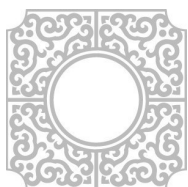


第一章



统计预测概述

第一节 统计预测的概念和作用

一、统计预测的概念

预测就是根据过去和现在估计未来、预测未来。统计预测属于预测方法研究范畴,即如何用科学的统计方法对事物的未来发展进行定量推测,并计算概率置信区间。在这种推测中,不仅有数学计算,而且有直觉判断。统计预测的方法论性质与统计学的方法论性质是一致的。

统计预测方法是一种具有通用性的方法。实际资料是预测的依据,经济理论是预测的基础,数学模型是预测的手段,它们共同构成统计预测的三个要素。统计预测可用于人类活动各个领域中的实质性预测。例如,用于预测经济的未来,就是经济预测;用于预测人类社会的未来,就是社会预测;等等。明确统计预测和各种实质性预测之间的联系和区别是十分重要的。下面以统计预测和经济预测为例,说明两者的联系和区别。

两者的主要联系是:

- (1)它们都以经济现象的数值作为其研究的对象。
- (2)它们都直接或间接地为宏观和微观的市场预测、管理决策、制定政策和检查政策等提供信息。
- (3)统计预测为经济定量预测提供所需的统计方法论。实践证明,如果没有科学的统计预测方法,经济定量预测就难以取得迅速的发展和较准确的结果。同时,统计预测也对经济预测结果的准确性进行研究,以便使预测方法得到不断完善。

两者的主要区别是:

- (1)从研究的角度来看,统计预测和经济预测都以经济现象的数值作为其研究对象,但着眼点不同。前者属于方法论研究,其研究的结果表现为预测方法的完善程度;后者则是对实际经济现象进行预测,是一种实质性预测,其结果表现为对某种经济现象的未来发展做出判断。
- (2)从研究的领域来看,经济预测是研究经济领域中的问题,统计预测则被广泛地应用于人类活动的各个领域。

为了使统计预测和决策方法在社会主义市场经济体制下发挥更大的作用,本书在阐述方法后,列举了大量的经济预测实例。

二、统计预测的作用

在市场经济条件下,预测的作用是通过各个企业或行业内部的行动计划和决策来实现的。预测与决策和行动计划之间的关系在于:预测在决策之前,行动计划在决策之后。预测为决策提供依据,是决策科学化的前提;而正确的决策又给合理的预测提供实现机会。行动计划是预测、决策之后的产物,又是预测、决策实现的桥梁。预测人员是情报和信息的生产者,而决策人员和计划人员则是情报或信息的消费者。

统计预测作用的大小取决于预测结果所产生效益的多少。影响预测作用大小的因素是多种多样的,主要有:(1)预测费用的高低。预测费用包括设计预测程序费用、资料搜集和整理等调查费用、资料使用费用和计算费用以及研究人员的劳务费用等。显而易见,费用的高低直接影响了预测结果效益的好坏。(2)预测方法的难易程度。它与预测费用的高低有着密切的联系,如方法简单易懂,则费用就低;反之,方法复杂难用,费用就高。(3)预测结果的精确程度。通常情况下,准确性高的预测比准确性低的预测作用更大一些。也就是说,花费更多的时间和金钱有可能得到一个较好的预测结果。但是,是否值得花这部分额外的代价去取得额外的精确性,是需要考虑的。虽然有办法去评估所提高精度的价值,但在一种特殊的方法被做出之前,要想知道它究竟能在多大程度上提高预测精度是困难的。同时,还要注意到提高精度的好处常在于减少风险,而不是降低费用。由此可见,预测费用的高低、预测方法的复杂程度以及预测结果的精度是影响预测作用的三大主要因素。

就统计预测方法而言,其最基本的作用在于把历史资料中同时并存的基本轨迹和误差分开,以研究其形态的变化。把轨迹分离出来的办法,就是对资料拟合某种模型,使模型尽可能准确而全面地反映出有规律的轨迹。误差又称为残差或剩余项。残差必须呈现某种随机性。研究残差的随机性是统计预测的一项重要内容。

第二节 统计预测方法的分类及其选择

一、统计预测方法的分类

按预测方法的性质,大致可分为定性预测法、回归预测法和时间序列预测法三类。

(一)定性预测法

定性预测法是以逻辑判断为主的预测方法。这类方法主要是通过预测者所掌握的信息和情报,结合各种因素对事物的发展前景做出判断,并把这种判断定量化。它普遍适用于对缺乏历史统计资料的事件进行预测,或对趋势转折进行预测。具体方法有德尔菲法、主观概率法、领先指标法、厂长(经理)评判意见法、推销人员估计法、相互影响分析法和情景预测法等。

(二)回归预测法

回归预测法是研究变量与变量之间相互关系的一种数理统计方法,应用回归分析从一个或几个自变量的值去预测因变量的值。回归预测中的因变量和自变量在时间上是并进关系,即因变量的预测值要由并进的自变量的值来旁推。这类方法不仅考虑了时间因素,而且考虑了变量之间的因果关系。具体有一元线性回归预测法、多元线性回归预测法、非线性回归预测法等。

(三)时间序列预测法

时间序列预测法是一种考虑变量随时间发展变化规律并用该变量以往的统计资料建立数

学模型做外推的预测方法。由于时间序列预测法所需要的只是序列本身的历史数据,因此这类方法应用得非常广泛。具体有时间序列分解分析法、移动平均法、指数平滑法、趋势外推法、自适应过滤法、平稳时间序列预测法、灰色预测法、状态空间模型和卡尔曼滤波等。

(四)统计预测方法小结

实际应用中,往往各种预测方法交叉运用、相互渗透,很难做出截然的划分。因此,对上述分类不能绝对化。例如,回归预测法和时间序列预测法的共同特点都是偏重于统计资料,以便建立数学模型进行预测。我们把以数学模型为主的预测方法习惯上称为定量预测法。因此,统计预测方法又可归纳成定性预测法和定量预测法两类。

按预测的时间长短,可分为近期预测、短期预测、中期预测和长期预测。一般来讲,近期预测是指1个月以内的预测,短期预测是指1~3个月的预测,中期预测是指3个月到2年的预测,2年以上的预测称为长期预测。向未来预测的时间越长,预测误差就越大。

按预测是否重复,可分为一次性预测和反复预测。在根据某种预测模型进行外推预测时,有的模型可以一次计算出所需的远近任何时期的预测值,称为一次性预测,如回归预测法和时间序列趋势外推预测法等;另外一些模型每次只能前测一期,称为反复预测,如指数平滑法、自适应过滤法等。

二、统计预测方法的选择

在选择预测方法时,应考虑三个主要问题,即合适性、费用和精确性。任何一种预测方法都是建立在一定的假定条件之上的,而任何一种假定条件都无法囊括现实世界中错综复杂的关系,因此,必须考虑方法的适用条件。也就是说,一个合适的方法不仅要适合于影响被预测项目的因素,而且也应适合于预测的环境和条件。例如,一种适合于一次性预测的方法可能不适用于库存量控制的预测。在精确性和费用问题上,要权衡两者的轻重,要依靠自己的判断能力去断定应该用多大的力量,从而决定使用哪一种方法。

表1-1概括了各种预测方法的特点,以供事前选择之用。

表 1-1 各种预测方法的特点

方 法	章	时间范围	适用情况	计算机硬件最低要求	应做工作
定性预测法	2	短、中、长期	对缺乏历史统计资料或趋势面临转折的事件进行预测	计算器	需做大量的调查研究工作
一元线性回归预测法	3	短、中期	自变量与因变量两个变量之间存在着线性关系	计算器	为两个变量收集历史数据,此项工作是此预测中最费时间的的事情
多元线性回归预测法	3	短、中期	因变量与两个或两个以上自变量之间存在着线性关系	在两个自变量情况下,可用计算器,多于两个自变量的情况用计算机	为所有变量收集历史数据是此项预测最费时间的部分
非线性回归预测法	3	短、中期	因变量与一个或多个自变量之间存在某种非线性关系	在两个变量情况下可用计算器,多于两个变量的情况下用计算机	必须收集历史数据,并用几个非线性模型试验

续表

方 法	章	时间范围	适用情况	计算机硬件 最低要求	应做工作
趋势外推法	4	中期到 长期	当被预测项目的有关 变量用时间表示时,用 非线性回归	与非线性回归预测 法相同	只需要因变量的历史资料, 但用趋势图作试探时很费时间
分解分析法	4	短期	适用于一次性的短期 预测或使用其他预 测方法前消除季节变 动的因素	计算器	只需要序列的历史资料
移动平均法	5	短期	不带季节变动的反复 预测	计算器	只需要因变量的历史资料, 但初次选择权数时很费时间
指数平滑法	5	短期	具有或不具有季节变 动的反复预测	在用计算机建立模 型后进行预测时,只 需计算器就行了	只需因变量的历史资料,是 一切反复预测中最简易的方 法,但建立模型所费的时间 与自适应过滤法不相上下
自适应 过滤法	6	短期	适用于趋势形态的性 质随时间而变化,而且 没有季节变动的反复 预测	计算机	只需因变量的历史资料,但 制定并检查模型规格很费时间
平稳时间 序列预测法	7	短期	适用于任何序列的发展 形态的一种高级预 测方法	计算机	计算过程复杂、烦琐
干预分析 模型预测法	8	短期	适用于当时时间序列受 到政策干预或突发事件 影响的预测	计算机	收集历史数据及影响事件
景气预测法	9	短、中期	适用于时序趋势延续 及转折预测	计算机	收集大量历史资料和数据, 并需大量计算
灰色预测法	10	短、中期	适用于时序的发展呈 指数型趋势	计算机	收集对象的历史数据
状态空间 模型和卡尔 曼滤波	11	短、中期	适用于各类时序的预 测	计算机	收集对象的历史数据,并建 立状态空间模型

第三节 统计预测的原则和步骤

一、统计预测的原则

在统计预测中的定量预测要使用模型外推法。使用这种方法有以下两条重要原则:

(一)连贯原则

所谓连贯原则,是指事物的发展是按一定规律进行的,在其发展过程中,这种规律贯彻始终,不应受到破坏,它的未来发展与其过去和现在的发展没有什么根本的不同。

(二)类推原则

所谓类推原则,是指事物必须有某种结构,其升降起伏变动不是杂乱无章的,而是有章可

循的。事物变动的这种结构性可用数学方法加以模拟,根据所测定的模型,类比现在,预测未来。

由以上两条原则可知,统计资料的稳定结构是应用统计预测的必要条件。在准备使用统计方法进行预测时,必须对所占有的大量统计资料进行认真审核,查明有无可用某种模型测定的稳定结构。凡是没有一定结构,或虽有结构但很不稳定,经常出现突然变化的资料,是很难据以进行预测的。

二、统计预测的步骤

一个完整的统计预测研究,一般要经过以下几个步骤:

(一)确定预测的目的

预测的目的不同,所需的资料和采用的预测方法也有所不同。例如,对人民群众生活水平,既可从其收入方面来预测,也可从其消费结构方面来预测,还可从物价变动对其生活的影响程度等方面来预测。有了明确的目的,才能据以搜集必要的统计资料,采用合适的统计预测方法。

(二)搜集和审核资料

准确的统计资料是统计预测的基础。预测之前,必须掌握大量的、全面的、准确有用的数据和情况。为了保证统计资料的准确性,还必须对资料进行审核、调整和推算。对审核、调整后的资料,要进行初步分析,画出统计图形,以观察统计数据的性质和分布,作为选择适当预测模型的依据。

(三)选择预测模型和方法

资料审核、调整后,根据资料结构的性质,选择合适的模型和方法来预测。在资料不够完备、精度要求不高时,可采用定性预测法;在掌握的资料比较完备、进行比较精确的预测时,可运用一定的数学模型,如采用回归预测法和时间序列预测法等。

(四)分析预测误差,改进预测模型

预测误差是预测值与实际观察值之间的离差,其大小与预测准确程度的高低成反比。预测误差虽然不可避免,但若超出了允许范围,就要分析产生误差的原因,以决定是否需要对预测模型和预测方法加以修正。

(五)提出预测报告

即把预测的最终结果编制成文件和报告,向有关部门上报或以一定的形式对外公布,即提供和发布预测信息,供有关部门和企业决策时参考和应用。

统计预测在具体应用中的完整程序如图 1-1 所示。

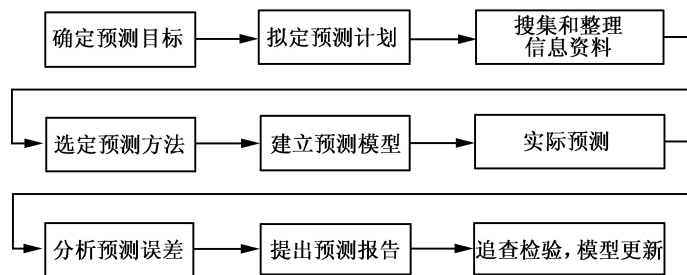


图 1-1 统计预测程序



本章小结

预测就是根据过去和现在估计未来、预测未来。统计预测属于预测方法研究范畴。

1. 统计预测方法是一种具有通用性的方法。实际资料是预测的依据,经济理论是预测的基础,数学模型是预测的手段,它们共同构成统计预测的三个要素。

2. 在市场经济条件下,预测的作用是通过各个企业或行业内部的行动计划和决策来实现的。统计预测作用的大小取决于预测结果所产生效益的多少。影响预测作用大小的因素主要有:(1)预测费用的高低;(2)预测方法的难易程度;(3)预测结果的精确程度。

3. 统计预测方法可以分为定性预测方法和定量预测方法两类,其中定量预测方法又可大致分为回归预测法和时间序列预测法。另外,还可按预测时间长短分为近期预测、短期预测、中期预测和长期预测;按预测是否重复,分为一次性预测和反复预测。

选择统计预测方法时,主要考虑三个问题,即合适性、费用和精确性。

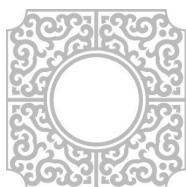
4. 一个完整的统计预测研究,一般要经过以下几个步骤:(1)确定预测的目的;(2)搜集和审核资料;(3)选择预测模型和方法;(4)分析预测误差,改进预测模型;(5)提出预测报告。



思考与练习

1. 什么叫预测?试述统计预测的三要素。
2. 试述统计预测与经济预测的联系和区别。
3. 试述统计预测的分类。
4. 定量预测使用外推法时有哪些重要原则?
5. 统计预测研究有哪些步骤?
6. 简述统计预测的作用和影响预测作用的三大主要因素。

第二章



定性预测法

第一节 定性预测概述

一、定性预测的概念和特点

定性预测是指预测者依靠熟悉业务知识、具有丰富经验和综合分析能力的人员与专家,根据已掌握的历史资料和直观材料,运用个人的经验和分析判断能力,对事物的未来发展做出性质和程度上的判断;然后,再通过一定的形式综合各方面的意见,作为预测未来的主要依据。

定性预测的特点在于:(1)着重对事物发展的性质进行预测,主要凭借人的经验以及分析判断能力。它是一种十分实用的统计预测方法,特别是在对预测对象所掌握的历史统计资料不多,或影响因素复杂,难以分清主次,或对主要影响因素难以定量分析等情况下,定性分析方法将是适用性很强的方法。(2)着重对事物发展的趋势、方向和重大转折点进行预测。它主要适用于下列情况的预测:国家经济形势的发展、经济政策的演变、市场总体形势的演变(如卖方市场向买方市场的过渡)、科学技术的发展与实际应用、新产品的开发、企业未来的发展方向、企业经营环境分析和战略决策方向、企业市场经营组合的改变,等等。

