

摘 要

近年来，随着经济的飞速发展，我国的航运事业也有了巨大的市场，船舶也逐渐向大型化、自动化、集装箱化方面发展。这些都为原来制定的有关航运方面的规定和标准提出了新的要求。所以对跨航道桥梁和架空线缆的通航净空高度进行研究也显得非常重要。

本文从收集实船数据入手，并对收集到的数据按船型进行归类，将实船数据的空载水面以上至最高固定点高度和船舶的载重吨进行非线性回归，找出它们之间的定量关系，建立数学模型。同时本文还对影响富裕高度的各项因素进行了定性分析，给出了各个小的分量的计算方法，尤其是船舶在风浪中的垂荡值的计算。这样当设计跨海桥梁和架空线缆的通航净空高度时，就可以按给定的船型在给定的吨位级别，就可以快速的得到他的空载水面上以上的最大高度，并且根据各个海区的实际情况，将各个影响富裕高度的因素考虑进去，就可以得到一个较为理想的净空高度值，为计跨海桥梁和架空线缆的通航净空高度提供一个新的参考标准。

关键词：净空高度；非线性回归；船舶最大高度；数学模型；悬链线

Abstract

In recent years, with the fast development of economy in our country, shipping industry is confronted with a enormous market demand. The evolution of morden ships is to be more bigger and more automatous. All of these above require new criterions for shipping industry to take the place of the old ones, because the old ones are becoming more restrictive to shipping industry. So it is necessary to research the clearance of a bridge and wire over strait.

This thesis starts with collectting data of ships, sort out them according to ship type, and then use the mathematical methods to find out the correlation between airdraft and ship's tonnage and upbuild mathematical model. At the same time ,this thesis also research all factor concerned with the safety heights and supply with mathematical methods to calculate them. when we design clearance of a bridge or wire over strait ,we can consult this.

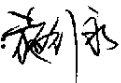
Key Words: non-linear curve regress; clearance height of over strait bridges and wires; airdraft, catenary; mathematical mode

大连海事大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：本论文是在导师的指导下，独立进行研究工作所取得的成果，撰写成博士/硕士学位论文“跨海桥梁和架空线缆净空高度的研究”。除论文中已经注明引用的内容外，对论文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本论文中不包含任何未加明确注明的其他个人或集体已经公开发表或未公开发表的成果。

本声明的法律责任由本人承担。

论文作者签名： 2006年3月18日

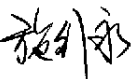
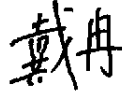
学位论文版权使用授权书

本学位论文作者及指导教师完全了解“大连海事大学研究生学位论文提交、版权使用管理办法”，同意大连海事大学保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权大连海事大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。

保密，在_____年解密后适用本授权书。

本学位论文属于： 保密

不保密（请在以上方框内打“√”）

论文作者签名： 导师签名：
日期：2006年3月18日

第一章 绪论

1.1 本文研究的背景

随着我国经济的快速发展，我国的航运事业也有突飞猛进的发展，新建码头泊位如雨后春笋一样，遍布在祖国的海岸线上。据有关统计显示，经过近几年的水工基础设施建设，目前中国年吞吐量万吨以上的港口达二千余个，其中万吨级深水泊位五百九十一一个，对外开放港口一百二十多个，每年接纳世界一百多个国家和地区的船舶三万六千多艘。上海、青岛、深圳、广州和天津港已跻身于世界五十大集装箱港口之列。目前，中国设立的中外合资航运企业已达一百二十多家，中国政府已和五十二个国家签订了航运协定。

航运事业的飞速发展，导致很多航道两侧地区各种资源交流和共享的需求也越来越多，有些岛屿上的资源已不能满足各种社会生产和人们生活的需要，需要从大陆得到更多的资源支持，这就需要建设更多的跨海桥梁和架空线缆满足资源共享和交流。而且近年来造船业也在快速的发展，为了提高船舶营运效率，船舶逐渐向大型化、自动化方向发展，这些都给跨海桥梁和架空线的净空高度设计提出了新的要求。

在所有的临、跨、过航道的建筑设施中,桥梁与架空线缆对航道的的影响最大,而且桥梁和架空线缆的通航净空高度直接影响航道的通航能力和船舶航行安全,如果航道上的桥梁和架空线缆的通航净空高度不能满足航道远期通航需要会限制航道附近的航运事业的发展。1968年建成的某长江大桥按通航五千吨级船舶标准设计的,净空高度24米,现已成为长江下游航道上的卡脖,严重阻碍了长江航运和地区经济的发展。所以对桥梁和架空线缆的通航净空高度的控制,将对航道的发展和船舶的航行安全带来深远的影响。

1.2 国内外研究现状及存在的问题

1.2.1 国内外研究现状

目前，国内外对这方面研究的较少，主要是根据经验和有关规定来确定跨航

道桥梁和架空线的净空高度。根据以往经验,确定跨航道大桥净空高度,其中最为主要的因素是代表船型的确定。代表船型一般应依据跨海大桥设计使用年限内通航水域附近港口的发展规划来确定,由此确定大桥所跨越的自然航道通航船舶等级状况(即满载时可通航吨位船舶、压载时可通航吨位船舶以及空载时可通航吨位船舶)及通航船舶种类。并广泛收集不同种类和吨位等级的船舶在不同载重状态下的水线以上高度,经过整理和比较之后可确立基本船型。当通航的代表船型确定后,主要是考虑富裕高度的选取。富裕高度包含了很多不太确定的因素,选取时应全面考虑,根据以往经验主要从以下几个方面考虑:

