的那如果说不为零，那这样的估计是不是就不是一个好的估计了？说它可能还会我估计出来这个参数贝塔一，因为我的模型是什么？Y等于8，Y等于0。加缪。如果说我们这个X和6之间还有一些关系，那大家就可以发现，有可能我估计出来这个V很低的值有一些漏了对吧？也有可能是在缪这里？

发言人 01:21:16

假设说这个X跟Y跟缪之间会存在一个相互影响关系的话，就是说你直观的就能感觉这个贝塔尼估计出来什么不准不好。那这个是我们如果是不想要成回归，我们估计出来结果它就有这样的一个性质。它的残差项之间，它的什么样本方差为这个也是后续如果我们做这个估计好坏的时候，大家可以去检验，尤其是采纳项可以算出来的对吧？你的这个采纳项目的值，你回归模型构建完了之后，你每一个真实值距离这个样本回归线的点的距离就是EI的这个值。每一个其实都可以算出来，就是它相当于是不是也是一组数了。那这种数跟你XI它你是不是就可以求一个相关性的值。那它俩你要求它俩如果是显著不相关，那是不是就是一个比较好的结果？

发言人 01:22:27

如果说它俩存在相关，那这个时候你的麻烦是不是就大了？那可能你就要去调整模型，换变量，或者说去做一些其他的处理了。我们看这个二层，这样几个性质综合来讲就是什么？首先我们这个样本回归线经过均值的点对吧？这个也是今天第一次给大家讲的是我们不管是实战的，其实研究的结果它都是买一个均值的概念。那它估计出来这条线呢肯定是经过把样本的均值点，而他残差的均值是等于零的，它估计值的均值也等于这个观测值的均值采纳项EI与XF相关，咱家像EI跟外貌也不相关，就跟我估计出来这个外貌也不相关。所以大家可以看到，基于上面这样一些性质，我们就不需要考虑这个数据的生成方式。所以它就被称为最小号成估计，最小二乘估计量。它的这样一个数值的性质就是我没有数据的时候，它也存在这样一些基本这样一些数值的性质。

发言人 01:23:59

代入数据我们就可以看它这个什么对角方程估计，它也存在一些统计的性质。首先是它的线性性，线性性是什么意思呢？它的意思是我们这个贝塔容貌是我们估计出来的这两个参数，那是否是这个什么？YI它的线性函，你说它是不是这个水晶面料外还有这样线性函数。我们今天是不是也讨论了什么，我们什么样叫做一个线性函数，对吧？如果说它是一个线性函数的那它就具备这个线性性。

发言人 01:24:46

其次我们看它是不是存在这样一个无偏性，无偏性的含义是什么呢？就是我们估计的这样一个参数的估计，这样的他的期望值是不是等于总体参数，它的真实值，也就是我被他冒他的期望是不是真实值有效性还是指当我们这个参数估计量在所有的这样一个线性无偏估计量中，他还要有一个最小的方差性。如果说你估计出来这个值是一个无偏的，也就说它是等于真实值。但是你估计的好坏是不是也有一个检验标准，就是你可能在这个什么很大的范围内，它都有可能是对的。

发言人 01:25:39

比如说你得到的这个什么收入对消费的影响是0.6，那这个0.6大家给我统计就知道它是一个什么均值。然后因为我什么估计出来的这个YI悲惨零和悲惨一，它都是什么随机变量。所以它比如说你估计出来个0.6是合理的那你估计出来0.67是合理的吗？估计出来0.68是不是合理的呢？或者说你不会穿0.9是不是合理？它因为它是一个随机变量，所以大家学过那个性质31个。如果说你的方差很大的话，就有可能是什么？估计出来0.6是合理的，估计出来0.9也是合理的那这样的话你估计出来的这个含义是不是就不不不不就就就就就不太好了，对吧？

发言人 01:26:37

因为我们想得到，比如说0.6是合理的，比如说你的这个加减0.01这样一个范围内变动的这样一个参数，我们想估计出来一个像那好啊，但是结果你估计出来什么？0.6加减0.3，这也就意味着什么？0.3到0.9都有可能是你估计出来的这个参数值。那这样的话，是不是就不是一个好的估计了？所以我们在检验这样一个最小化程它是不是好的路径的时候，我们也要看我的这个天窗固定量，我也要让它方差尽量小一点，尽量小一点，那它就是一个有效性的这样一个感觉。这些性质它就是最想要成估计量这样一个小样本性质。如果说满足这些性质，它的估计量它就称为这是什么最优线性无偏估计量，这个也就是高斯马尔定理讲的内容。在这个最小二乘估计统计性质里面们也思考一下这两个问题。一个是为什么OS估计量它是贝塔0和贝塔0，贝塔0和贝塔一，我们估计出来的这两个参数值它是或者这样这两个参数它是随机变量的。