定性预测的方法有很多,但从应用的广泛性、实用性和有效性角度来看,主要有德尔菲法、主观概率法、领先指标法、厂长(经理)评判意见法、推销人员估计法和相互影响分析法等。

二、定性预测和定量预测之间的关系

定性预测和定量预测各有优点和缺点。

定性预测的优点在于:注重事物发展在性质方面的预测,具有较大的灵活性,易于充分发挥人的主观能动作用,且简单迅速,省时、省费用。其缺点是:易受主观因素的影响,比较注重人的经验和主观判断能力,从而易受人的知识、经验和能力的多少大小的束缚和限制,尤其是缺乏对事物发展做数量上的精确描述。

定量预测的优点在于:注重事物发展在数量方面的分析,重视对事物发展变化的程度做数量上的描述,更多地依据历史统计资料,较少受主观因素的影响,可以利用电子计算机对统计方法和数学方法做大量的计算处理。其缺点是:比较机械,不易灵活掌握,对信息资料的质量和数量要求较高,而且不易处理有较大波动的信息资料,更难以预测事物性质的变化。

定性预测和定量预测并不是相互排斥的,而是可以相互补充的,在实际预测过程中,应该

把两者正确地结合起来使用。具体地说,就是在占有比较完备的统计资料的条件下,可以先用一定的统计方法进行加工处理,找出有关变量之间的规律性联系,作为预测未来的一个重要依据。但是,任何数学方法或统计方法的应用,都是以过去的信息资料为基础来预测未来,如果在预测期内发生了重大变化,出现了新的重大影响因素,如政府或管理部门的政策、方针有重大改变,企业的市场经营战略或市场经营组合有重大改变,市场上出现了强大的竞争对手,或出现了过去的信息资料所没有反映的其他重要情况,则对定量预测方法所得到的结果,还要根据以上新产生的因素加以修正。这就需要依靠熟悉情况和业务的人员和专家,运用定性预测方法,提出修正意见。而在使用定性预测方法的同时,也要尽可能地采用数学方法,对事物发展变化的趋势、方向、程度和转折点出现的时间等做出数量上的测算。可见,在实际统计预测工作中,只有把定性预测方法和定量预测方法正确地结合起来,相互补充、相互检验和修正,才能取得较好的预测效果。

第二节 德尔菲法

一、德尔菲法的概念和特点

德尔菲法是根据有专门知识的人的直接经验,对研究的问题进行判断、预测的一种方法,也称专家调查法。它是美国兰德公司于1964年首先用于预测领域的。德尔菲是古希腊传说中的神谕之地,城中有座阿波罗神殿可以预卜未来,因而借用其名。德尔菲法一般适用于长期预测。

德尔菲法具有以下三个特点:

(1)反馈性。反馈表现在多次作业、反复、综合、整理、归纳和修正,但不是漫无边际,而是有组织、有步骤地进行。

(2)匿名性。由于专家是背靠背提出各自意见的,因而可免除心理干扰影响。把专家看成相当于一架电子计算机,脑子里贮存着许多数据资料,通过分析、判断和计算,可以确定比较理想的预测值。

(3)统计性。对各位专家的估计或预测数进行统计,然后采用平均数或中位数统计出量化结果。

二、德尔菲法的预测程序

德尔菲法的一般预测程序如下:

第一步,提出要求,明确预测目标,用书面形式通知被选定的专家、专门人员。在这里,选择专家是关键。专家一般是指掌握某一特定领域知识和技能的人。人数不宜过多,一般在8~20人左右为宜。要求每位专家讲明有什么特别资料可用来分析这些问题以及这些资料的使用方法。同时,也向专家提供有关资料,并请专家提出进一步需要哪些资料。

第二步,专家接到通知后,根据自己的知识和经验,对所预测事物的未来发展趋势提出自己的预测,并说明其依据和理由,书面答复主持预测的单位。

第三步,主持预测的单位或领导小组根据专家的预测意见,加以归纳整理,对不同的预测值,分别说明预测值的依据和理由(根据专家意见,但不注明哪个专家的意见),然后再寄给各位专家,要求专家修改自己原有的预测,以及提出还有什么要求。

第四步,专家等人接到第二次通知后,就各种预测意见及其依据和理由进行分析,再次进行预测,提出自己修改的预测意见及其依据和理由。如此反复往返征询、归纳、修改,直到意见基本一致为止。修改的次数,根据需要决定。

三、运用德尔菲法预测时应遵循的原则

第一,问题要集中,要有针对性,不要过分分散,以便使各个事件构成一个有机整体。问题要按等级排队,先简单,后复杂;先综合,后局部,这样易于引起专家回答问题的兴趣。

第二,调查单位或领导小组意见不应强加于调查的意见之中,要防止出现诱导现象,避免专家的评价向领导小组靠拢,以至于得出迎合领导小组观点的预测结果。如果是这样,其预测的可靠性是值得怀疑的。

第三,避免组合事件。如果一个事件包括两个方面,一方面是专家同意的,另一方面是专家不同意的,这样,专家将难以做出回答。

四、德尔菲法的优缺点

德尔菲法的优点在于:(1)可以加快预测速度,节约预测费用。(2)可以获得各种不同但有价值的观点和意见。(3)适用于长期预测和对新产品的预测,在历史资料不足或不可测因素较多时尤为适用。

德尔菲法的缺点在于:(1)对于分地区的顾客群或产品的预测可能不可靠。(2)责任比较分散。(3)专家的意见有时可能不完整或不切合实际。

五、德尔菲法的应用案例

【例 2—1】 某商业股份有限公司要向外地购进一批新产品,这种产品在本地还没有销售记录。于是,该公司成立调查领导小组,并聘请业务经理、商品专家和推销员等 9 位专家,预测全年可能的销售量。首先将产品的样品、特点和用途做详细介绍,并将同类产品的价格和销售情况做介绍,发给书面意见书,让他们提出个人的判断。经过 3 次反馈,得出结果如表 2—1 所示。

表 2—1 专家意见反馈综合表 单位:台

专家编号	第一次判断			第二次判断			第三次判断		
	最低销售量	最可能销售量	最高销售量	最低销售量	最可能销售量	最高销售量	最低销售量	最可能销售量	最高销售量
1	1 000	1 500	1 800	1 200	1 500	1 800	1 100	1 500	1 800
2	400	900	1 200	600	1 000	1 300	800	1 000	1 300
3	800	1 200	1 600	1 000	1 400	1 600	1 000	1 400	1 600
4	1 500	1 800	3 000	1 200	1 500	3 000	1 000	1 200	2 500
5	200	400	700	400	800	1 000	600	1 000	1 200
6	600	1 000	1 500	600	1 000	1 500	600	1 200	1 500
7	500	600	800	500	800	1 000	800	1 000	1 200
8	500	600	1 000	700	800	1 200	700	800	1 200
9	800	1 000	1 900	1 000	1 100	2 000	600	800	1 200
平均数	700	1 000	1 500	800	1 100	1 600	800	1 100	1 500

(1)如果按照 9 位专家第三次判断的平均值计算,则预测这个新产品的平均销售量为:

$$\frac{800+1\ 100+1\ 500}{3}=1\ 133(\text{台})$$

(2)将最可能销售量、最低销售量和最高销售量分别按 0.50,0.20 和 0.30 的概率加权平均,则预测平均销售量为:

$$1\ 100 \times 0.50 + 800 \times 0.20 + 1\ 500 \times 0.30 = 1\ 160(\text{台})$$

(3)用中位数计算,可将第三次判断按预测值高低排列如下:

最低销售量:	600	700	800	1 000	1 100	
最可能销售量:	800	1 000	1 200	1 400	1 500	
最高销售量:	1 200	1 300	1 500	1 600	1 800	2 500

中间项的计算公式为 $\frac{n+1}{2}$ (n =项数)。

最低销售量的中位数为第三项,即 800。

最可能销售量的中位数为第三项,即 1 200。

最高销售量的中位数为第三、四项的平均数,即 1 550。

将最可能销售量、最低销售量和最高销售量分别按 0.50,0.20 和 0.30 的概率加权平均,则预测平均销售量为:

$$1\ 200 \times 0.50 + 800 \times 0.20 + 1\ 550 \times 0.30 = 1\ 225(\text{台})$$

选择使用平均数或中位数的原则是:如果数据分布的偏态比较大,一般使用中位数,以免受个别偏大或偏小的判断值的影响;如果数据分布的偏态比较小,一般使用平均数,以便考虑到每个判断值的影响。按照此原则,此例采用平均数为宜。

第三节 主观概率法

一、主观概率的概念

主观概率是人们对根据某几次经验结果所做的主观判断的量度。简单地说,就是凭经验或预感而估算出来的概率。一般在专家预测时,对于专家最佳推测的实现可能性,应用主观概率加以评定。

主观概率与客观概率不同。客观概率是根据事件发展的客观性统计出来的一种概率。例如,根据 50 年的统计资料,某地区每年 7 月、8 月、9 月的雨天数分别为 12~15 天、10~12 天、9~12 天。这样,某地区 7 月份的雨天气候概率为 0.4~0.5,8 月份的雨天气候概率为 0.33~0.4,9 月份的雨天气候概率为 0.3~0.4。这个概率对农业生产的安排、建设工程的开工安排等具有重要作用。雨天太多给工程建设、交通运输等带来不便,因此,很需要了解和掌握各地雨天或晴天的统计概率。而主观概率是凭人们某一次或几次经验的特定结果所持的个人信念量度。当然,个人主观概率与个人知识水平、工作经验、判断能力等都有密切的关系。例如,某一业务员根据今年的情况做个人判断,认为某一商品明年销售增长的速度可能是 8%。主观概率是用数值来表明本人对事件发生的可能程度的判断。主观概率必须符合概率论的基本原

理,即:

$$0 < P_i < 1 \quad \sum P_i = 1 \quad (i=1,2,3,\dots)$$

由于主观概率是个人的主观判断,反映个人对某事件的信念程度,因此,以下两点应引起注意:

(1)由于每个人的主观认识能力不同,对同一事件在同一条件下出现的概率,不同的人可能会提出不同的数值。

(2)对于主观概率是否正确是无法核对的。例如,一个企业管理人员说,明年利润上升的可能性是 90%,另一个说,明年利润上升的可能性只有 60%,这时无法判断谁提出的概率是正确的。如果明年利润果然上升了,也不能证明上升的概率是 90%或是 60%。正因为存在着不同个人的主观概率和无法核对主观概率的准确程度,就有必要寻求合理的或最佳的估计概率。因此,在统计预测中,常要调查比较多的人的主观估计判断,并了解他们提出主观概率的依据。

主观概率与客观概率的差别在于它们反映客观实际的程度不同,这种差别只有相对的意义。事实上,我们并不能求得绝对反映客观实际的概率数字,即使是根据过去大量的统计资料或实验数据计算出来的数字,也总是有限的、相对的,不可能包括所反映的全部事实。并且,作为计算根据的资料总是过去的资料,而市场的情况时时刻刻在发生着变化,过去的资料并不能完全反映现在和将来,只能作为过去的演变规律用于判断未来。事物总是有联系的,但又总会有发展变化。许多统计学家认为,客观概率和主观概率是不确定程度由小到大连续排列的一个连续状态的两个部分。因此,估计主观概率的方法,在统计预测中越来越受到重视。

二、主观概率法的预测步骤及其应用案例

主观概率法是一种适用性很强的统计预测方法,可应用于人类活动的各个领域。现以具体案例来加以说明。

【例 2—2】 某商业集团公司打算预测 2016 年 11 月份的商品销售额,要求预测误差不得超过 ± 6 万元。现用主观概率法进行预测。

(一)准备相关资料

将过去若干年该商业集团公司的商品销售额资料以及当前市场情况等有关资料,汇集整理成供专家参考的背景材料。

(二)编制主观概率调查表

编制主观概率调查表的目的是为了获得可以用来预测 2016 年 11 月份销售额的资料以及得到对未来销售额增长趋势有关看法的主观概率。在调查表中要列出不同销售额可能发生的不同概率。概率要在 0 与 1 之间分出多个层次,如 0.10,0.20,0.30, \dots ,0.99 等。一般用累积概率。由被调查者填写可能实现的销售额,如表 2—2 所示。

表 2—2 主观概率调查表					被调查人姓名_____编号_____				
累积概率	0.010 (1)	0.125 (2)	0.250 (3)	0.375 (4)	0.500 (5)	0.625 (6)	0.750 (7)	0.875 (8)	0.990 (9)
商品销售额 (万元)									

表 2—2 中第(1)栏累积概率为 0.010 的商品销售额是可能的最小数值,表示小于该数值

的可能性只有 1%；表 2-2 中第(9)栏累积概率为 0.990 的商品销售额是可能的最大数值，说明商品销售额大于该数值的可能性只有 1%；表 2-2 中第(5)栏累积概率为 0.500 的商品销售额是最大值与最小值之间的中间值，说明商品销售额大于和小于该数值的机会都是 50%。

(三) 汇总整理

它是按事先准备好的汇总表，请各个调查人填好后，加以汇总，并计算出各栏平均数。此例共调查了 10 人。主观概率汇总表如表 2-3 所示。

表 2-3 主观概率汇总表

被调查人 编 号	累积概率								
	0.010 (1)	0.125 (2)	0.250 (3)	0.375 (4)	0.500 (5)	0.625 (6)	0.750 (7)	0.875 (8)	0.990 (9)
	商品销售额(万元)								
1	190	193	194	198	200	202	204	205	208
2	178	189	192	194	198	200	204	205	225
3	184	189	192	193	202	204	206	208	220
4	194	195	196	197	198	199	200	201	202
5	198	199	200	202	205	208	210	212	216
6	168	179	180	184	190	192	194	196	198
7	194	198	200	206	208	212	216	219	224
8	180	185	186	189	192	195	198	200	205
9	188	189	190	191	192	193	194	195	196
10	200	202	202	205	207	209	212	213	220
平均数	187.4	191.8	193.2	195.9	199.2	201.4	203.8	205.4	211.4

(四) 判断预测

根据主观概率汇总表，可以做出以下判断：

第一，该集团公司 2016 年 11 月份的商品销售额最低可达 187.4 万元，小于这个数值的可能性很小，只有 1%。

第二，该集团公司 2016 年 11 月份的商品销售额最高可达 211.4 万元，超过这个数值的可能性也只有 1%。

第三，可以用 199.2 万元作为 2016 年 11 月份该集团公司商品销售额的预测值。这是最大值与最小值之间的中间值。其累积概率为 50%，是商品销售额期望值的估计数。

第四，取预测误差为 6 万元，则预测区间为：(199.2-6)~(199.2+6)，即商品销售额的预测值在 193.2 万元~205.2 万元之间。

第五，预测商品销售额在 193.2 万元~205.2 万元之间，在第(3)栏到第(8)栏的范围之内，其发生概率相当于 $0.875 - 0.250 = 0.625$ 。也就是说，商品销售额在 193.2 万元~205.2 万元之间的可能性为 62.5%。扩大预测误差的范围，可以提高实现的可能性。例如，要求误差在 ±12 万元以内，则预测区间为 187.2 万元~211.2 万元之间，在第(1)栏到第(9)栏的范围之内，其相应概率为 $0.990 - 0.010 = 0.98$ ，即商品销售额在 187.2 万元~211.2 万元之间的可能性达到 98%。

第四节 定性预测的其他方法

一、领先指标法

各种经济现象或经济指标之间存在着各种各样的联系,彼此间都有一定的制约关系。在经济周期的演变过程中,它们在时间顺序上是有先有后的。例如,原材料价格的变动先于制成品价格的变动,教育事业的发展先于科学技术的发展等。