(1) 船舶驾驶安全高度

船舶驾驶安全高度是一个综合性的因素,其中包括驾驶员的心理素质。船舶越大,波浪越高,船舶过桥时航行条件越差,要求的安全高度也越大。

(2)水流以及波浪引起的船舶纵摇和垂荡使船舶水线面以上至最高固定点高度产生较大的变化。

(3)船舶水线面以上至最高固定点高度的计量误差有以下情况造成:不同的载重量,船舶大修后发生的高度变化,船舶高度测量时的误差,船舶在淡水、海水中的吃水差等;

富裕高度的取值,国际上一般取2~4米,视船舶的大小和水域的环境而定。通航海轮的内河江面宽阔,可能形成较大的风浪,考虑到这些河流多为重要的水上运输航线,行驶的船舶较大,因此富裕高度取2米。长江、黄河、黄浦江和瓯江等河流上新建、拟建的桥梁富裕高度均取2米。在无掩护的海域,即使风浪不大,也容易受外海涌浪的影响,富裕高度取2米较为合适。

波浪较大的开敞海域,船舶纵摇和垂荡幅度大,船舶驾驶的安全高度也要求更高,同时由于航道的重要程度高,航道对国民经济的发展作用大,过往船舶航行密度大,船舶吨位大,在该水域建设桥梁和架空线缆,富裕高度应有较高的要求,通常取4米。

1.2.2 存在的问题

大家都知道一说到经验,就是些模式化的东西,不可否认大部分经验还是很有指导作用的,可是我国的海岸线自鸭绿江至北仑河口,长达1.8万多公里。此有指导作用的,可是我国的海岸线自鸭绿江至北仑河口,长达1.8万多公里。此

外岛屿海岸线长 1.4 万多公里，大陆海岸线总长 3.2 万多公里。同时海岸线曲折，较大海湾有 150 多个。地区不一样，气候也不一样，不同海区航道附近的风、流也是差别很大。同时随着航运事业的快速发展，船舶逐渐向大型化、自动化方向发展，以前总结的经验是否适应新的大型船舶通航的桥梁和架空线缆，这些都是有待解决的问题。所以对跨海桥梁和架空线缆的通航净空高度的研究也变得重要起来。

1.3 本文主要研究的内容及方法

1.3.1 本文主要研究的内容

本文主要研究的是跨海大桥和架空线缆的通航净空高度，主要从以下几个方面进行研究：

(1) 船舶空载水线面以上至最高固定点的高度与船舶吨位关系的研究

在此部分笔者将收集大量的实船数据，并以实船基本数据为依据，对基本数据作统一处理，再运用非线性曲线拟合的方法建立数学模型，找出各类型船舶的空载水线面以上至最高固定点高度与船舶吨位的关系。理论上讲，本文将会得到具体的船型，在吨位给定的情况下其空载水线面以上至最高固定点高度的一个估值，为建设跨航道桥梁和架空线的净空高度的设计针对不同等级的航道提供一个新的参考标准。

(2) 富裕高度的研究

此部分主要对影响富裕高度值的因素进行分析和研究，重点是对不同船型、不同吨位的船舶在风浪中的垂荡值进行研究。同时考虑到大桥受到重压会发生微小的下沉，以及架空线缆受到温度的影响，其垂度也会发生微小的下沉和上升，这些都对富裕高度的取值有较大影响，应认真分析和研究。

(3) 跨海桥梁和架空线缆通过能力的研究

因为无论是桥梁的通航桥孔还是架空线缆，各个点的通航净空高度都是不一样的，若航道的宽度很大，此方面的计算将显得很重要。此部分主要研究通航桥孔和架空线缆各个点的通航净空高度的计算和分析。

1.3.2 本文研究的方法

(1) 本文首先介绍了船舶主尺度和跨航道桥梁及架空线缆通航净空高度的基本概念、船型介绍，因本文的主旨是对不同船型的船舶空载水线面以上最大高度与载重量基本数据进行统计回归，得出它们之间的定量关系，找出各种船型的船舶空载水线面以上至最高固定点的分布规律。

(2) 介绍了回归分析的基本理论，回归分析是一种从事物变化因果关系出发来进行预测的方法，亦称为“因果预测法”。它是通过对变量的观测数据进行统计分析，来确定变量之间联系的一种有效方法。

(3) 介绍了回归模型的选择方法，在数学模型的选择上我们使用两个重要的回归统计量F统计量和决定系数 r^2 来衡量模型的优良程度，并选择其中最优的一个作为我们的回归模型。

(4) 数据处理及数学模型的建立，利用相关软件对数据进行了处理，建立了不同船型的各个吨位级的船舶空载水线面以上最大高度与载重量之间关系的数学模型，同时做出了不同船型的各个吨位级的船舶空载水线面以上最大高度与载重量之间关系的散点图与拟合图，从图上可以直观的看出不同船型的船舶空载水线面以上最大高度与载重量的关系。