发言人 01:28:58

榛子。

发言人 01:29:04

为什么？为什么我们OIS估计出来的这个参数贝塔零贝塔一，它是这个随机变量的。

发言人 01:29:23

就是我们之前讲过这个。

发言人 01:29:37

这样一个方程里面。X是什么？X这边它是固定的还是随机的？它是确定的，所以是固定的对吧？如果给定X对它是固定的那Y是什么呢？充电量它是随机的还是固定的，因为它是什么随机的对吧？那缪呢？它叫随机误差项，对吧，他有随机的那这里被害人和被告1？

发言人 01:30:25

那是随机的，是随机变量吗？是吗？是的，对吧？为什么？

发言人 01:30:47

对，这是好现状。那就是说我们这个什么随机变量，它是跟什么我们这个数据相关，对吧？它是由我们数据估计出来的这样一个就是它具有一个什么随机变量的这样一个性质。

发言人 01:31:07

其次我们第二个问题就是为什么我们要去研究OLS估计量它的期望和方差呢？就是我们估计出来的这个什么OS估计量是什么？贝塔0和贝塔一，为什么我们要去看他的期望方差的？

发言人 01:31:34

心意。

发言人 01:31:41

希望是在协。

发言人 01:31:49

就相当它范围对吧？变动的范围好几个。实际上我们这个估计量长是什么？看它的什么一阶段、二阶段、三阶段、四阶段对吧？那一阶距它起码最常用的，我们就是这个，一般来讲我们最常用的期望和方向，它衡量的贝塔0和贝塔一，它这两个什么随机变量的什么中心值这样的中心值和你什么变化的这样一种程度。我们就可以用这个期望和方差点，我就可以给它体现出来，我们也回到这样一个例子，我们现在就终于找到了这样一条好的蓝线对吧？然后来去代表所有新样本值的这样一些点。所以说我们这个师生比和学生考试成绩之间的关系，我们就得到了。

发言人 01:32:51

假如我们估计就用刚刚学到的这个什么求每个点到线距离之和最小的这样一个目标函数，对它进行贝塔01阶求导的，贝塔二一阶求导等于0。然后两个式子一求求出贝塔零的这个Y均值减贝塔12X均值。然后贝塔一等于什么X和Y的协方差除以X的方差。得到了这个什么贝塔0和贝塔一之后，我们是不是就可以把这个值带进去，把这条线画出来了。画出来了之后，我们得到了这个贝塔一的值-2 028。那这个-2.28它表示什么含义呢？它表明了我们什么声势比每增加1，平均来讲我们的考试成绩会降低2.28个单位还要把学生放在更多绅士，比他们叫师生比。那绅士比如果提高，还有这个学生成绩就下降2.28个点。

发言人 01:34:02

那么这个负的2.28，它实际上是什么？你估计出来。那显然很显然，它这个值跟你的这个什么真实值臻或者说真实的这样一种规律，它是不是可能他之间是会存在一个差异的。所以我们通过现在已经通过什么这条就是求解，我得到了这个数，那我怎么样来评判这个数，这是对的。怎么评判我们得到的这个什么2.28，-2.28它就是一个正确的职能。或者说看这个-2.28是不是很近似的去会很好的去估计近视的我们这个真实的这个贝塔一，我们就引入了刚刚这样一个定义，对吧？它叫做无天性。

发言人 01:34:57

所以说我们统计学里面，大家都讲过了？我们这个呼吸量它的气泡就等于我们这个杯贝塔一贝塔一害。我说我们这个贝塔一害，它作为一个平均行为的一个描述，它具体的某一个特别的或者说某一组它的数据，你可以算出来这个贝塔旱它有可能是偏离我们这个贝塔。那无偏性它就是什么？我们得到了这个贝塔12.2-2.28，它是等于我们什么真实的贝塔因子。同样的，贝塔零也是我们期望它的什么估计出来或者预测出来这个值它是等于真实。