领先指标法就是将各种经济指标分为三种类型:(1)领先指标;(2)同步指标;(3)滞后指标。

根据上述分类,可以通过领先指标以预测同步指标或滞后指标。这种方法的特点在于它不但可以预测经济的发展趋势,而且可以预测其转折点。

领先指标法的预测步骤如下:

(1)根据预测指标找出领先指标。例如,预测化工产品的价格变动问题,就可以把石油价格变动作为领先指标。

(2)画出领先指标、同步指标、滞后指标的时间序列图,如图 2—1 所示。

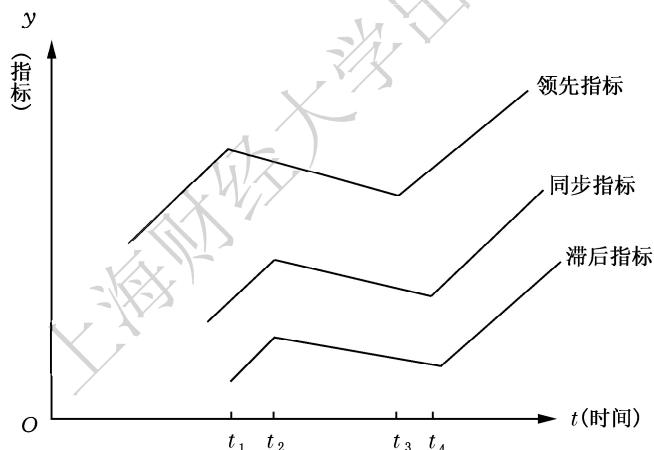


图 2—1 各类指标的时间序列图

图中: t_1 为领先指标出现最高点的时间, t_2 为同步指标出现最高点的时间, t_3 为领先指标出现最低点的时间, t_4 为同步指标出现最低点的时间, $t_2 - t_1 = \tau$ 为领先时间。

(3)进行预测。通过领先时间的估算,得到领先时间 τ ,若领先指标于 $t = t_3$ 时达到最低点,则可预测同步指标将于 $t = t_3 + \tau$ 时达到最低点。

二、厂长(经理)评判意见法

厂长(经理)评判意见法,就是由企业的总负责人把与市场有关或熟悉市场情况的各种负责人员(包括主管供销、生产、财务、产品开发与研究等领导人员)和中层管理部门的负责人(包括计划科、销售科、财务科、采购供应科等部门的管理决策人员)召集起来,让他们对未来的市场发展形势或某一重大市场问题发表意见,做出判断。然后,将各种意见汇总起来,进行分析

研究和综合处理,最后得出市场预测结果。

这种预测方法的主要优点如下:

(1)迅速、及时和经济,不需要经过复杂的计算,也不需要多少预测费用,就可以及时得到预测结果。

(2)由于这种方法集中了各个方面熟悉市场情况的有经验的高中级经营管理人员的意见,因此,可以发挥集体的智慧,使预测结果比较准确可靠。

(3)使用这种方法不需要有大量的统计资料,更适合于对那些不可控制因素较多的产品进行销售预测。

(4)如果市场情况发生了变化,可以立即进行修正。

正因为有以上一些优点,所以,这种预测方法得到了广泛的应用。据美国协商委员会对161家公司使用预测方法进行调查的结果,采用这种方法进行预测的公司占54%。其中,生产资料生产公司中有50%,消费品生产公司中有64%,服务性公司中有72%。在我国工商企业中,这种方法也得到了一定程度的应用。

这种预测方法的主要缺点如下:

(1)预测结果容易受主观因素的影响。

(2)对市场变化、顾客的愿望等问题了解不细,因此,预测结果比较一般化。

采用这种方法的一个重要前提是:厂长(经理)及参与预测的有关领导人员有较高的知识、较丰富的经验以及对市场的洞察能力和分析能力。如果不具备这样的条件,就不能做出正确的预测。

厂长(经理)评判意见法的适用性很强,可以在各种场合应用。由以下的例子便可说明。

【例2-3】某纺织厂厂长召集主管销售、财务、计划和生产等部门的负责人,对下一年度某种纺织品的销售前景做出估计。几个部门负责人的初步判断如表2-4所示。

表 2-4 各部门负责人评判表

部 门	各种销售量 估 计	销售量 (件)	概 率	期望值(件) (销售量×概率)
销售部门 负责人	最高销售量	3 000	0.2	600
	最可能销售量	1 800	0.6	1 080
	最低销售量	1 600	0.2	320
	总期望值		1.0	2 000
计划财务部门 负责人	最高销售量	2 000	0.3	600
	最可能销售量	1 800	0.5	900
	最低销售量	1 500	0.2	300
	总期望值		1.0	1 800
生产部门 负责人	最高销售量	2 000	0.2	400
	最可能销售量	1 700	0.5	850
	最低销售量	1 200	0.3	360
	总期望值		1.0	1 610

部门负责人判断意见提出之后,有以下两种综合处理的方法:

(1)绝对平均法。

下一年度某种纺织品的销售量预测值为:

$$\frac{2\,000+1\,800+1\,610}{3}=1\,803(\text{件})$$

(2)加权平均法。

根据各部门负责人对市场情况的熟悉程度以及他们在以往的预测判断中的准确程度,分别给予不同部门负责人不同的评定等级,在综合处理时,采用不同的加权系数。例如,销售部门一般比较熟悉市场情况,其预测估计的可靠性较高,因此,在综合三个部门的判断意见时,应加强销售部门负责人的判断意见在综合结果中的影响。例如,销售部门负责人的加权系数为2,其他两个部门负责人的加权系数各为1,从而下一年度某纺织品的销售预测值为:

$$\frac{2\,000 \times 2 + 1\,800 + 1\,610}{4}=1\,852.5(\text{件})$$

厂长(经理)评判意见法也适用于对商品的规格品种、性能用途、款式花色、质量与服务、新产品开发前景,对消费者和用户的消费心理、习惯、购买意向,对价格变动、竞争形势、流通渠道演变、库存控制等做出预测分析。

三、推销人员估计法

推销人员估计法就是在做统计预测时,把本企业所有的推销人员(包括代理商、经销商和分支机构的推销人员)都找回来,让他们对自己负责的销售区(或产品)下一季度或下一年度的销售额做出估计,然后把他们每个人的估计销售额汇总起来,做出本企业下一季度或下一年度销售额的预测。

推销人员估计法的主要优点如下:

(1)它与厂长(经理)评判意见法一样,不需要经过复杂的计算,因此,预测速度比较快,也比较节省费用。

(2)由于推销人员一直在市场中活动,因此,他们对市场情况特别是对他所在地区的市场情况很熟悉,而且,对原有顾客的需要情况和对潜在顾客的情况心中有数,所以,他们的预测结果比较准确可靠。正因为如此,这种方法的应用颇为流行。

当然,这种预测方法也有缺点,它同经理人员评判法一样,具有主观因素,容易受个人偏见的影响。假如有些推销人员对形势发展的认识比较乐观,他们估计的预测数字就可能偏高;反之,有些推销人员如果对形势发展抱悲观态度,他们估计的预测数字就可能偏低。特别当有些企业把完成销售任务同评定推销人员的业绩结合在一起时,就会给预测结果带来更大的影响。因为推销人员怕把数字估计高了,将来完不成销售任务而得不到奖励,因此,不愿意把那些可能争取到的销售额估计进去。这样,就会使整个预测数字不准确。

为了避免这种缺点,许多企业在采用这种方法进行预测时,除要求推销人员要有正确的态度外,还采取以下两种办法来纠正偏差:一种办法是让预测与销售人员的业绩评定分开,使预测尽量客观一些;另一种办法是根据每个推销人员历年来估计的数字同实际销售额之间的差额打一个折扣(即调整系数),使之更接近于实际。

【例 2—4】 某公司有三个推销人员,他们对自己负责的销售区下一年度的销售额估计如表 2—5 所示。

表 2-5 推销人员估计值汇总表

推销员	各种销售额估计(万元)		概 率	期望值(万元)
甲	最高销售额	2 000	0.2	400
	最可能销售额	1 600	0.5	800
	最低销售额	1 200	0.3	360
	总期望值		1.0	1 560
乙	最高销售额	1 500	0.3	450
	最可能销售额	1 400	0.6	840
	最低销售额	1 200	0.1	120
	总期望值		1.0	1 410
丙	最高销售额	1 500	0.3	450
	最可能销售额	1 300	0.6	780
	最低销售额	1 100	0.1	110
	总期望值		1.0	1 340

把这三个推销人员的估计值汇总在一起,就可以得到企业下一年度销售额预测值为 4 310 万元。

根据以往的经验,推销员甲每年估计的数字都比实际销售额高出 20%,销售员乙每年估计的数字都比实际销售额低 10%左右,推销员丙每年估计的数字都比实际销售额低 30%左右,这样就要对他们自己估计的销售额按各自偏差的比例进行调整。例如,推销员甲估计的期望值为 1 560 万元,去掉高出的 20%(312 万元),结果为 1 248 万元;推销员乙估计的期望值为 1 410 万元,他要增加 10%(141 万元),结果为 1 551 万元;推销员丙估计的期望值为 1 340 万元,他要增加 30%(402 万元),结果为 1 742 万元。经过调整以后,再把三者的结果加起来,便得到明年的预测销售额为 4 541 万元。

四、相互影响分析法

相互影响分析法就是从分析各个事件之间由于相互影响而引起的变化以及变化发生的概率,来研究各个事件在未来发生的可能性的一种预测方法。

利用相互影响分析法进行预测的步骤为:

- (1)通过预测者的主观判断,获得各种有关事件的发生概率。
- (2)用矩阵的形式描述各种事件相互之间的逻辑关系,用概率的变化表示各事物相互影响的强度,并分析各事件之间相互发生影响作用的时间。
- (3)根据各事件之间相互影响的结果,修正各事件的发生概率,做出最后预测。

【例 2-5】 在评定今后 15 年内能源政策的各种方案时,考虑了三种可能出现的相互关联的事件 A_1, A_2, A_3 及其相应发生的概率 P_1, P_2, P_3 。假定:

A_1 代表使用各种煤炭代替石油消耗,其概率 $P_1=0.8$ 。

A_2 代表降低国内石油价格,其概率 $P_2=0.4$ 。

A_3 代表控制环境(水源、空气)污染,其概率 $P_3=0.3$ 。

三个事件相互关系矩阵表如表 2—6 所示。

表 2—6 三个事件相互关系矩阵表

事 件	发生概率	事件间的相互影响		
		A_1	A_2	A_3
A_1	P_1	—	↑	↑
A_2	P_2	↓	—	—
A_3	P_3	↓	↓	—

在表 2—6 中,向上的箭头表示正的影响,说明一事件的发生增进了另一事件发生的概率。例如,在第一行中,如果事件 A_1 发生,即使用煤炭代替石油消耗,则这一事件将影响对石油的需求,石油需求量下降,事件 A_2 即降低石油价格发生的可能性增大, A_2 发生的概率相应提高了;同时,事件 A_3 即必须控制环境污染的概率也因事件 A_1 的出现而相应提高了,因为煤是一种“肮脏”的燃料,与石油相比更易于对环境造成污染。

在表 2—6 中,向下的箭头表示负的影响,说明一事件的发生将降低或者消除相关事件发生的概率。例如,若出现事件 A_2 ,即降低石油价格,则因烧油比烧煤便宜,对煤的需求便不再迫切了,那么,事件 A_1 出现的概率就会相应降低;若出现事件 A_3 ,即加强对环境污染的控制,则事件 A_1 即用煤代替石油消耗就要受到一定影响,事件 A_2 即降低石油价格也会受到一定抑制,更加干净的能源(如电能、风能、太阳能、核能等)将会更加被重视。

在表 2—6 中,符号“—”表示无关或无影响。

最后,还要根据概率计算方式,求出正、负影响的程度数值,修正各种事件的发生概率。

【例 2—6】 某齿轮厂通过市场调查,获知未来两年对本厂齿轮的需求量将比 2011 年增加 30%。如何满足市场需求,有以下三种方案:

- (1)进行扩建,关键工序采用高效率的自动化设备。
- (2)改进劳动组织,提高工人的技术熟练程度,采用新的奖励制度等,提高劳动生产率。
- (3)增加工人,普遍开设两班制,关键工序实行三班制。

要求预测这三种方案在未来两年发生的可能性如何。

设 A_1, A_2, A_3 表示三种方案; P_1, P_2, P_3 表示三种方案各自发生的概率; T_1, T_2, T_3 表示三种方案相互影响发生的时间。

可根据各方案间的相互逻辑关系和已知条件列出矩阵,如表 2—7 所示。

表 2—7 三种方案相互影响矩阵表

未来事件	发生概率	各事件间的相互影响								
		A_1			A_2			A_3		
		A_1	P_1	T_1	A_2	P_2	T_2	A_3	P_3	T_3
A_1	$P_1=0.3$	—	0	0	↑	+0.3	1	↓	-0.5	1
A_2	$P_2=0.5$	↑	+0.2	0	—	0	0	↓	-0.1	1
A_3	$P_3=0.8$	↓	-0.3	0	↓	-0.2	0	—	0	1

在分析相互影响前,原定三种方案的发生概率 $P_1=0.3, P_2=0.5$ 和 $P_3=0.8$,是根据最

近两年本厂的形势和资金情况确定的。本厂进行扩建、采用自动化设备的可能性最小,发生的概率只有 0.3;用提高劳动生产率的办法提高产量的把握程度不大,只有 50%的可能;最切实可行的方案是增加工人,其发生概率为 0.8。

在分析相互影响后发现,若采用自动化设备的方案,对提高劳动生产率有正的影响,而对增加工人则有负的影响,这种影响在时间上要滞后 1 年才能发生,并且采用自动化设备还必须筹措资金。若采用增加工人的方案,对采用自动化设备有负的影响,其影响程度将使采用自动化设备的概率降到 0;对提高劳动生产率也有负的影响,使提高劳动生产率的可能性大大减少。

根据各种方案之间的相互影响,修正后的概率为:

$$P_1 = 0.3 + 0.3 - 0.5 = 0.1$$

$$P_2 = 0.5 + 0.2 - 0.1 = 0.6$$

$$P_3 = 0.8 - 0.3 - 0.2 = 0.3$$

因此,可以预测,未来两年该厂进行扩建、采用自动化设备的可能性很小,只有 0.1;增加工人的可能性也不大,只有 0.3;最大的可能是采用提高劳动生产率的办法来提高产量,满足市场需求。为此,企业在制定未来两年的规划时,应当把立足点放在改进劳动组织、提高工人技术熟悉程度、提高劳动生产率这一点上。

第五节 情景预测法

定性预测方法有一个共同的缺点,即主观性较强,而定量分析方法容易受模型假设条件的限制,且只能对定量数据进行分析,无法考虑定性因素的影响,尽管有时可通过设虚拟变量将定性因素定量化,但效果不佳。另外,定量预测是根据适用的模型得到对将来某种状况的分析,而面对有多种情况的预测,定量预测是不够的,如在对我国经济发展的分析过程中,我国的经济状况在改革前后有很大的变化,因此,改革成为影响经济发展的重要因素。我国目前正处于政治体制与经济体制的变革时期,在这种背景下,预测问题变得更加复杂,因此,单凭定量分析是难以反映错综复杂的经济关系的。然而,只凭定性预测又没有一定的数据根据,不利于决策者进行分析,所以,客观上需要寻找一种定性和定量相结合的分析方法,情景预测法正是解决这一问题的有效方法。