(5) 分析研究架空线缆和桥梁通航孔的弧度，得出一个算式，利电脑软件将架空线缆和桥梁通航桥孔水平面上各个点的通航净空高度计算出来。并划出通航高度分布图，可以直观地看到架空线缆各段下的通航高度，并将该高度分布图标到海图上，作为船舶穿越跨海大桥和架空线缆航行时的依据。

第二章本文涉及到的相关概念与算法

2.1 船舶主尺度的基本概念

船舶主尺度是用以表示船舶大小和特征的几个典型尺度，包括有船长、船宽、船深（或船高）和吃水等。船舶主尺度按不同用途和丈量规则可分为最大尺度、登记尺度和船型尺度等三种，如下图所示。

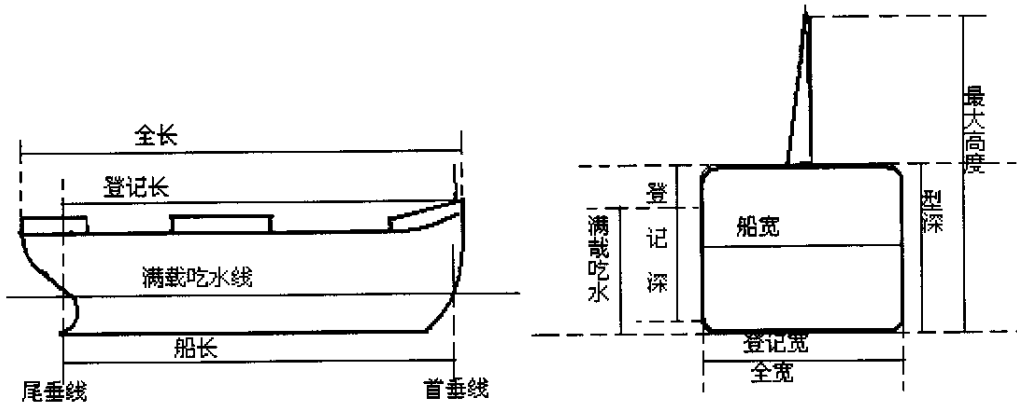


图 2.1 船舶尺度示意图

Fig.2.1 Measure of ships

2.1.1 最大尺度

也称全部尺度或周界尺度。这种尺度在实际工作中有着重要的意义。它决定某一船舶能否停靠一定长度的码头、通过或进入一定长度和宽度的船闸或船坞，决定船舶在狭窄航道和港内的安全操纵和避让，以及能否顺利通过横跨航道上的桥梁和架空电缆等。

船舶总长：是从船首最前端量到船尾最后端的水平距离。若船首最前端与船尾最后端两端外板上有永久性固定突出物（如顶推装置）也应包括在内。

(1) 最大宽度：是指船体最宽部分处两舷的船壳板外缘间的水平距离。若有固定的舷边设备或护舷材也应包括在内。

(2) 最大高度：是指从龙骨下边到最高桅顶的垂直距离，也称连桅高度。它包括任何桅顶构件最高点的垂直距离。

(3) 由最大高度减去船舶实际吃水，即得船舶在水面上的高度。船舶水面上

高度决定船舶能否通过航道上的桥梁或架空电缆。

2.1.2 登记尺度

它是丈量船舶、计算船舶吨位的尺度，该尺度登记在船舶丈量证书上，表明船舶大小。根据我国颁布的“1985年海船丈量规范”的规定，国际航行船舶与国内航行船舶其丈量方法不同。

2.1.2.1 对国际航行船舶：

1. 登记长度：指自龙骨板上缘的最小型深85%处水线长度的96%，或沿该水线从首柱前缘量至上舵杆中心长度，取两者中较大者。

2. 登记宽度：在登记长度中点处两舷船壳板内缘间的水平距离。

3. 登记深度：在登记长度中点处两舷侧处，自平板龙骨板上缘至甲板下的垂直距离。

2.1.2.2 对国内航行船舶：

1. 登记长度：在量吨甲板上表面，从首柱前缘量至舵后缘的水平距离。没有舵柱的船舶量至舵杆中心。

2. 登记宽度：在中剖面的最宽处两舷的壳板内缘间的水平距离。

3. 登记深度：在登记长度中点处，自舷侧处甲板下缘至内底板上缘的垂直距离。无内底板船量至肋板上缘。

2.2 船型尺度及船舶分类

2.2.1 船型尺度

也叫理论尺度或计算尺度。船舶设计中主要是用船型尺度，它是计算船舶稳性、吃水差、干舷高度、船舶系数和水对船舶阻力时使用的尺度。

1. 型长：也叫船长。指在夏季载重线上自船首柱前缘至船尾柱后缘的水平距离（也叫两垂线间长）。没有尾柱的船舶量至舵杆中心。

2. 型宽：在船体最宽处两舷壳板内缘之间的水平距离。

3. 型深：在船长中点处沿船舷由平板龙骨上缘至上甲板下缘的垂直距离

2.2.2 船舶分类

(1) 散货船

散装船是专门用来装运煤、矿砂、盐、谷物等散装货物的船舶。与杂货船不同的地方是它运输的货物品种单一、货源充足、装载量大。依照不同的散装货品种，装卸时可采用大抓斗、吸粮机、装煤机、皮带输送等专门的机械。不象杂货船那样装的是包装或箱装等杂货，规格大小不一，理货时间长，运输效率低。因此散货船比杂货船的运输效率高，装卸速度快。散装货船的特点是：驾驶室和机舱都设在屋艏部；货舱口比杂货船的货舱口宽大；内底板与舷侧用斜边板连接，使货物能顺利地向货舱中央集中；有较多的压载水舱，作为空载返航时压载之用。散货船都为单甲板船，甲板下面两舷与舱移动，防止船的稳性变化。空载时还可以利用顶边舱作为压载水。