发言人 01:35:50

是的，我们在建立这样一个无限性的时候，也需要满足一些奖励。首先就是模型对于这个贝塔零贝塔一来讲，它都是什么线性的。假设二是我们这个样本来自于总体，而且它是一个什么随机的样本。假设三是我们这个X的样本，它的结果或者说XY它不是完全相同的数值。就有假设说现在大家家长都串通起来了，这个全班统一，就可能有家长联盟什么的，大家都防止大家攀比或者说过度消费，那统一给大家一个月都是1000块钱，大家就可以看到我们什么。

发言人 01:36:45

如果说我们构建的这个模型里面Y等于这个X我这个X大家家长串通了，每大家我收集的什么班加上50个人出去，每一个人负X都是1000。那这个时候大家可以想到我这个模型还能估计吗？就估计不了了，对吧？那个X都没有变动的，那个变动的信息让我放到这个模型里来求解，对吧？所以我们也需要讲一个什么解释，变量它的样本还是有变差的，不能完全一样。

发言人 01:37:24

假定四条件均值这样进行，假定说我给定对吧？就我现在知道大家收入了这个已经是确定的了。这样一个信息改进之后，我这个缪就是我的随机扰动项，我这个期望是零。你说我其他因素或者说其他的误差加合起来是为零的。

发言人 01:37:51

下面我们去看这个无偏性的这样一个表达，或者说一个好的这样一个估计出来的贝塔一的那他要表先表示可以先把它表示成这样一个随机变量的形式，然后这样你就可以看出来它为什么是一个这样一个随机变量的表达。因为它一号这个式子刚刚我们是不是给它通过这个最小二乘估计，我们给它写出来，对吧？它是什么？X和Y的协方差除以X的方差这样一个形式。而且我们通过这个什么分子里面的这样一个相乘，我们把这个YI给它带入进去。那是不是我们就有什么乘以这个贝塔0加贝塔EXI加缪I的这样一个形式。我们再进一步的去把它这个分子拆解出来，我们就可以得到威海明乘以XXI减X均值的这样一个表达式。

发言人 01:39:03

再加上A乘1乘以XXI减XXX均值乘以XI再加上最后什么XI减X均值乘以6X所以我们把这个什么分子里面，我们用Y两个数据去除了之后，就可以把它分子分解成这样一个形式。因为我们什么之前得到的什么XI我们可以看到XI减去X均值的话，它求和起来是不是等于6，对吧？所以我们相乘就可以得到。那这个时候所以他衣帽就可以写成马贝塔一真实值。也就是说我们这个data 1 head，就是贝塔一貌似是什么？

发言人 01:40:00

我们通过样本数据得到了这样一个产出的贡献值，它可以表达成贝塔一真实值，加上我们后面这个尾巴的形式这个尾巴形式是不是X和缪的协方差除以什么X的方差的形式？那如果说我的这个什么贝塔一的期望等于贝塔一，那是不是就相当于什么X和缪，它的什么协方差等于0，所以这个时候它就是一个无偏性。同样我们也可以看到随机的变量，也就是说同一组的XI不同的缪I它也会导致不同的YI对吧？所以不同的组合XI和YI就会得到不同的这贝塔0和贝塔一，它就是个随机变量。好，那我们这节课。

发言人 01:41:49

这证明的话，大家可以看一下，大概要就会。好像都不低。因为去年我还想着前年给大家出简单了，去年增加了考题，也增加难度，吃不透，对吧。但我感觉不会很难，因为我们我们这个课时也少，就不方便到大家去很多证明数据。

发言人 01:42:29

感觉我的商品链接115块94。我去年是考了好多，其实考的挺好的。因为我增加难度，也增加了那个14题。因为反正大家都行了，因为大家太嘚瑟了，都是结果还都考的挺。

发言人 01:42:54

你好。对。

发言人 01:43:04

没有开玩笑。然后另外一个就是。去那有有U盘吗？那你们谁U盘？有U盘的肯定开了，你自己的没有？而且我都拷给你了，我直接拷给你们，不是直接拷给我们。你可以上传到学妹系列的那个资料里，你不知道。就是你可以上传到这儿。你们用教材了吗？没有。

发言人 01:43:52

以前的教材的。然后你可以穿上这个，而且我们也没有发这个。然后把资料传到这儿，或者是直接查到章节，电脑有个网页，手机上有啊。对，然后电脑的话你直接收学费，学费也直接收学费期间就行了，他会有网络。