一、情景预测法的概念和特点

情景预测法是 20 世纪 70 年代兴起的一种预测技术,又称剧本描述法。“情景”一词最早出现在 20 世纪 60 年代末凯恩和维拉的《2000 年》一书中。该书将情景分析定义为:用于着重研究偶发事件及决策要点的一系列假设事件。情景预测法是对将来的情景做出预测的一种方法,它把研究对象分为主题和环境,通过对环境的研究,识别影响主题发展的外部因素,模拟外部因素可能发生的多种交叉情景,以预测主题发展的各种可能前景。

情景预测法首先是构造一个“无突变”情景 A,即在假定当前的环境不发生重大变化的条件下研究对象的未来情景;然后分析情景 A 的环境因素,就各因素的不同取值从而对 A 造成不同的影响,由此产生了情景 B 和情景 C,进而可以得到 A, B, C, AB, AC, BC 六种情景;同时,还可假设有突发事件 D,它对情景 A、B、C 又有不同程度的影响,从而又产生了 AD, BD, CD, ABD, ACD, BCD 六种情景,由于环境因素的不同取值,还可得到其他多种情景,但情景的范围

是确定的,为 BUCUD,详见图 2-2。

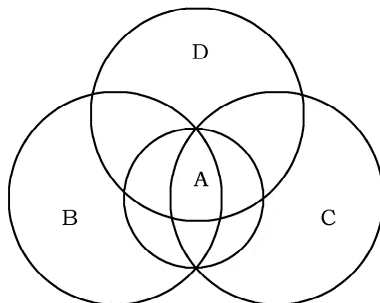


图 2-2 情景示意图

情景预测法在分析过程中根据不同情景可采用不同的预测方法,使定量、定性分析相结合,这样就弥补了定性预测和定量预测各自的缺陷。

情景预测法不同于一般方法,其特点主要表现在以下几个方面:

第一,适用范围很广,不受任何假设条件的限制,只要是对未来的分析,均可使用。

第二,考虑问题周全,又具有灵活性。它尽可能地考虑将来会出现的各种状况和各种不同的环境因素,并引入各种突发因素,将所有的可能尽可能地展示出来,有利于决策者进行分析。

第三,通过定性分析与定量分析相结合,为决策者提供主、客观相结合的未来情景。它通过定性分析寻找出各种因素和各种可能,并通过定量分析提供一种尺度,使决策者能更好地进行决策。

第四,能及时发现未来可能出现的难题,以便采取行动消除或减轻它们的影响。

二、情景预测法的一般方法

常见的情景预测法有未来分析法、目标展开法、间隙分析法三种。

(一)未来分析法

未来分析法立足于现在,着眼于未来,是最常用的一种方法。未来分析法通常视未来为三种情景:无突变情景、悲观情景、乐观情景。一般而言,未来分析法先假设目前的状况会持续发展,预测以这样的发展状况未来会出现什么样的情景,即得到无突变情景;再找出对未来情景有影响的各种环境因素,让它们进行不同程度的变化,从而得到有利的环境和不利的环境;最终分析在有利环境和不利环境下分别得到什么样的乐观情景和悲观情景。

例如,预测我国人民的生活水平。根据过去的历史和目前的状况,我们可以预测在不久之后,我国将消除贫困,人民生活水平进一步提高,这是一种情景。分析该情景的环境因素有国家政策、国际环境影响、自然条件、人的观念等。假设国内外条件均有利于我国经济的发展,而自然条件也非常有利于农业生产,则在这一有利环境下可得到乐观情景是:国民经济加速发展,人民生活水平有很大提高,提前进入“小康”,实现共同富裕。再假设国际环境对我国经济发展形成一定阻碍,又碰上自然灾害,且在扶贫工作中又出现了一些困难,从而不利于国民经济的发展,导致了悲观情景的出现:国民经济发展速度有所减慢,人民生活水平提高的速度也相应减缓。

我们还可有其他关于未来的预测,不同的环境与主题相互作用,得到不同的未来预测。

(二)目标展开法

目标展开法与未来分析法不同,它立足于未来,分析现在,即已确定好目标,去分析如何达成这一目标。在分析过程中,可根据总目标设计出各子目标,再分析实现这些目标需满足的环境、条件,并从中寻找一条最佳路径。

例如,确定在未来一年内经济增长速度为10%,现对这一目标进行分析。经济增长主要受资金、劳动力、科技进步三个因素的影响。这三者对经济增长的不同贡献率的组合,均可实现这一目标。我们可根据现实状况设置不同的组合,如三者贡献率依次为30%、30%、40%,或50%、10%、40%,又或30%、10%、60%等。根据不同的组合分析它们的可行性,从中选取最优的实现途径。当然,不同状况下的最优选择也不同,如改革前可能会选取偏重于劳动力、资金的投入,而在改革后更重视对科技的贡献率。其中,还要考虑资金、劳动力、科技进步各自的环境因素的影响,但不同的环境、不同的组合,最终的目标均为10%的增长速度。

(三)间隙分析法

间隙分析法立足于现在和未来,寻找中间途径。它主要是先根据目前状况,预测如此发展的话,将来会怎样,再根据两者的状况决定中间的路该怎样走。这与目标展开法有类似之处,但间隙分析法更强调阶段性,如分别考虑5年、7年、10年这些不同阶段下应怎样做。

三、情景预测法的一般步骤

这里结合具体案例,说明情景预测法的一般步骤。

第一步,确定预测主题。

若现在要求对我国的人力资源状况进行预测,在进行预测前,先分析预测的主题。主题应包括两个方面:人力资源的数量预测和人力资源的质量预测。

第二步,根据预测主题,寻找资料,充分考虑主题将来会出现的状况。

针对我国目前的人力资源状况,经过分析,认为未来我国人力资源可能出现的状况为:素质偏低,数量偏多;素质较高,数量充裕;素质较高,但人力资源结构不够合理;素质高低相差悬殊,人力资源缺乏等。

第三步,寻找影响主题的环境因素,要尽可能周全地分析不同因素的影响程度。

人力资源的数量主要受人口出生率、死亡率的影响,同时还受到健康状况、医疗卫生水平、饮食、营养状况、地理环境、国家计划生育政策等因素的影响。

人力资源的质量则受文化教育程度、国家对人力资源的开发、出国人员和回国人员数量、教育结构的合理性、人的自我价值的实现等多种因素的影响。

在做这些分析的同时,还要考虑这些环境因素本身还受到哪些因素的影响,如国家对人力资源的开发受国家财政、国家的师资力量、国家对人力资源不同的需求等因素的影响,这样才能使分析尽可能地周密、详细。

第四步,将上述影响因素归纳为几个影响领域,分析在不同影响领域下主题实现的可能性,同时分析是否有突发事件的影响,若有,影响如何。

对人力资源的影响因素进行归纳,可归纳为以下几个方面:人口的数量及年龄构成,受教育程度,在校学生状况,国家对教育的重视程度等。突发事件则可能为:自然灾害造成人口的减少;经济萧条引起对人力资源需求的减少,从而影响人力资源状况。

第五步,对各种可能出现的主题状态进行预测。

在这一步骤中可采用不同的方法,对有数据的主题可采用定量分析方法,否则进行定性分

析,也可定量与定性相结合进行预测。

对人力资源可采用定量分析方法预测未来的人口数和各种学历的人数,还可采用定性分析方法分析国家对教育的重视程度,并采用主观打分法对各影响因素进行打分,最终对我国未来的不同的人力资源状况进行综合评分,并分析其实现的可能性,这可通过专家调查法得到。

四、情景预测法的实证分析

改革开放 30 几年来,我国经济、政治、社会、科技等各个领域都发生了翻天覆地的变化,综合国力大大增强,人民生活水平显著提高,能反映这些变化的最重要标志是我国的国内生产总值。国内生产总值(GDP)反映了全社会最终产品和劳务的价值总量,它具有最完整的物质内容,是整个国民经济核算的基本总量和中心环节。从 1978 年到 2014 年,我国的 GDP 年均增长率达到 9%以上,经济实力明显增强。对我国 GDP 的增长情况进行较好的预测和分析,对评价我国未来发展趋势具有十分重要的作用。

(一)确定预测主题

国内生产总值常被公认为衡量国家经济状况的最佳指标。国内生产总值是指一个国家领土范围内的所有常住单位,在一定时期内生产最终产品和提供劳务价值的总和。此外,国内生产总值增长率也经常被用来反映一国的经济增长程度,在此仅选取国内生产总值作为我们的变量。

(二)分析未来情景

未来的情景随具体情况的不同而不同,我国经济在改革开放前后呈现出巨大的差异,主要分为计划经济时期和市场经济时期两个阶段,而在市场经济时期又可分为由计划经济向市场经济转轨时期和市场经济发展时期。国内生产总值在这几个阶段的不同情况将在后面的具体分析中做详细的说明。

(三)寻找影响因素

一国的国内生产总值受到各方面因素的影响,主要有经济体制、资本、劳动力、技术进步、消费和进出口等方面的因素。

1. 经济体制

经济体制是经济增长的重要保障。如果一个国家采取的经济体制符合该国的生产力发展水平,就能促进其国内生产总值的增长。我国从新中国成立到改革开放前实行的是高度集中的计划经济体制,使国家有限的财力、物力迅速地集中起来,为稳固社会主义制度奠定了物质基础。但是,完全由政府计划为主导的经济体制在推动经济增长方面具有很大的局限性。改革开放后,市场的逐渐开放为经济发展提供了有利的环境和基础,经济效率提高,国内生产总值以较快的速度保持增长。

2. 资本

资本对经济发展的作用毋庸置疑。不论是发达国家还是发展中国家,必须有资本投入才能使经济发展和增长,充足的资本投入是国内生产总值不断提高的必要条件。资本一般可分为固定资产投资与存货增加,其中固定资产投资占主要部分。

3. 劳动力

我国是人口大国,每年劳动力人口数量的不断增加对国内生产总值的持续增长有很大的贡献。而随着世界各国经济增长方式由粗放型转向资本、知识密集型,单纯的劳动力数量上的增加对 GDP 增长的贡献正在降低,同时还造成劳动力结构性过剩等一系列问题。另一方面,

人力资本(即劳动力的教育程度、知识积累能力等)则越来越受到人们的重视,对国内生产总值的影响在不断提高。

4. 技术进步

现代科学技术已经成为影响经济增长的决定性因素,技术进步对 GDP 增长的贡献已渐渐超过资本和劳动力的作用。高科技的发展水平已经成为一个国家综合国力的主要因素,是衡量一个国家发达与否的重要标志。

5. 消费

消费需求对一国的经济发展有非常重要的作用。消费需求增加会扩大内需,促进生产,吸引资本投入,最终导致国内生产总值上升。相对于市场经济时期,由于在计划经济时期我国实行计划消费,消费对经济的作用没能很好地体现出来。

6. 进出口

进出口与国内消费一样,对一国国内生产总值具有很大的影响。净出口值的增加会通过促进产业结构调整、吸引外资流入、拉动需求和增加就业等方面促进产值增加。中国加入 WTO 后,进出口因素对 GDP 的影响作用越来越明显。

7. 其他

除了上述主要因素外,经济周期波动、国际经济形势、自然灾害等各种因素均会对国内生产总值产生影响。

(四)具体分析

由于我国的国内生产总值变化有较明显的阶段性,故在分析时也分阶段进行,其中 1952~1977 年为计划经济阶段,1978~1990 年为经济体制转轨阶段,1991~2014 年为市场经济发展阶段。另外,1949~1952 年为新中国成立后的恢复阶段,故在分析时从 1952 年的数据开始。

1. 计划经济阶段:1952~1977 年

新中国成立初期,计划经济体制为国民经济的恢复和发展做出了巨大的贡献。在 1953~1957 年第一个五年计划期间,我国国内生产总值不断提高,人均 GDP 从 119.4 元增长到了 167.8 元。而在之后的“大跃进”期间,国民经济发展极度不均衡,造成了从 1961 年开始 GDP 的下降。之后长达十年的“文化大革命”,更是搅乱了我国国民经济的正常发展。

1956 年,全国范围内对于农业、手工业和资本主义工商业进行的社会主义改造顺利完成,中国进入社会主义建设时期。以这年为基础进行比较,由表 2-8 可得到,1956 年,中国 GDP 为 1 029 亿元,人均 GDP 为 165.6 元;到“文化大革命”结束的 1976 年,我国 GDP 为 2 961.5 亿元,人均 GDP 为 318.2 元。20 年的时间,中国的经济规模才增加不到 2 倍,人均 GDP 增加不到 1 倍。在这一阶段,国民经济波动较大,发展缓慢。

因此,在这一阶段可能出现三种未来情景:

(1)无突变情景,即由于之前国民经济发展遭到阻碍,各经济部门结构还需要进行协调,近期内经济不会复苏,国内生产总值仍然徘徊在低水平。

(2)悲观情景,即由于在“文化大革命”结束时国民经济已濒临崩溃边缘,经济形势将继续下滑。

(3)乐观情景。即由于“文化大革命”时期的结束,国家的发展重点回到经济上,经济开始复苏,国内生产总值的增长率开始提高。

分析这三种情景,在这一阶段,占主体的应为无突变情景和悲观情景。

表 2—8 1952~1977 年中国 GDP、人均 GDP 数据表

年 份	国内生产总值 (亿元)	人均国内 生产总值 (元)	年 份	国内生产总值 (亿元)	人均国内 生产总值 (元)
1952	679	119.4	1965	1 717.2	240.1
1953	824.2	141.8	1966	1 873.1	254.7
1954	859.4	144.4	1967	1 780.3	235.9
1955	910.8	149.6	1968	1 730.2	223.4
1956	1 029	165.6	1969	1 945.8	244.4
1957	1 069.3	167.8	1970	2 261.3	276.3
1958	1 308.2	200.3	1971	2 435.3	289.5
1959	1 440.4	216.3	1972	2 530.2	293.5
1960	1 457.5	218.5	1973	2 733.4	309.9
1961	1 220.9	184.9	1974	2 803.7	311.4
1962	1 151.2	172.9	1975	3 013.1	328.8
1963	1 236.4	181.2	1976	2 961.5	318.2
1964	1 455.5	208.4	1977	3 221.1	341.4

注：以上数据来自中国经济统计数据库。

2. 经济体制转轨阶段：1978~1991 年

党的十一届三中全会后，经济体制的全面改革和现代化建设的全面开展，使国民经济有了较快的发展。由表 2—9 可以看到，1978 年，我国国内生产总值为 3 645.2 亿元，人均 GDP 为 381 元；到 1991 年，我国 GDP 已达 21 781.5 亿元，人均 GDP 为 1 893 元。国内生产总值增长了约 5 倍，人均国内生产总值约翻了 4 倍。1978~1991 年，我国 GDP 平均每年增长率都超过 9%，高于同一时期其他国家的增长速度。

从 1978 年实行改革开放以来的这 14 年间，我国共经历了 1978~1984 年、1985~1987 年和 1988~1992 年三个经济波动周期。由于我国的特殊国情，我国 GDP 增长的每次波动都伴随着国民经济“过热—过冷—过热”的频繁调整，在一定程度上影响了经济运行的质量。

在对这一阶段的情景预测中，同样可分为三种未来情景：

(1) 无突变情景，即经济继续保持稳步增长的趋势。

(2) 乐观情景，即随着市场的全面开放、国家更多的有利政策出台，国民经济将以更快的速度发展，国内生产总值的增长率将进一步提高。

(3) 悲观情景，即由于经济过热，国民经济再次陷入“过热—过冷”的经济周期中，国内生产总值增长率有所下降。

表 2-9 1978~1991 年我国 GDP、人均 GDP 及固定资产投资额数据表

年 份	国内生产总值 (亿元)	人均国内生产总值 (元)	全社会固定资产投资总额 (亿元)
1978	3 650.2	382	—
1979	4 067.7	420	—
1980	4 551.6	464	910.9
1981	4 898.1	493	961
1982	5 333	529	1 230.4
1983	5 975.6	584	1 430.1
1984	7 226.3	697	1 832.9
1985	9 039.9	860	2 543.2
1986	10 308.8	966	3 120.6
1987	12 102.2	1 116	3 791.7
1988	15 101.1	1 371	4 753.8
1989	17 090.3	1 528	4 410.4
1990	18 774.3	1 654	4 517
1991	21 895.5	1 903	5 594.5