(2) 油船

油船是专门运载石油类液货的船只，它在外形上和布置上很容易与一般的干货船区别开来。油船上层建筑和机舱设在艏部，上甲板纵中部位，布置纵通全船的输油管及步骤。石油分别装在各个油舱内，油船在装卸石油时是用油泵和输油管输送的，因此它不需要货吊杆和起货机，甲板上也不需要大的货舱开口。油船的机舱设在艏部，这样可以避免浆轴通过油舱时可能引起的轴隧漏油和挥发出可燃气体引起爆炸的危险。此外，机舱设在艏部，烟囱排出烟时带出的火星向后吹走，不致落入油舱的通气管内引起火灾。油船的油舱内装有蒸汽加热管路，当温度氏时石油的粘度增加，不容易流动，有了加热管加温舱石油，就可使石油流动，便于装卸。油船船体系结构通常是单层甲板，单层低结构，但也有双层底结构，油船的干舷很小，满载航行时，甲板离水面很近。

(3) 集装箱船

集装箱船运输是将杂货件预先装在金属或塑料制成的货柜体箱中，这种货柜体箱称为集装箱。装货时将客观存在直接装在船口上，然后转运到目的地，这种特制的船舶称为集装箱船。这种运输方法大大地提高了杂货运输的效率。集装箱船上所载的集装箱都已形成标准化。每箱按不同规格，可装货 3-10 吨。一台专用的起货机每小时可装卸 1 千吨箱货，一艘万吨的集装箱船只用 10 个小时即可装卸

完毕。集装箱船每年的运输量比同样大小的杂货船平均可提高一倍左右。不过这种货船一定要在港口有专用装卸设备和具有陆上联运条件时才能发挥作用。

(4) 杂货船

货船有干货船和液货船之分。杂货船是干货船的一种，它是装载一般包装、袋装、箱装和桶装的普通货物船，一般所称的万吨级货船，是指它的载货量在一万吨左右或一万吨以上，而其总载重量和满载排水量则还要大得多。万吨杂货船一般都是双层甲板船，有 4-6 个货舱，每个货舱的甲板上设有货舱口，货舱的两旁装有能起至 5-20 吨的吊货杆。有些船上还备有起吊重货的重型吊杆，起重起力可达 60-150 吨。货船按机舱位置的不同，有所谓中机型船，艏机型船和中后机型船，中机型船的机舱位置在船体中央部分，艏机型船的机舱设在船的艏部，中后机型船的机舱在偏艏部方向。万吨级大舱口远洋货船具有大的货舱口和能起至 120 吨的 V 型起重吊杆，能适应装运特大、特长、特重货物的需要。

(5) 滚装船

滚装船是一种载货汽车或拖车直接由码头开进大舱或由大舱开到码头。进行装卸货的货船。由于减少了码头中转、装卸的环节，因而运输效率大大提高，有的滚装船，它的大舱口内装载汽车，上层甲板上还可载运旅客。

(6) 液化气体船

液化气体船是液货船中的一种，这种船上装有特殊的高压液舱。运输时，先把天然气或石油气体液化，再用高压泵打入液舱内。液舱一般分薄壁液舱和球形液舱。液化天然气在运输途中要蒸发，可把这部分蒸发的天然气送到锅炉燃烧，减小损耗，所以液化天然气船的动力装置都选用蒸汽轮机。

(7) 冷藏船

冷藏船是专门运输易腐的鲜货船，例如装运新鲜的鸡、鸭、鱼、肉、蛋、水果、蔬菜的冷冻食品等。冷藏船好象一座水上活动的冷库。冷藏船按所装货物的品种不同，要求不同的冷藏温度。对冷藏舱和冷藏装置的要求是：1) 冷藏舱的上下甲板之间或甲板和舱底之间的高压不能太大，以防止上层物压坏下层货物，2) 冷藏舱周围的壁应有良好的绝热装置；3) 冷藏舱的绝热层必须有良好的防潮设备；4) 船上的制冷压缩机及其设备在船舶摇摆、振动以及高温高湿的条件下必须保证正常工作，与盐雾、海水接触的部件要采取防腐措施；5) 船舶冷藏负荷变化

较大，对制冷压缩机能量的要求比陆地冷库高得多，需要留有适当的余量。

2.3 跨航道桥梁和架空线缆的通航净空高度的定义

跨航道桥梁通航净空高度^[1]是指代表船型的船舶或船队安全通过桥孔的最小高度,起算面为设计通航最高水位(大部分以平均大潮高潮面为准)。系统净空数值为代表船型空载水线以上至最高固定点高度与富裕高度之和。即:大桥净空高度值=代表船型空载水线以上至最高固定点高度+富裕高度(如图 2-2)所示。这一高度应保证在允许航行的天气条件下,任何时候、任何情况代表船型的船舶和船队都能安全通过。考虑到船舶有空载过桥的情况,净空高度应保证代表船型在空载状态也能顺利通航。