注：以上数据来自中国经济统计数据库。

3. 市场经济发展阶段：1992~2014 年

这一阶段，我国的经济体制改革目的进一步明确，国民经济稳定、高速地增长，特别是 1992 年邓小平南方讲话后，国民经济的发展更是上了一个新台阶。从 1992 年到 1996 年，国内生产总值增加了 164%，人均国内生产总值增加了 153%，固定资产投资额增加了 183%。为了防止经济过热，党中央提出要适当减缓经济增长速度，故在后面的几年中，国民经济增长速度略为减缓。进入 21 世纪后，伴随着加入 WTO，我国经济建设日趋完善，市场日益规范，经济稳步上升的态势愈加明显。到了 2006 年，我国 GDP 已经是 1992 年时的 8 倍左右（见表 2-10）。2008 年的北京奥运会、2010 年的上海世博会都极大地刺激了投资和经济的发展。2013 年，中国（上海）自由贸易试验区经国务院正式批准设立。2014 年中国证监会正式批复开展互联互通机制试点，建立上海与香港市场交易互联互通机制。因此，在可预见的未来，国内生产总值仍然将快速、稳定增长。

在对这一阶段的情景预测中，同样可分为无突变情景、乐观情景和悲观情景。

(1) 无突变情景，即经济将继续按照目前的发展速度增长，实现国内生产总值的稳步提高。

(2) 乐观情景，即经济在奥运会或世博会的刺激下，有进一步的提高，国内生产总值将以更快的速度保持增长。

(3) 悲观情景，即由于目前经济有过热趋势，国民经济会在经济周期的作用下放慢增长速度。

分析这三种情景，在这一阶段，占主体的应为乐观情景和无突变情景。

表 2-10 1992~2014 年我国 GDP、人均 GDP 及固定资产投资额数据表

年 份	国内生产总值 (亿元)	人均国内生产总值 (元)	全社会固定资产投资总额 (亿元)
1992	27 068.3	2 324	8 080.1
1993	35 524.3	3 015	13 072.3
1994	48 459.6	4 066	17 042.1
1995	61 129.8	5 074	20 019.3
1996	71 572.3	5 878	22 913.5
1997	79 429.5	6 457	24 941.1
1998	84 883.7	6 835	28 406.2
1999	90 187.7	7 199	29 854.7
2000	99 776.3	7 902	32 917.73
2001	110 270.4	8 670	37 213.49
2002	121 002	9 450	43 499.91
2003	136 564.6	10 600	55 566.61
2004	160 714.4	12 400	70 477.4
2005	185 895.8	14 259	88 773.62
2006	217 656.6	16 602	109 998.2
2007	268 019.4	20 337	137 323.94
2008	316 751.7	23 912	172 828.4
2009	345 629.2	25 963	224 598.77
2010	408 903	30 567	251 683.77
2011	484 123.5	36 018	311 485.13
2012	534 123	39 544	374 694.74
2013	588 018.8	43 320	446 294.09
2014	636 462.7	46 652	512 760.7

注:以上数据来自中国经济统计数据库。

(五)预测

在对 2015 年国内生产总值进行预测时,我们选取了全社会固定资产投资总额作为解释变量对 GDP 做回归。因为固定资产投资额作为资本的主要组成部分,对国内生产总值具有非常重要的影响,同时考虑到多变量因素模型存在的多重共线性等一系列问题,在此只简单地选取全社会固定资产投资总额一个解释变量。

1. 模型的建立

由于市场经济发展阶段的数据较少,故在建模时将转轨阶段的数据一起考虑。为了区分两个不同的阶段,这里增设一个虚拟变量,经济体制转轨阶段时虚拟变量取 1,市场经济发展阶段时虚拟变量取 2。另外,由于 1978 年、1979 年全社会固定资产投资总额的数据缺省,故为统一起见,所用数据为 1980~2014 年的数据(所用数据见表 2-9、表 2-10)。

我国国内生产总值的预测模型为:

Z ——虚拟变量。

这里将未来的情景分为两种：

与 2013 年相比,2014 年的全社会固定资产投资额增长了 14.89%,则 2015 年全社会固定资产投资总额为:

$$512\,760.7 \times (1 + 14.89\%) = 589\,110.77 (\text{亿元})$$

将 2015 年全社会固定资产投资总额代入我国国内生产总值预测模型,可得到 2015 年国内生产总值预测值:

$$Y = -42\,982.26 + 1.247 \times 589\,110.77 + 50\,359.62 \times 2 = 792\,358.11 (\text{亿元})$$

(2) 固定资产投资总额的增长有其自身的规律性,这里将全社会固定资产投资总额作为一时间序列,由于其具有较稳定的增长趋势,故用指数平滑法对全社会固定资产投资总额的增长速度进行预测,得到 2015 年全社会固定资产投资总额的增长速度为 16.46%,则 2015 年全社会固定资产投资总额为:

$$512\,760.7 \times (1 + 16.46\%) = 597\,161.11 (\text{亿元})$$

同样代入我国国内生产总值预测模型中,可得情景 2 的预测值:

$$Y = -42\,982.26 + 1.247 \times 597\,161.11 + 50\,359.62 \times 2 = 802\,396.88 (\text{亿元})$$

根据以上分析,可得到如表 2-11 所示的计算结果。

表 2-11 2015 年我国国内生产总值情景预测 单位:亿元

对象 \ 情景	情景 1 $d=14.89\%$	情景 2 $d=16.46\%$
国内生产总值 GDP	792 358.11	802 396.88

注:其中 d 为全社会固定资产投资总额年增长速度。



本章小结

定性预测是指预测者依靠熟悉业务知识、具有丰富经验和综合分析能力的人员与专家,根据已掌握的历史资料和直观材料,运用个人的经验和分析判断能力,对事物的未来发展做出性质和程度上的判断,然后再通过一定的形式综合各方面的意见,作为预测未来的主要依据。

1. 德尔菲法是根据有专门知识的人的直接经验,对研究的问题进行判断、预测的一种方法,也称专家调查法。它是美国兰德公司于1964年首先用于预测领域的。

德尔菲法具有反馈性、匿名性和统计性的特点,选择合适的专家是做好德尔菲预测的关键

环节。

2. 主观概率是人们凭经验或预感而估算出来的概率。它与客观概率不同,客观概率是根据事件发展的客观性统计出来的一种概率。在很多情况下,人们无法计算事件发生的客观概率,因而只能用主观概率来描述事件发生的概率。

主观概率法是一种适用性很强的统计预测方法,可以用于人类活动的各个领域。采用主观概率法有以下步骤:(1)准备相关资料;(2)编制主观概率调查表;(3)汇总整理;(4)判断预测。

3. 领先指标法就是通过将经济指标分为领先指标、同步指标和滞后指标,并根据这三类指标之间的关系进行分析预测。领先指标法不仅可以预测经济的发展趋势,而且可以预测其转折点。

4. 厂长(经理)评判意见法就是由企业的总负责人把与市场有关或熟悉市场情况的各种负责人员和中层管理部门的负责人召集起来,让他们对未来的市场发展形势或某一重大市场问题发表意见,做出判断;然后,将各种意见汇总起来,进行分析和综合处理;最后,得出市场预测结果。

5. 推销人员估计法就是将不同销售人员的估计值综合汇总起来,作为预测结果值。由于销售人员一般都很熟悉市场情况,因此,这一方法具有一些显著的优势。

6. 相互影响法就是从分析各个事件之间由于相互影响而引起的变化以及变化发生的概率,来研究各个事件在未来发生的可能性的一种预测方法。

7. 情景预测法是一种新兴的预测法,由于它不受任何条件限制,应用起来灵活,能充分调动预测人员的想象力,考虑较全面,有利于决策者更客观地进行决策,在制定经济政策、公司战略等方面有很好的应用。但在应用过程中一定要注意具体问题具体分析,同一个预测主题,由于其所处环境不同,最终的情景可能会有很大差异。



思考与练习

1. 什么是定性预测法? 定性预测有哪些特点?
2. 简述定性预测与定量预测的关系。
3. 德尔菲法有哪些特点? 又有哪些优点和缺点?
4. 若用德尔菲法预测 2020 年家用电脑的普及率,你准备:

(1)如何挑选专家? 挑选多少专家?

(2)设计咨询表应包含哪些内容?

(3)怎样处理专家意见?

(4)为了提高专家意见的回收率,你准备采用什么办法?

5. 什么叫主观概率? 什么问题适合采用主观概率法进行预测?

6. 情景预测法有何特点? 进行情景预测一般有哪些步骤?

7. 什么叫相互影响分析法? 试举一实例说明之。

8. 某时装公司设计了一种新式女时装,聘请了三位最有经验的时装销售人员来参加试销和时装表演活动,预测结果如下:

甲:最高销售量是 80 万件,概率 0.3

最可能销售量是 70 万件,概率 0.5

最低销售量是 60 万件,概率 0.2

乙:最高销售量是 75 万件,概率 0.2

最可能销售量是 64 万件, 概率 0.6

最低销售量是 55 万件, 概率 0.2

丙: 最高销售量是 85 万件, 概率 0.1

最可能销售量是 70 万件, 概率 0.7

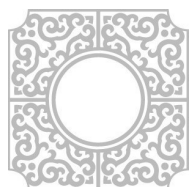
最低销售量是 60 万件, 概率 0.2

请运用推销人员估计法预测销量。

9. 上海市国内生产总值 GDP 与固定资产投资历史数据如下, 请根据可能情形对 2010 年上海国内生产总值进行预测。

年份	国内生产总值(亿元)	固定资产投资(亿元)	年份	国内生产总值(亿元)	固定资产投资(亿元)
1952	36.66	1.98	1978	272.81	27.91
1953	51.71	3.65	1979	286.43	35.58
1954	54.70	3.25	1980	311.89	45.43
1955	53.64	3.43	1981	324.76	54.60
1956	63.61	3.76	1982	337.07	71.34
1957	69.60	5.20	1983	351.81	75.94
1958	95.61	11.32	1984	390.85	92.30
1959	128.49	15.61	1985	466.75	118.56
1960	158.39	17.64	1986	490.83	146.83
1961	101.78	7.21	1987	545.46	186.30
1962	84.72	3.83	1988	648.30	245.27
1963	90.69	5.32	1989	695.54	214.76
1964	100.70	7.22	1990	756.45	227.08
1965	113.55	7.75	1991	893.77	258.30
1966	124.81	7.23	1992	1 114.32	357.38
1967	110.04	4.61	1993	1 511.61	653.91
1968	123.24	4.58	1994	1 971.92	1 123.29
1969	142.30	7.45	1995	2 462.57	1 601.79
1970	156.67	10.90	1996	2 902.20	1 952.05
1971	164.86	11.36	1997	3 360.21	1 977.59
1972	170.98	13.22	1998	3 688.20	1 964.83
1973	185.35	16.24	1999	4 034.96	1 856.72
1974	193.45	22.43	2000	4 551.15	1 869.67
1975	204.12	32.54	2001	4 950.84	1 994.73
1976	208.12	24.52	2002	5 408.76	2 187.06
1977	230.36	18.00	2003	6 250.81	2 452.11

第三章



回归预测法

第一节 一元线性回归预测法

一元线性回归预测法是指成对的两个变量数据分布大体上呈直线趋势时,采用适当的计算方法,找到两者之间特定的经验公式,即一元线性回归模型,然后根据自变量的变化,来预测因变量发展变化的方法。

一、建立模型

一元线性回归模型可表述为:

$$y_i = b_0 + b_1 x_i + u_i \quad (3-1)$$

(3-1)式中, b_0 、 b_1 是未知参数; u_i 为剩余残差项或称随机扰动项,引进随机扰动项 u_i 是为了包括对因变量 y_i 的变化有影响的所有其他因素。

在运用回归预测法时,要求满足一定的假定条件,其中最重要的是关于 u_i 须具有的 5 个特性:(1) u_i 是一个随机变量;(2) u_i 的平均值为零,即 $E(u_i) = 0$;(3)在每一个时期中, u_i 的方差为一常量,即 $D(u_i) = \sigma_u^2$;(4)各个 u_i 间相互独立;(5) u_i 与自变量无关。

二、估计参数

要将一元线性回归模型用于预测,就需要估计出 b_0 、 b_1 这两个未知参数,建立以下一元线性回归预测式:

$$\hat{y}_i = b_0 + b_1 x_i \quad (3-2)$$

一个好的估计量应满足一致性、无偏性和有效性的要求。

线性回归模型参数的估计方法通常有两种,即普通最小二乘法 and 最大似然估计法。最常用的是普通最小二乘法。

最小二乘法的意义在于使:

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \quad (3-3)$$

达到最小。

(3-3) 式中, y_i 是实际值,而 \hat{y}_i 是理论值或称估计值。

根据数学分析中的求极值原理,要使 $\sum_{i=1}^n u_i^2$ 为最小,只需在(3-3)式中分别对 b_0 、 b_1 求偏

导数,并令其等于零。

求得的 b_1 和 b_0 的两个公式为:

$$b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} \quad (3-4)$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (3-5)$$

【例 3-1】 某饮料公司发现,饮料的销售量与气温之间存在着相关关系,即气温越高,人们对饮料的需求量就越大。表 3-1 的第(2)栏和第(3)栏列出了饮料销售量和气温的观察值。该数据是某饮料公司通过实际记录所得到的。

表 3-1 饮料销售量的回归分析表

时期	销售量 (箱)	气温 (度)	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	\hat{y}	$(y - \hat{y})^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	430	30	150	9	2 500	409	441
2	335	21	270	36	2 025	322	169
3	520	35	1 120	64	19 600	458	3 844
4	490	42	1 650	225	12 100	526	1 296
5	470	37	900	100	8 100	477	49
6	210	20	1 190	49	28 900	312	10 404
7	195	8	3 515	361	34 225	195	0
8	270	17	1 100	100	12 100	283	169
9	400	35	160	64	400	458	3 364
10	480	25	-200	4	10 000	361	14 161
总和	3 800	270	9 855	1 012	129 950		33 897
总和/ n	380	27	985.5	101.2	12 995		
符号	\bar{y}	\bar{x}	σ_{xy}	σ_x^2	σ_y^2		

根据表 3-1 的数据,可求得:

$$b_1 = \frac{9\,855}{1\,012} = 9.74$$

$$b_0 = \bar{y} - b_1 \bar{x} = 380 - 9.74 \times 27 = 117$$

现在,我们可以用回归模型来预测。如果预计明天的气温是 35°C ,则饮料销售量的期望值为:

$$\hat{y} = b_0 + b_1 x = 117 + 9.74x = 117 + 9.74 \times 35 = 458(\text{箱})$$

三、进行检验

(一)标准误差

标准误差是回归直线即估计值与因变量值间的平均平方误差。其计算公式为:

$$SE = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} \quad (3-6)$$

根据表 3-1 的数据,求得标准误差为:

$$SE = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} = \sqrt{\frac{33\,897}{10-2}} = 65 (\text{箱})$$

(二) 可决系数

可决系数是衡量因变量与自变量关系密切程度的指标,表示自变量解释因变量变动的百分比。它取值于 0 与 1 之间,并取决于回归模型所解释的 y 方差的百分比。可决系数(R^2)的公式为:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (3-7)$$