架空线缆通航净空高度是指代表船型的船舶货船队安全通过架空线的最小高度,起算面为设计通航最高水位。系统净空数值为代表船型空载水线以上至最高固定点高度与富裕高度之和。即:架空线缆净空高度=代表船型空载水线以上至最高固定点高度+富裕高度。如图(2.2)所示。不过这里的富裕高度里增加了线缆的电气安全距离。

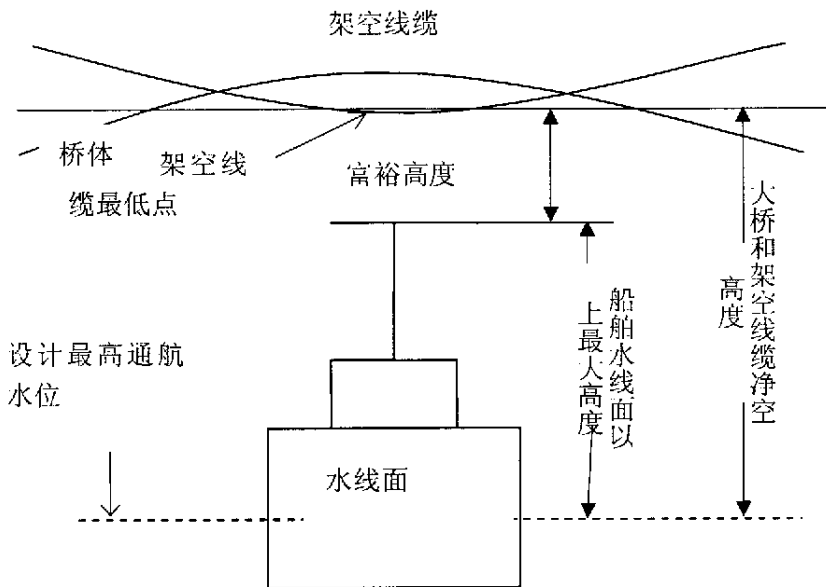


图2.2跨航道桥梁与架空线缆净空高度示意图

Fig.2.2 Sketch map of clearance height of bridges and wires over strait

2.4 架空线缆相关概念

2.4.1 相关概念

(1) 架空线

悬挂在塔杆上的相线和顶上的避雷线叫架空线，力学计算时可以将它们看作匀质柔索，可以任意弯曲，只能承受拉力而不能承受压力。

(2) 档距 l

相邻两杆间的水平距离，称之为档距，用 l 表示。 l 的大小取决于技术、地理条件、可靠性等因素。

(3) 垂度 f

通常指导线中一个档距内最低点与连接两悬点水平线（或斜线）之间的垂直距离。影响 f 值大小的因素有导线材料的容重、覆冰情况、环境温度、档距大小等

2.4.2 架空线缆垂度计算的三种方法和误差分析

(1) 平抛物线法

材料力学中介绍此法。但未指明这就是平抛物线法。它对垂直载荷作如下假定：不论是等高悬挂点还是有高差悬挂点的档距，均认为垂直载荷沿档距长度均匀分布，而真实情况是载荷沿曲线均匀分布。如图（2.3）所示，为悬挂点等高的架空线力学模型，在图（2.3）坐标系下可求

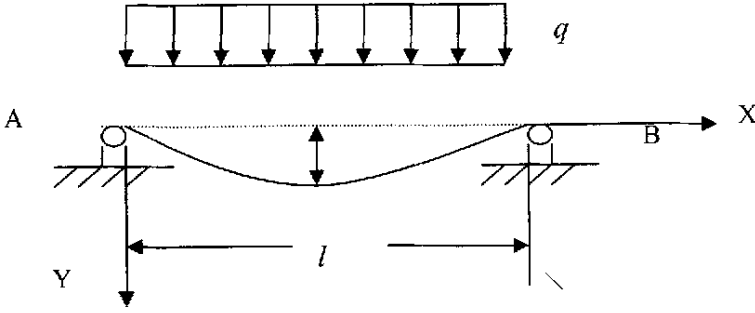


图 2.3 悬点等高的力学模型

Fig.2.3 Model at the same height

得：

任意点垂度：
$$y = \frac{4f}{l^2}(lx - x^2) \quad (2.1)$$

最大垂度：
$$f = \frac{ql^2}{8H} \quad (2.2)$$

式中：

q ——导线受到铅锤方向的载荷集度 (N/m)，包括自重、覆冰、风载等。

(2) 斜抛物线法

在较陡的山坡上架线或落差很大的山谷与山顶间架线，如果仍用上面所说的平抛物线法计算垂度，结果与实际数值会相差很大，此时可采用斜抛物线法。此法的条件是 $0.1 < \frac{h}{l} \leq 0.25$ ，它是假定二悬挂点已存在高差 h ，此时线上垂直载荷视为沿斜档（即而悬挂点间的空间直