(3-7)式的可决系数是用 1 减去 y 对回归直线的方差(未解释离差)与 y 的总方差的比值。

上例饮料问题的可决系数为:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} = 1 - \frac{33\,897}{129\,950} = 0.74$$

计算结果表明,气温因素变化占饮料销售量变动的 74%。在预测实践中, R^2 常用于模型的比较,人们往往采纳 R^2 最高的模型,这是因为 R^2 高,就意味着该模型把 y 的变动解释得好。

实际上,有一个更为简捷的可决系数的计算公式,计算结果同上式相同。其计算公式为:

$$R^2 = \left[\frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}} \right]^2 \quad (3-8)$$

(三) 相关系数

相关系数是另一个被广泛用来测定拟合优度的指标。它的计算公式为:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}} \quad (3-9)$$

饮料销售量与气温的相关系数为:

$$r = \frac{9\,855}{\sqrt{1\,012} \times \sqrt{129\,950}} = 0.86$$

由公式可知,可决系数只是相关系数的平方,但这两种度量方法提供了相互补充的信息。相关系数与可决系数的主要区别在于:相关系数有正负。正相关系数意味着因变量与自变量以相同的方向增减。如果直线从左至右上升,则相关系数为正;如果直线从左至右下降,则相关系数为负。尽管相关系数的意义不如可决系数那样明显,但也有类似的意义。相关系数越接近+1 或-1,因变量与自变量的拟合程度就越好。

(四) 回归系数显著性检验

在求出回归系数后,需要进行回归系数的显著性检验。回归系数的显著性检验是用 t 参数检验的:

$$t_b = \frac{b}{S_b} \quad (3-10)$$

其中: $S_b = SE / \sqrt{\sum (x - \bar{x})^2}$, t 服从自由度为 $n-2$ 的 t 分布, 取显著性水平 α , 如果 $|t_b| > t_\alpha$, 则回归系数 b 显著。

利用表 3-1 的数据, 可以计算出:

$$S_b = \frac{65}{31.81} = 2.05$$

$$t_b = \frac{9.74}{2.05} = 4.75$$

取 $\alpha = 0.05$, 查表(附表二)得: $t_{0.05}(8) = 2.306$, 显然 $t_b > t_{0.05}(8) = 2.306$, 因此, 回归系数 b 显著。

(五) F 检验

将总离差 $\sum (y - \bar{y})^2$ 分解, 可以分解为回归偏差和剩余残差两部分, 即:

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2 \quad (3-11)$$

自由度 $n-1$ 也可以分解为两部分, 即回归自由度 1 和残差自由度 $n-2$ 。将回归偏差和剩余残差各自除以它们的自由度后加以比较, 便可得到检验统计量 F , 即:

$$F = \frac{\sum (\hat{y} - \bar{y})^2 / 1}{\sum (y - \hat{y})^2 / (n-2)} \quad (3-12)$$

F 服从 $F(1, n-2)$ 分布, 取显著性水平 α , 如果 $F > F_\alpha(1, n-2)$, 则表明回归模型显著; 如果 $F < F_\alpha(1, n-2)$, 则表明回归模型不显著, 回归模型不能用于预测。利用表 3-1 的数据, 可以计算出:

$$\begin{aligned} \sum (\hat{y} - \bar{y})^2 &= \sum (y - \bar{y})^2 - \sum (y - \hat{y})^2 \\ &= 129\,950 - 33\,897 \\ &= 96\,053 \end{aligned}$$

$$F = \frac{96\,053/1}{33\,897/8} = 22.67$$

取显著性水平 $\alpha = 0.05$, 查表(附表三)得: $F_{0.05}(1, 8) = 5.32$, 因 $F > F_{0.05}(1, 8)$, 方程通过 F 检验。

(六) 德宾—沃森统计量

如前所述, 回归模型的剩余项 u_i 之间应该是相互独立的。也就是说, 各个 u_i 之间不存在自相关问题。如果存在自相关问题, 那么, 用回归模型进行预测就要失真。德宾—沃森统计量 ($D-W$) 是检验模型是否存在自相关的一种有效方法, 其公式为:

$$D-W = \frac{\sum_{i=2}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2} \quad (3-13)$$

式中:

$$u_i = y_i - \hat{y}_i$$

把(3-13)式计算的 $D-W$ 值, 与德宾—沃森给出的不同显著性水平 α 的 $D-W$ 值之上限 d_U 和下限 d_L (它们与样本容量 n 和自变量个数 m 有关) 进行比较, $D-W$ 的取值域在 $0 \sim 4$ 之间。

$D-W$ 小于等于 2 时, $D-W$ 检验法则规定:

如果 $D-W < d_L$, 认为 u_i 存在正自相关;

如果 $D-W > d_U$, 认为 u_i 无自相关;

如果 $d_L < D-W < d_U$, 不能确定 u_i 是否有自相关。

$D-W$ 值大于 2 时, $D-W$ 检验法则规定:

如果 $4 - d_L < D-W$, 认为 u_i 存在负自相关;

如果 $4 - d_U > D-W$, 认为 u_i 无自相关;

如果 $4 - d_L < D-W < 4 - d_U$, 不能确定 u_i 是否有自相关。

由图 3-1 可看出, $D-W$ 值等于 2 时为最好。根据经验, $D-W$ 统计量在 1.5~2.5 之间时, 表示没有显著自相关问题。

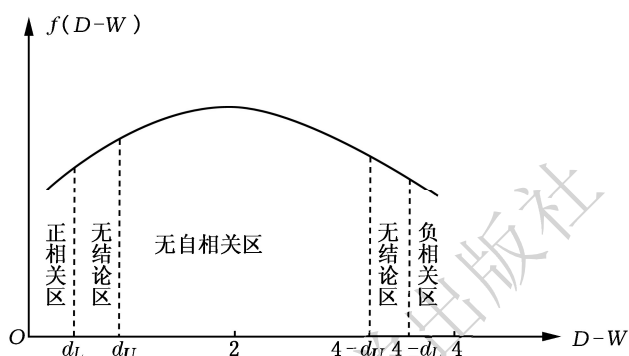


图 3-1 $D-W$ 统计量的范围与有有序列相关的范围关系图

我们利用表 3-1 中的 y 和 \hat{y} 数据, 计算了表 3-2 中饮料销售量的 $D-W$ 统计量。

表 3-2 $D-W$ 统计量计算表

i	y_i	\hat{y}_i	$y_i - \hat{y}_i$	$[(y_i - \hat{y}_i) - (y_{i-1} - \hat{y}_{i-1})]^2$	$(y_i - \hat{y}_i)^2$
1	430	409	-21	—	441
2	335	322	13	64	169
3	520	458	62	2 401	3 844
4	490	526	-36	9 604	1 296
5	470	477	-7	841	49
6	210	312	-102	9 025	10 404
7	195	195	0	10 404	0
8	270	283	-13	169	169
9	400	458	-58	2 025	3 364
10	480	361	119	31 329	14 161
总计	—	—	—	65 862	33 897

$$D-W = \sum_{i=2}^n (u_i - u_{i-1})^2 / \sum_{i=1}^n u_i^2 = \frac{65\,862}{33\,897} = 1.94$$

计算结果表明, 本例不存在自相关问题。如果检验结果发现有自相关问题, 就必须对数据进行调整。例如, 运用以下变量能对饮料销售量进行新的回归分析:

y : 本期饮料销售量相对上一期的增量;

x : 本期气温相对上一期的增量。

如果调整之后仍存在自相关问题,就有必要用本章第三节探讨的方法进行非线性回归或用时间序列法来进行模拟分析。

四、进行预测

建立回归模型的主要目的之一是根据自变量的变化来预测或估计因变量的变动情况。具体的估计的形式可分为点估计和区间估计两种。

(一)点估计

这种方法比较简单,只要将给定的自变量值代入所建立的一元线性回归模型,便可得到因变量的一个对应的估计值。本例中,当气温为 35°C 时,饮料的销售量计算如下:

$$\begin{aligned}\hat{y} &= 117 + 9.74x \\ &= 117 + 9.74 \times 35 = 458 (\text{箱})\end{aligned}$$

(二)区间估计

在估计中,有时需要给出精确度,这就要用到在一定概率保证程度下的区间估计方法。根据一元线性回归模型的性质可知, \hat{y} 的抽样分布服从正态分布,但 σ_y^2 未知,需要用它的估计值 S_y^2 来代替,这样对 \hat{y} 进行区间估计时就要区分大样本与小样本。

(1)如果要估计的是因变量的平均水平 μ_y ,则所估计的区间称为置信区间。计算公式为:

$$\text{大样本: 置信区间} = \hat{y} \pm z_{\frac{\alpha}{2}} SE_{\hat{\mu}} \quad (3-14)$$

$$\text{小样本: 置信区间} = \hat{y} \pm t_{\frac{\alpha}{2}} SE_{\hat{\mu}} \quad (3-15)$$

(2)如果要估计的是某个特定的因变量 y 值,则所估计的区间称为预测区间。其计算公式为:

$$\text{大样本: 预测区间} = \hat{y} \pm z_{\frac{\alpha}{2}} SE_{\hat{y}} \quad (3-16)$$

$$\text{小样本: 预测区间} = \hat{y} \pm t_{\frac{\alpha}{2}} SE_{\hat{y}} \quad (3-17)$$

其中, $SE_{\hat{\mu}} = SE \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$, $SE_{\hat{y}} = SE \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ (证明略), x_0 为用于预测 y 的 x 值, $t_{\frac{\alpha}{2}}$ 是置信度为 $(1-\alpha)$ 、自由度为 $(n-2)$ 的 t 分布的临界值。

本例中,在小样本条件下,如果取显著性水平 $\alpha = 0.1$,在气温 35°C 时,饮料销售量的置信度为 90% 的置信区间为:

$$\hat{y} \pm 1.860 \times 65 \times \sqrt{\frac{1}{10} + \frac{(35-27)^2}{1012}} = 458 \pm 48$$

即当气温为 35°C 时,有 90% 的把握可以估计饮料销售量的平均变动范围在 410~506 箱之间。

同样的条件下,饮料销售量的预测区间为:

$$\hat{y} \pm 1.860 \times 65 \times \sqrt{1 + \frac{1}{10} + \frac{(35-27)^2}{1012}} = 458 \pm 130$$

即当气温为 35°C 时,有 90% 的把握认为饮料销售量的变动范围在 328~588 箱之间。

第二节 多元线性回归预测法

以上讨论了两个变量因素之间的回归预测问题。然而,客观事物的变化往往受多种因素

的影响,即使其中一个因素起着主导作用,但有时其他因素的作用也是不可忽视的。在实际问题中,大多数影响因变量的因素不是一个,而是多个。我们把包括两个或两个以上自变量的回归称为多元回归。这一节中,我们首先讨论两个自变量的模型,借以说明多元回归的使用,然后推广到三个或三个以上自变量的回归模型。

一、估计参数

【例 3—2】 承例 3—1,该饮料公司的许多零售点设在体育比赛场地,该公司明白,当比赛一边倒时,观众就会比往常喝得多一些,因为这时观众就有时间注意到口渴,而不是把注意力完全集中在比赛上。因此,可以利用比赛结束时的比分差作为第二个自变量,其预测模型为:

$$\text{饮料销售量} = b_0 + b_1 \text{气温} + b_2 \text{比分差}$$

模型中可以包括进去的自变量个数是没有限制的。通常,使用一个以上的自变量可以使预测精度大大提高。

上述预测模型用符号可表达为:

$$\hat{y} = b_0 + b_1 x + b_2 z \quad (3-18)$$

式中, y 是因变量, x 和 z 是自变量, b_0 、 b_1 和 b_2 是回归系数。回归系数的计算可以通过解以下联立方程求得:

$$\sum y = nb_0 + b_1 \sum x + b_2 \sum z \quad (3-19)$$

$$\sum xy = b_0 \sum x + b_1 \sum x^2 + b_2 \sum xz \quad (3-20)$$

$$\sum zy = b_0 \sum z + b_1 \sum xz + b_2 \sum z^2 \quad (3-21)$$

表 3—3 列出饮料销售量、气温和 10 场比赛结束时的比分差。

利用表 3—3 的数据,可得:

$$3\ 800 = 10b_0 + 270b_1 + 100b_2$$

$$112\ 455 = 270b_0 + 8\ 302b_1 + 2\ 771b_2$$

$$41\ 545 = 100b_0 + 2\ 771b_1 + 1\ 302b_2$$

解联立方程,得:

$$b_0 = 39.195\ 3 \quad b_1 = 9.064\ 1 \quad b_2 = 9.607\ 4$$

若预计明天比赛的气温为 35℃,体育节目播音员预计比分差为 8,那么预测这次比赛的饮料销售量为:

$$\hat{y} = 39.195\ 3 + 9.064\ 1 \times 35 + 9.607\ 4 \times 8 = 433(\text{箱})$$

二、拟合优度

用于度量简单回归模型中的精确性指标,同样也适用于多元回归。

(一)标准误差

同一元线性回归的情况一样,标准误差是对 y 值与模型估计值之间离差的一种度量,它是计算置信区间估计值和其他拟合优度的基础指标。其计算公式为:

$$SE = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 3}} \quad (3-22)$$

将表 3—4 中的有关数据代入(3—21)式,得:

二元回归分析计算表

时期 (i)	销售量 (y)	温度 (x)	比方差 (z)	xy	xz	zy	x^2	z^2	$(y-\bar{y})^2$	$(x-\bar{x})^2$	$(z-\bar{z})^2$	$(y-\bar{y})(x-\bar{x})$	$(y-\bar{y})(z-\bar{z})$	$(x-\bar{x})(z-\bar{z})$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	430	30	12	12 900	360	5 160	900	144	2 500	9	4	150	100	6
2	335	21	10	7 035	210	3 350	441	100	2 025	36	0	270	0	0
3	520	35	22	18 200	770	11 440	1 225	484	19 600	64	144	1 120	1 680	96
4	490	42	6	20 580	252	2 940	1 764	36	12 100	225	16	1 650	-440	-60
5	470	37	8	17 390	296	3 760	1 369	64	8 100	100	4	900	-180	-20
6	210	20	2	4 200	40	420	400	4	28 900	49	64	1 190	1 360	56
7	195	8	9	1 560	72	1 755	64	81	34 225	361	1	3 515	185	19
8	270	17	8	4 590	136	2 160	289	64	12 100	100	4	1 100	220	20
9	400	35	6	14 000	210	2 400	1 225	36	400	64	16	160	-80	-32
10	480	25	17	12 000	425	8 160	625	289	10 000	4	49	-200	700	-14
合 计	3 800	270	100	112 455	2 771	41 545	8 302	1 302	129 950	1 012	302	9 855	3 545	71

$$SE = \sqrt{\frac{6\,612}{7}} \doteq 31$$

把这个数据与由一元线性回归获得的标准误差数字 65 相比,多元回归的标准误差缩小了一半多。在对准确性要求更高的预测中,就能表现出这种误差缩小的好处。

表 3-4 拟合优度和德宾-沃森检验

时期 (<i>i</i>)	销售量 (<i>y</i>)	温度 (<i>x</i>)	比分差 (<i>z</i>)	$(y - \bar{y})^2$	\hat{y}	$u_i = y - \hat{y}$	$u_i^2 = (y - \hat{y})^2$	$(u_i - u_{i-1})^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	430	30	12	2 500	426	4	16	—
2	335	21	10	2 025	326	9	81	25
3	520	35	22	19 600	568	-48	2 304	3 249
4	490	42	6	12 100	478	12	144	3 600
5	470	37	8	8 100	451	19	361	49
6	210	20	2	28 900	240	-30	900	2 401
7	195	8	9	34 225	198	-3	9	729
8	270	17	8	12 100	270	0	0	9
9	400	35	6	400	414	-14	196	196
10	480	25	17	10 000	429	51	2 601	4 225
总数	3 800			129 950			6 612	14 483
总数/ <i>n</i>	380			12 995				