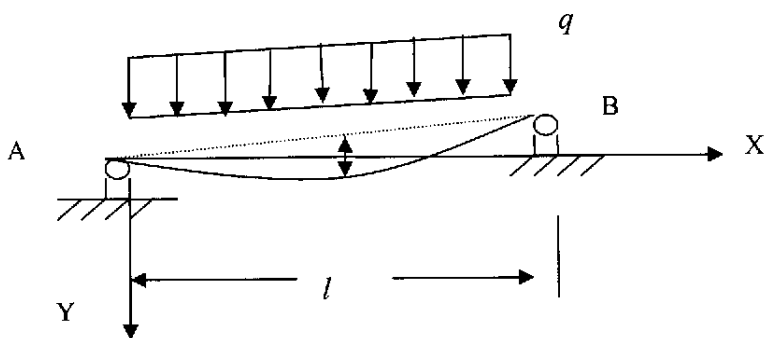


图 2.4 悬点不等高的力学模型
Fig.2.4 Model at the different height

线距离) 均匀分布。如图 (2.4) 所示, 实际档距为 l , 斜档距 $l' = AB = \frac{l}{\cos \varphi}$ 载
荷 q 沿 AB 均匀分布, 折算到 X 轴上均匀分布则为 $q' = q \cos \varphi$, 再利用前面提过
的抛物线法所得到的结果, 则可以得:

任意点垂度:

$$y = \frac{4f'}{l^2}(lx - x^2) = \frac{q}{2H \cos \varphi}(lx - x^2) \quad (2.3)$$

档距中点垂度:

$$f' = \frac{q'l'^2}{8H} = \frac{q \cos \varphi}{8H} \left(\frac{l}{\cos \varphi}\right)^2 = \frac{ql^2}{8H \cos \varphi} \quad (2.4)$$

式中:

q 、 l 、 H 、含义同前面平抛物线中规定。

φ ——斜档 l' 与水平轴 X 夹角 $\operatorname{tg} \varphi = \frac{h}{l}$ 。

(3) 悬链线计算方法

架空线的形状就像两端悬挂着的项链, 所以架空线可以按悬链线进行计算。悬链线的自重 (覆薄冰时也一样) 沿曲线长均匀分布设为 q , 坐标原点设在架空线的最低点, 并使 X 轴与悬链线相切, Y 轴与悬链线相切, Y 坐标恰与 AB 的中线重合, 如图 (2.5) 所示

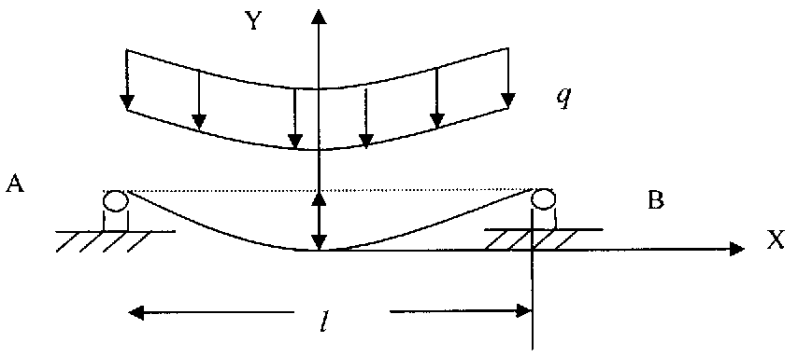


图 2.5 悬连线自重的力学模型

Fig.2.4 Model of the catenary

导线中张力的水平分量仍为 H 。由高等数学中的知识知悬链线方程：

$$y = \frac{H}{q} \left(\operatorname{ch} \frac{H}{q} x - 1 \right) \quad (2.5)$$

档距中点垂度的求法是令 $X = \frac{l}{2}$ 带入上式，即得：

$$f = \frac{H}{q} \left(\operatorname{ch} \frac{ql}{2H} - 1 \right)$$

我们知道公式 $\operatorname{ch} x = \frac{e^x + e^{-x}}{2}$ 又

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^n}{n!}$$

则有
$$\operatorname{ch} x = \frac{e^x + e^{-x}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{2n!}$$

因此将 f 展成级数形式，即

$$f = \frac{H}{q} \left[\left(1 + \frac{q^2 l^2}{8H^2} + \frac{q^4 l^4}{384H^4} + \cdots \right) - 1 \right] = \frac{ql^2}{8H^2} + \frac{q^3 l^4}{384H^3} + \cdots \quad (2.6)$$

$$\frac{ql^2}{8H}$$

从级数形式可见右边倒数第二项 $\frac{ql^2}{8H}$ ，就是用平抛物线求得的结果，由此可见，悬链线求垂度是精确的方法。

当两端悬挂点不等高时，即存在高差 h ， $\operatorname{tg}\varphi = \frac{h}{l}$ ，此时只需以

$q' = q \cos \varphi, l' = \frac{l}{\cos \varphi}$ ， H 不变带入 $f = \frac{H}{q'} \left(\operatorname{ch} \frac{q'l'}{2H} - 1 \right)$ 中得不等高档距中点垂度计算式

$$\begin{aligned} f_{\varphi} &= \frac{H}{q \cos \varphi} \left(\operatorname{ch} \frac{ql}{2H} - 1 \right) \\ &= \frac{ql^2}{8H \cos \varphi} + \frac{q^3 l^4}{384 H^3 \cos \varphi} + \dots \end{aligned} \quad (2.7)$$

(4) 总结

悬链线法未对载荷作任何假设，因此它是求垂度的最佳方法，同时悬链线法的垂度级数形式中可以包含平抛物线和斜抛物线的结果，并且悬链线法的使用范围适用架空线任何情况。所以本文计算有关架空线缆的垂度均采用悬链线法。

2.5 跨海桥梁和架空线缆通航能力的计算方法

2.5.1 跨海桥梁通过能力的计算

如图 (2.2) 所示，由于跨海桥梁的净空高度的确定不是在桥体底部的最高点而是在通航净宽的范围确定的，同时现在的跨海大桥大都采用斜拉或者是吊桥，桥梁底部各点的净空高度的变化不大，所以在里笔者不进行研究。

2.5.2 跨海架空线缆通过能力的计算

2.5.2.1 计算跨海架空线缆通过能力的意义

(1) 求架空线缆最低点的位置。受地理条件和其他原因限制，跨海线缆两端的高度有时候会不一样。有以上给出计算悬链线最低点的方法知，当线缆两端的高度不相等时，其最低点不在水平线中点的位置，但是在确定架空线缆的通航净空高数值时要知道线缆的最低点位置，所以要进行最低点位置计算。

(2) 计算线缆的各点的通过能力。因为有时候跨海架空线缆的跨度非常大，这样计算各点的通航净空高度就会很有意义，将各个点的通过能力算出来，标到海图上。

2.5.2.2 架空线缆通过能力的计算方法

在此，以本人参加项目《220KV 舟山与大陆联网工程输电线路跨越螺头水道通航安全评估》时计算架空线缆的通过能力为例，将架空线缆的通过能力计算方法介绍一下。

(1) 项目工程简介

舟山与大陆联网跨海工程大跨越工程采用 220KV 架空线路将舟山岛和大陆连接起来，采用的路径为：从舟山本竹岛经盘峙岛、雌山、小猫岛、大猫岛、凉帽山、外峙岛进入北仑区白蜂镇。该路经有三处跨越，分别位于舟山本岛—盘峙岛、盘峙岛—和尚山—雌山、大猫岛—凉帽山—外峙岛。路径全长 20.13km，其中三处跨越的情况如下表

表 2.1 工程项目跨越一览表

Table 2.1 The span of the project

序号	跨越名称	耐张段长 (m)	跨越档距 (m)	备注
1	盘峙港	2260	1700	小山
2	和尚山	1230	790	小山
3	大猫山	6250	2750/1235	山/小山/海下平台