(二)可决系数

可决系数的计算公式为:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$$

同一元线性回归的情况一样, $R^2 = 0$ 意味着回归模型没有对 y 的变差做出任何解释, $R^2 = 1$ 则意味着回归模型对 y 的全部变差做出解释。此例中:

$$R^2 = 1 - \frac{6\,612}{129\,950} = 0.949$$

由此可见,此回归模型解释了饮料销售量变差的 94.9%,而简单回归模型只解释了饮料销售量变差的 74%。

(三)相关系数

对于多元回归可决系数而言,多元相关系数似乎是多余的,它并不提供任何新的信息,只是可决系数的平方根。

三、自相关和多重共线性问题

(一) 自相关检验

自相关是多元回归和简单回归所共同面临的问题。正如简单回归的情况一样,可用德宾—沃森统计量作为一种检验指标:

$$D-W = \frac{\sum_{i=2}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2} \quad (3-23)$$

式中:

$$u_i = y_i - \hat{y}_i$$

表 3-4 中的第(8)栏和第(9)栏是计算德宾—沃森统计量的基础:

$$D-W = 14\,483/6\,612 = 2.19$$

采用前面提出的经验法则,若 $D-W$ 统计量在 1.5~2.5 之间,则表明不存在显著的自相关问题。

如果发现了自相关问题,通常是通过对所有的原始数据进行差分来消除它:

y = 对前一期饮料销售量的变化量

x = 对前一期温度的变化量

z = 对前一期比分差的变化量

然后用这些变量代替原始变量,进行多元回归分析。

(二) 多重共线性检验

多重共线性是多元回归中出现的问题,简单回归不存在此问题。由于各个自变量所提供的是各个不同因素的信息,因此,假定各自变量同其他自变量之间是无关的。例如,假定把比赛到一半时的比分差和比赛结束时的比分差作为两个自变量,则两者中的每一个都是比分接近程度的度量,它们可能存在相关关系。这种关系会导致建立错误的回归模型以及得出使人误解的结论。为了避免这个问题,有必要对自变量之间的相关与否进行检验。任何两个自变量 x 和 z 之间的相关系数为:

$$r_{xz} = \frac{\sum (x - \bar{x})(z - \bar{z})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (z - \bar{z})^2}} \quad (3-24)$$

相关系数 r_{xz} 的取值范围为 -1 到 1。绝对值为 0 表示不相关,绝对值越接近 1 表示相关性越强, -1 表示负相关, 1 表示正相关, r_{xz} 的绝对值越接近 0 则表示 x 和 z 之间越不可能存在多重共线性;越接近 1 表示 x 和 z 之间月可能存在多重共线性。

四、用多元线性回归模型进行预测

(一) 点估计

本例中,若预计明天的气温为 35℃,体育节目播音员预计比分差为 8,那么可以估计这次比赛的饮料销售量为:

$$\hat{y} = 39.195\,3 + 9.064\,1 \times 35 + 9.607\,4 \times 8 = 433(\text{箱})$$

(二) 区间估计

(1)如果要估计的是因变量的平均水平 μ_y ,则置信区间的公式为:

大样本:置信区间 $=\hat{y} \pm z_{\frac{\alpha}{2}} SE_{\hat{\mu}}$

小样本:置信区间 $=\hat{y} \pm t_{\frac{\alpha}{2}} SE_{\hat{\mu}}$

(2)如果要估计的是某个特定的因变量 y 值,则预测区间的公式为:

大样本:预测区间 $=\hat{y} \pm z_{\frac{\alpha}{2}} SE_{\hat{y}}$

小样本:预测区间 $=\hat{y} \pm t_{\frac{\alpha}{2}} SE_{\hat{y}}$

其中, $t_{\frac{\alpha}{2}}$ 是置信度为 $(1-\alpha)$ 、自由度为 $(n-3)$ 的 t 分布的临界值。 $SE_{\hat{\mu}}, SE_{\hat{y}}$ 的计算公式为:

$$SE_{\hat{\mu}} = SE \sqrt{C_0} \quad SE_{\hat{y}} = SE \sqrt{1+C_0}$$

$$C_0 = \frac{1}{n} + \frac{(x_{10} - \bar{x}_1)^2 \sum (x_2 - \bar{x}_2)^2 + (x_{20} - \bar{x}_2)^2 \sum (x_1 - \bar{x}_1)^2}{\sum (x_1 - \bar{x}_1)^2 \sum (x_2 - \bar{x}_2)^2 - [\sum (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)]^2} - 2$$

$$\times \frac{(x_{10} - \bar{x}_1)(x_{20} - \bar{x}_2) \sum (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sum (x_1 - \bar{x}_1)^2 \sum (x_2 - \bar{x}_2)^2 - [\sum (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)]^2}$$

式中, x_{10} 与 x_{20} 分别为给定的自变量 x_1 和 x_2 的值。常用的统计软件都可以直接给出因变量的置信区间和预测区间。

本例中,要计算置信度为 90% 的估计区间,查表得 $t_{0.10/2}(7) = 1.895$,从而饮料销售量的置信区间为:

$$\hat{y} \pm 1.895 \times SE_{\hat{\mu}} = 430 \pm 25$$

即当气温为 35℃、比分差为 8 时,平均的饮料销售量的值介于 408~458 箱之内的概率为 0.90。

同样的条件下,饮料销售量的置信度为 90% 的预测区间为:

$$\hat{y} \pm 1.895 \times SE_{\hat{y}} = 433 \pm 63$$

即当气温为 35℃、比分差为 8 时,有 90% 的把握估计饮料销售量的值在 370~496 箱之间。

五、两个以上自变量的多元回归模型

使用两个自变量和使用两个以上自变量的多元回归,在概念上是没有差别的,只是后者计算更复杂一些。下式是包括许多自变量的多元回归预测模型:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$$

系数 b_0 至 b_n 一般通过把因变量和自变量的观察值输入适当的计算机程序而求得。上述关于包括两个自变量的拟合优度的度量 and 置信区间的估计,同样也适用于三个或三个以上自变量的情形。

六、计算机在多元回归分析中的应用

随着回归分析中变量的增多,回归分析的计算量成倍增加,因此,多元回归分析必须借助于计算机。很多的计算机软件如 EXCEL 等都提供有回归分析功能。下面以一个简单的实例来介绍如何用 EXCEL 软件进行多元回归分析。

将表 3-5 的数据输入 EXCEL 的工作表中,用鼠标单击 EXCEL 工具栏,点取数据分析功能,选择其中的回归分析,并输入因变量和自变量的值域,按“确定”按钮即可得到如表 3-6 所示的结果。

表 3-5 工业主要指标在国民经济中的比重 单位: %

年份	工业部门创造的国民收入占国民收入总额的比重	工业劳动者占全社会劳动者的比重	工业固定资产原值占全社会固定资产原值的比重	工业定额流动资金占全社会定额流动资金的比重
	y	x_1	x_2	x_3
1991	48.90	15.80	65.20	30.90
1992	46.70	16.00	65.00	30.90
1993	45.80	15.90	65.10	30.40
1994	45.10	15.90	65.40	30.50
1995	44.50	16.50	65.20	32.90
1996	45.10	16.70	66.30	38.60
1997	45.50	17.50	64.90	39.50
1998	45.80	17.70	65.00	40.10
1999	46.10	17.80	64.30	40.60
2000	47.40	17.30	64.30	42.40
2001	45.80	17.10	64.90	51.20
2002	47.80	17.00	64.90	51.00
2003	49.40	17.20	64.15	52.65

表 3-6 用 EXCEL 计算的回归分析结果

汇总输出						
回归统计						
R(相关系数)	0.727 247					
R ² (可决系数)	0.528 889					
调整的 R 平方	0.371 852					
标准误差	1.193 941					
观察值个数	13					
ANOVA(方差分析)						
	自由度	SS	MS	F	显著水平 F	
回归	3	14.402 86	4.800 954	3.367 923	0.068 457	
残差	9	12.829 45	1.425 494			
总和	12	27.232 31				
	系数	标准差	t 统计量	p 值	下限 95%	上限 95%
截距	170.04	52.030 67	3.268 29	0.009 71	52.348 13	287.751 4
X 变量 1	-1.433 07	0.679 058	-2.110 38	0.064 03	-2.969 21	0.103 066
X 变量 2	-1.596 5	0.731 495	-2.182 51	0.056 941	-3.251 26	0.058 261
X 变量 3	0.106 943	0.058 312	1.833 973	0.099 864	-0.024 97	0.238 855

利用 EXCEL 的回归分析功能,还可以计算残差、制作残差图及线性拟合图等。

第三节 非线性回归预测法

以上两节,我们研究了变量之间呈线性关系的预测问题。然而,在社会经济现象中,有时两个因素之间的关系并不是呈线性的关系,这时就要选配适当类型的曲线,才符合实际情况。

一、选配曲线问题

选配曲线通常可以分以下两个步骤:

(一)确定变量间函数的类型

变量间函数关系的类型有的可以根据理论或过去积累的经验,事前予以确定。但是,当不能事先确定变量之间函数关系的类型时,就需要根据实际资料作散点图,从散点图中的分布形状选择适当的曲线来配合。

(二)确定相关函数中的未知参数

函数类型确定以后,接下来就需要确定函数关系式的未知参数。最小二乘法是确定未知参数最常用的方法。但在具体运用时,必须先通过变量变换,把非线性函数关系转化为线性关系。

【例 3—3】 表 3—7 是某年某市各百货商店的商品年销售额和商品流通费率,如果把这些数据绘制一张相关散点图,并用曲线连接起来,可以清楚地看出销售额逐渐增大,商品流通费率逐渐减少,如图 3—2 所示。

表 3—7 商品年销售额和商品流通费率等数据

商品年销售额 分组(万元)	组中值 x	商品流通费率 (%) y	$x' = \frac{1}{x}$	$(x')^2$	y^2	$x'y$
3 以下	1.5	7.0	0.666 7	0.444 49	49.0	4.666 90
3~6	4.5	4.8	0.222 2	0.049 37	23.04	1.066 56
6~9	7.5	3.6	0.133 3	0.017 77	12.96	0.479 88
9~12	10.5	3.1	0.095 2	0.009 06	9.61	0.295 12
12~15	13.5	2.7	0.074 1	0.005 49	7.29	0.200 07
15~18	16.5	2.5	0.050 6	0.003 67	6.25	0.151 50
18~21	19.5	2.4	0.051 3	0.002 63	5.76	0.123 12
21~24	22.5	2.3	0.044 4	0.001 97	5.29	0.102 12
24~27	25.5	2.2	0.039 2	0.001 54	4.84	0.086 24
合 计	—	30.6	1.387 0	0.535 99	124.04	7.171 51

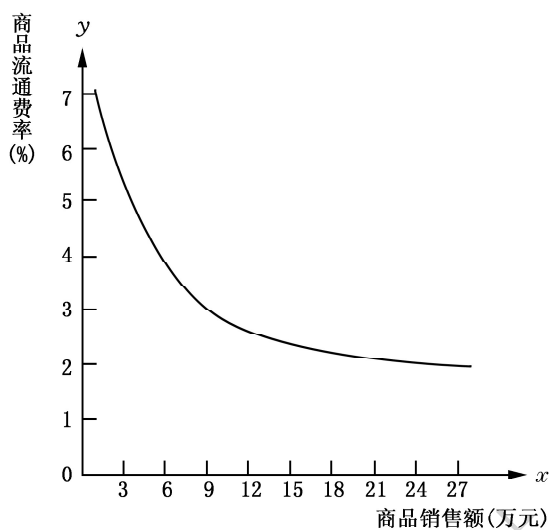


图 3-2 某年某市各百货商店商品年销售额和商品流通费率

因此,就要用双曲线回归预测方程式 $\hat{y} = a + b \frac{1}{x}$ 来描述它们之间的数量变化。

要解双曲线方程式中的 a 、 b 两个参数,必须先对其作线性变换,即假设 $x' = \frac{1}{x}$,那么:

$$\hat{y} = a + bx'$$

$$\sum x' = 1.3870$$

$$\sum y = 30.6$$

$$\bar{x}' = 0.15411$$

$$\bar{y} = 3.4$$

$$\sum (x')^2 = 0.53599$$

$$\sum y^2 = 124.02$$

$$\frac{(\sum x')^2}{n} = 0.21375$$

$$\frac{(\sum y)^2}{n} = 104.04$$

$$l_{x'x'} = \sum (x')^2 - \frac{(\sum x')^2}{n} = 0.32224$$

$$l_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 19.98$$

$$\sum x'y = 7.17151$$

$$\frac{\sum x' \sum y}{n} = \frac{1.3870 \times 30.6}{9} = 4.7158$$

$$l_{x'y} = \sum x'y - \frac{\sum x' \sum y}{n} = 2.45571$$

$$b = \frac{l_{x'y}}{l_{x'x'}} = \frac{2.45571}{0.32224} = 7.62075$$

$$a = \bar{y} - b\bar{x}' = 3.4 - 7.62075 \times 0.15411 = 2.22557$$

得回归方程为:

$$\hat{y} = 2.22557 + 7.62075 \left(\frac{1}{x} \right)$$

当商品年销售额为 13 万元时,则预测流通费率为:

$$\hat{y} = 2.225\ 57 + 7.620\ 75 \times \frac{1}{13} = 2.8(\%)$$

各百货商店商品年销售额与商品流通费率的相关系数为:

$$r = \frac{l_{x'y}}{\sqrt{l_{x'x'}l_{yy}}} = \frac{2.455\ 71}{\sqrt{0.322\ 24 \times 19.98}} = 0.968$$

我们也可以利用方差分析对回归方程进行显著性检验。检验结果如表 3—8 所示(F 检验临界值表见附表三)。

表 3—8 双曲线回归方差分析

变差来源	平方和	自由度	均方差 (F)	F_α	显著性
回 归	$U = bl_{x'y}$ $= 18.716\ 68$	1	$F = \frac{U}{\frac{Q}{(n-2)}}$ $= 103.30$	$F_{0.01}(1,7)$ $= 16.24$	16.24 小 于 103.30, 高度显著
剩 余	$Q = l_{yy} - bl_{x'y}$ $= 1.268\ 32$	7			
合 计	19.98	—	—	—	—

二、一些常见的函数图形

选择合适的曲线类型并不是一件轻而易举的工作,主要依靠专业知识和经验,当然也可以通过计算剩余均方差(估计标准误差)来确定。为了便于选择曲线类型,现给出常用的几种曲线,供使用时参考。

(一)幂函数 $y = ax^b$

如果 x 与 y 的关系可用下列曲线函数表示:

$$y = ax^b$$

那么这个函数称为幂函数。幂函数曲线如图 3—3 所示。

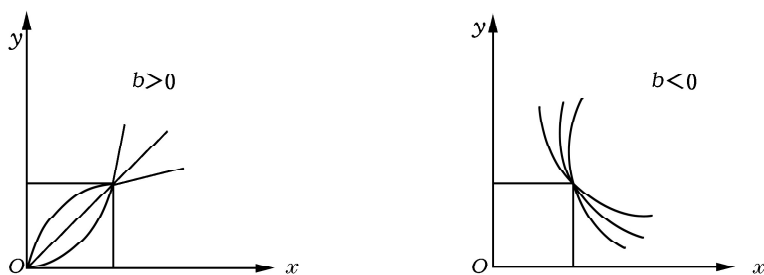


图 3—3 幂函数曲线 $y = ax^b$

如果把 x 与 y 的数值绘于对数坐标轴上,则成为直线。

令:

$$y' = \lg y, \quad x' = \lg x, \quad a' = \lg a$$

则:

$$y' = a' + bx'$$

(二)指数函数 $y = ae^{bx}$

指数函数方程为:

$$y = ae^{bx}$$

令:

$$y' = \ln y, \quad a' = \ln a$$

则:

$$y' = a' + bx$$

指数曲线如图 3-4 所示。

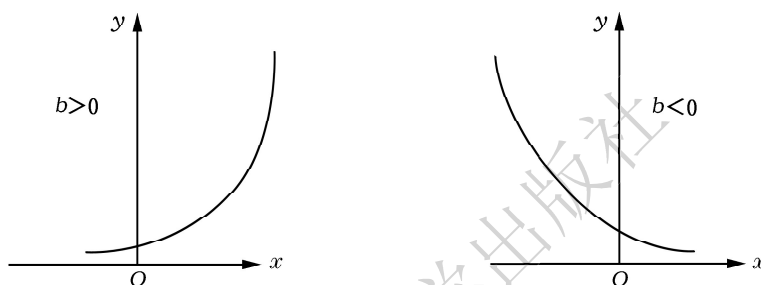


图 3-4 指数曲线 $y = ae^{bx}$

(三)抛物线函数 $y = a + bx + cx^2$

按最小二乘法确定 a, b, c 之值,应先解下列三个规范方程式:

$$\begin{cases} \sum y = na + b \sum x + c \sum x^2 \\ \sum xy = a \sum x + b \sum x^2 + c \sum x^3 \\ \sum x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4 \end{cases}$$

抛物线如图 3-5 所示。

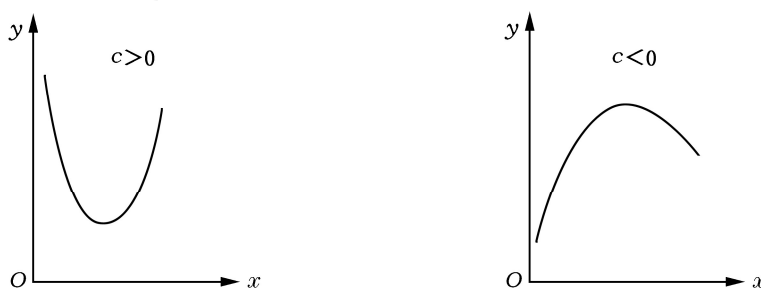


图 3-5 抛物线 $y = a + bx + cx^2$

(四)对数函数 $y = a + b \lg x$

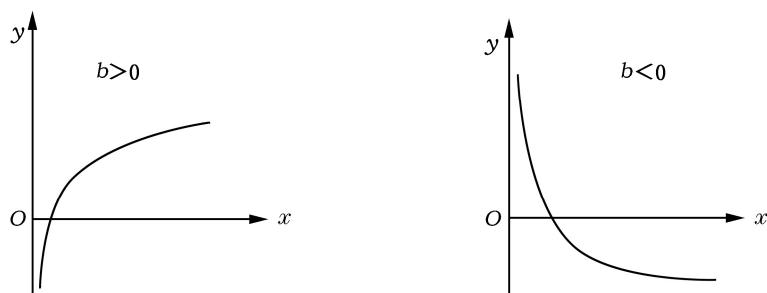
令:

$$x' = \lg x$$

则:

$$y = a + bx'$$

对数曲线如图 3-6 所示。

图 3-6 对数曲线 $y=a+b\lg x$

(五) S 形函数 $y = \frac{1}{a + be^{-x}}$

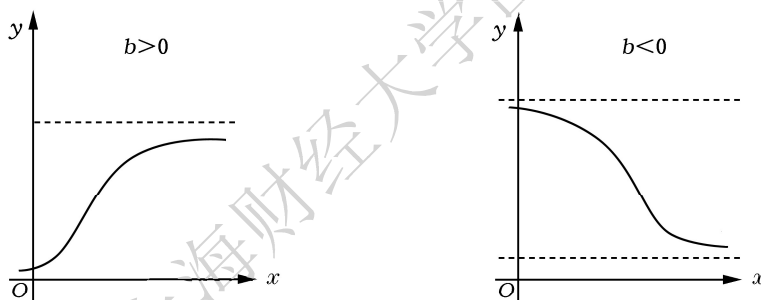
令:

$$y' = \frac{1}{y}, \quad x' = e^{-x}$$

则:

$$y' = a + bx'$$

S 形曲线如图 3-7 所示。

图 3-7 S 形曲线 $y = \frac{1}{a + be^{-x}}$

第四节 应用回归预测法应注意的问题

随着我国社会主义市场经济的发展,人们将不断地借助回归预测法对市场经济中出现的各种数量关系进行描述、分析、控制和预测。然而,在具体运用时,应注意以下几个问题:

一、关于定性分析问题

社会现象之间的相互关系和发展规律性,取决于马克思主义的理论分析。在理论上证实社会现象或社会现象指标之间确实存在相关关系性质之后,才能再利用回归预测法具体研究和测定社会现象相关关系的数量表现。回归分析不能代替马克思主义理论对社会现象相互关系的质的分析,只有在马克思主义理论的质的分析的基础上,才能测定社会现象在数量上的相互关系。这应该成为统计上运用回归预测法的一条基本原则。

因此,我们首先应该进行定性分析,在马克思主义理论指导下,依靠研究人员的理论知识、专业知识、实际经验和分析研究能力,来确定各种变量之间的相关关系及其影响程度。

应该指出,我们必须积极把握事物从量变到质变的“度”的界限。在许多情况下,现象之间只是在一定的范围内才具有相关关系,超出了这个范围,就可能成为荒谬。例如,施肥量和农作物生产量只有在一定范围内才具有正相关关系,施肥量超过了一定的限度,产量不但不会增加,反而会减少。农作物的耕作深度与每亩收割量之间的关系也是如此。因此,在运用时要注意它的作用与范围,超过了这个范围去推断或预测,可能会得出错误的结论。

二、关于回归预测不能任意外推的问题

回归分析的应用,仅仅是限于原来数据所包括的范围内,相关关系只限于 x 和 y 两方面的数值从最小值到最大值的范围;回归分析中的回归关系限于 x 方面原数值从最小值到最大值的范围(以 x 推算 y 时)。所谓外推,是指把相关关系或回归关系用于超出上述范围之外。由于原来资料只提供了一定范围内的数量关系,在此范围以外是否存在同样的关系,尚未得知。如果有进行外推的充分根据和需要,也应十分慎重,而且不能离开原来的范围太远。

统计学已阐明,回归方程实际观察值 y 的偏差 δ 不仅与显著性水平 α 有关(α 越小, δ 越大),与 n 有关(n 越大, δ 越小),而且与观察值 x_0 有关。当 x_0 靠近 \bar{x} 时, δ 就小;当 x_0 远离 \bar{x} 时, δ 就大,即 $\delta = \delta(x_0)$ 。假如分别作函数 $y = \hat{y} - \delta$ 和 $y = \hat{y} + \delta$ 的图形(见图 3-8),那么,它们把回归直线夹在中间,两头都呈喇叭形。

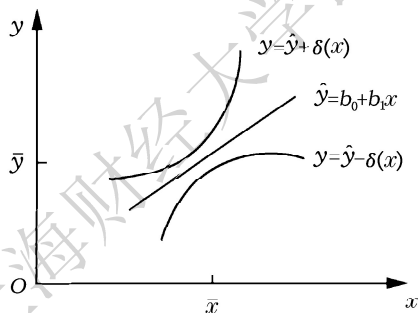


图 3-8 小样本预测区间

显而易见,当 x_0 处于 \bar{x} 时,回归直线效果最好。但是,回归直线任意外推,预测远期目标,误差就大,往往会得出不正确的结论。当然,这不能一概而论,一般来说,直线回归只适宜做中、短期预测,不宜于做长期预测。

三、关于对数据资料的要求问题

在利用回归分析进行预测时,还必须注意数据资料的准确性、可比性和独立性问题。

(一)关于数据资料的准确性问题

这个问题是容易理解的,只有借以预测的资料是正确可靠的,才能保证分析和预测的可靠性。如果数据是凭经验、拍脑袋估计出来的,就不能得出科学的分析结论。如果由于历史上的某些原因使得某一年度的资料明显不准确(如“大跃进”时期一些地区的数据资料),应该按照核实后的数据来计算,不可将错就错。在整理资料过程中,如果发现个别因素缺少某些年度的数字,可采用一定的统计方法(如比例推算法、统计插值法、调查估算法等)予以补齐。如果发现某一年度的数字畸高畸低,可利用数理统计中的控制理论,按照 3σ 原则对该

数字进行检验,如与总体平均数的离差超过 3δ ,那么该年度各个变量的数值就不能用来分析和推断。

(二)关于数据资料的可比性和独立性问题

我们应该保证 y_1, \dots, y_n 之间的指标数值所包含的经济内容、指标的口径、范围、计算方法和计量单位的一致性,并且,各年的指标应当是当年的生产成果。

(三)关于社会经济现象基本稳定的问题

回归分析是在假定社会经济条件没有发生重大变化、社会经济现象基本稳定的情况下进行的,即假定工艺、技术市场以及国家政策等相对稳定、没有突变的情况下进行的。如果由于某年某个企业发生了重大的技术革命,使生产量成倍地增长,那么,这一年前后的数字就不能合并在一起进行回归预测。

总之,社会经济现象的发展变化是复杂的,这就要求我们在进行回归预测时,必须估计到社会因素的变化,以修正分析的结论。



本章小结

1. 一元线性回归预测是指成对的两个变量数据分布大体上呈直线趋势时,运用合适的参数估计方法,求出一元线性回归模型,然后根据自变量与因变量之间的关系,预测因变量的趋势。

由于很多社会经济现象之间都存在相关关系,因此,一元线性回归预测具有很广泛的应用。进行一元线性回归预测时,必须选用合适的统计方法估计模型参数,并对模型及其参数进行统计检验。

2. 社会经济现象的变化往往受到多个因素的影响,因此,一般要进行多元回归分析,我们把包括两个或两个以上自变量的回归称为多元回归。

多元回归与一元回归类似,可以用最小二乘法估计模型参数。也需对模型及模型参数进行统计检验。选择合适的自变量是正确进行多元回归预测的前提之一,多元回归模型自变量的选择可以利用变量之间的相关矩阵来解决。

3. 在现实经济生活中,很多现象之间的关系并不是线性关系,对这种类型现象的分析预测一般要应用非线性回归预测,通过变量代换,可以将很多的非线性回归转化为线性回归。因此,可以用线性回归方法解决非线性回归预测问题。

4. 应用回归预测法时,应首先确定变量之间是否存在相关关系。如果变量之间不存在相关关系,对这些变量应用回归预测法就会得出错误的结果。正确应用回归分析预测时应注意:(1)用定性分析判断现象之间的依存关系;(2)避免回归预测的任意外推;(3)应用合适的的数据资料。



思考与练习

- 应用回归预测法进行预测时,应注意哪些问题?
- 某种商品的需求量与人均月收入的关系数据如下表所示:

人均月收入(元)	700	800	900	1 000	1 100	1 200	1 260	1 340
需求量(万元)	9.0	9.6	10.2	11.6	12.4	13.0	13.8	14.6

如果估计下月的人均月收入为 1 400 元,试预测下月该商品的需求量(取置信度 $\alpha=0.05$)。

3. 设某公司的每周广告费支出和每周销售额数据如下表所示:

每周广告费支出 (元)	4 100	5 400	6 300	5 400	4 800	4 600	6 200	6 100	6 400	7 100
每周销售额 (万元)	12.50	13.80	14.25	14.25	14.50	13.00	14.00	15.00	15.75	16.50

要求:

(1)广告费支出与销售额之间是否存在显著的相关关系?

(2)计算回归模型参数。

(3)回归模型能解释销售额变动的比例有多大?

(4)计算 $D-W$ 统计量。

(5)如下周的广告费支出为 6 700 元,试预测下周的销售额(取置信度 $\alpha=0.05$)。

4. 某超市顾客的付款时间与所购商品价值之间的关系数据如下表所示:

付款时间 (分钟)	3.6	4.1	0.8	5.7	3.4	1.8	4.3	0.2	2.6	1.3
商品价值 (元)	306	305	24	422	218	62	401	20	155	65

要求:

(1)付款时间与所购商品价值之间是否存在显著的相关关系?

(2)计算回归模型,并作相关统计检验。

(3)试构造当付款时间为 3 分钟时,所购商品价值的置信度为 99% 的置信区间。

5. 下表是国外某研究小组的调查结果。试问:受教育程度与年收入之间是否存在线性相关?并计算回归模型、可决系数 R^2 ,以及解释 R^2 的意义。

小 组	平均受教育年限 (X)	平均年收入 (1 000 美元)(Y)
1	3	3
2	4	4
3	5	5
4	7	7
5	6	9
6	9	11
7	11	13
8	12	20
9	16	35
10	15	40
11	21	50

6. 某公司下属企业的设备能力和劳动生产率的统计资料如下表所示:

企业代号	01	02	03	04	05	06	07	08	09	10	11	12	13	14
设备能力 (千瓦/人)	2.8	2.8	3.0	2.9	3.4	3.9	4.0	4.8	4.9	5.2	5.4	5.5	6.2	7.0
劳动生产率 (万元/人)	6.7	6.9	7.2	7.3	8.4	8.8	9.1	9.8	9.8	10.7	11.1	11.8	12.1	12.4

该公司现计划新建一家企业,设备能力为 6.2 千瓦/人,试预测其劳动生产率,并求出其 95%的置信区间。

7. 某公司 10 家下属企业的产量与生产费用之间的关系如下表所示:

产量(万件)	40	42	48	55	65	79	88	100	120	140
生产费用(万元)	150	140	160	170	150	162	185	165	190	185

要求:

(1)计算可决系数 R^2 和相关系数。

(2)计算回归模型参数,并作统计检验。

8. 下表是从 10 个地区调查得来的数据。因变量是表示喜欢某种牌号牙膏的居民百分比,自变量则是该地区居民的人均年收入(x_1 ,以千元计量)和关于居民受教育水平的某种量度(x_2)。

要求:

(1)计算回归模型参数估计值。

(2)计算可决系数。

(3)计算 $D-W$ 统计量。

(4)设 $x_1=5$ 和 $x_2=6$,构造 Y 的置信度为 95%的预测区间。

地 区	喜欢该品牌牙膏 的百分比(Y)	人均年收入 (x_1)	教育指数 (x_2)
1	61.6	6	6.3
2	53.2	4.4	5.5
3	65.5	9.1	3.6
4	64.9	8.1	5.8
5	72.7	9.7	6.8
6	52.2	4.8	7.9
7	50.2	7.6	4.2
8	44	4.4	6.0
9	53.8	9.1	2.8
10	53.5	6.7	6.7

9. 某大企业人事处长想利用职员的年龄和经验做自变量来预测他们的技术水平。下表所列数据是来自一个由 15 名职员构成的随机样本。

要求:

(1)求多元回归方程。

(2)设 $x_1=2, x_2=25$,求 Y 的置信度为 95%的置信区间。

技术水平(Y)	经 验(x_1)	年 龄(x_2)
15	0	21
15	0	18
21	0	22
28	1	24
30	1	25
35	1	25
40	1	26
35	2	34
30	2	25
45	2	38
50	3	44
60	3	51
45	4	39
60	4	54
50	5	55