其中大猫山跨越档距达 2750 米，跨越整个螺头水道，是目前国内最大的跨越档距，本次计算以此为例。工程地理位置示意图如下：

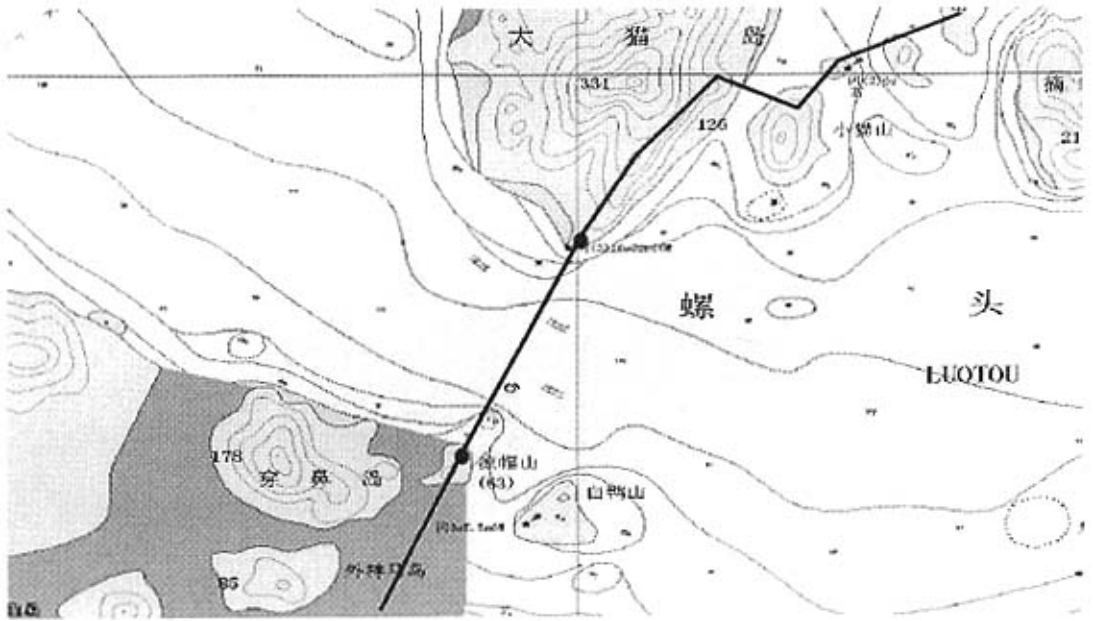


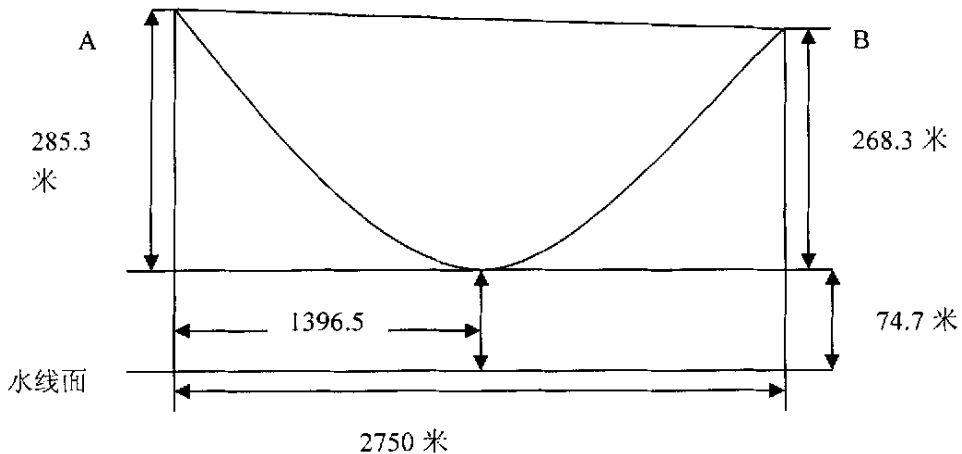
图 2.6 工程地理位置示意图

Fig. 2.6 Geographic position of the project

工程尺度参数数值如下：

$$\frac{H}{q}$$

跨度档距 l 为 2750 米。 q 为 3588，一端距水线面的垂直距离为 275.3 米，另一端距水线面的垂直距离为 258.3 米。工程尺度示意图如下：



A 点比 B 点高出 17 米

图 2.7 架空线缆尺度示意图

Fig.2.2 Sketch map of clearance height of bridges and wires over strait

(2) 数据计算

由上面给出的不等高档距中点垂度计算式

公式:

$$f_{\varphi} = \frac{H}{q \cos \varphi} \left(\operatorname{ch} \frac{ql}{2H} - 1 \right)$$

$$= \frac{ql^2}{8H \cos \varphi} + \frac{q^3 l^4}{384H^3 \cos \varphi} + \dots \quad (2.8)$$

令 $A = \frac{q}{H}$ 则公式 (3.8) 可变为 $f_{\varphi} = \frac{A}{\cos \varphi} \left(\operatorname{ch} \frac{l}{2A} - 1 \right)$

又知: $\operatorname{tg} \varphi = \frac{17}{2750} = 0.006$ 可知 $\varphi \rightarrow 0$ 即可将公式近似的变为

$$f_{\varphi} = A \left(\operatorname{ch} \frac{l}{2A} - 1 \right)$$

又知由知 $A=3588$, 算出最低点的垂度为 $f=266.8$

以最低点为原点激励如下图所示的坐标系, 则该悬链线公式为

$$f_x = A \operatorname{ch} \left(\frac{x}{A} \right) - A, \text{ 则有}$$

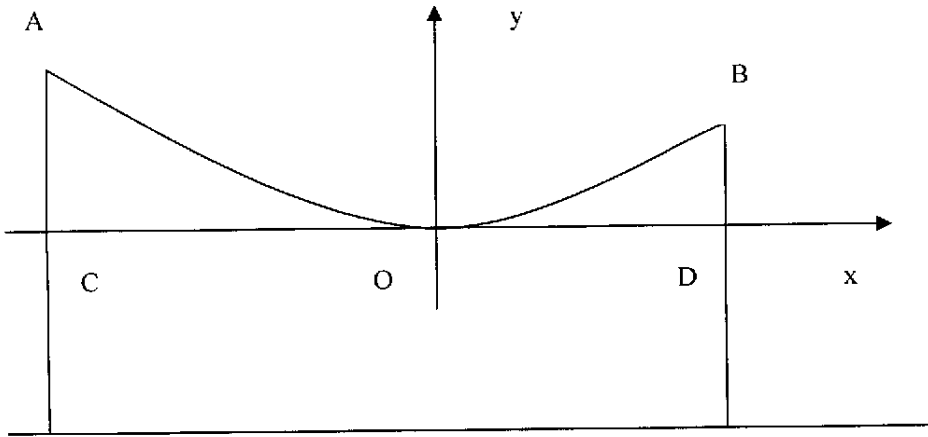


图 2.8 悬链线方程坐标示意图
Fig.2.2 Sketch map of wires over strait

$$f_A - f_B = 17$$

$$OA + OB = 2750$$

解上面两个方程组成的方程组得

$$OA = 1396.5 \text{米}$$

$$OB = 1353.5 \text{米}$$

根据上面的计算方法，利用悬链线计算垂度的方法，对架空线缆在航道上各段距线缆最低点的垂直距离进行计算，再加上线缆最低点距水线面的垂直距离，即得架空线缆各点距水线面的垂直距离（由于船舶在航行时将占用一定的宽度，所以在计算时选取每相距 100 米为一个取值点），其计算结果如下：

表 2.1 架空线缆各点距水线面的垂直距离

Tab.2.1 Upright height of the wires over strait at different position

距最低点距离 (m)	0	100	200	300	400	500	600
线缆各点距水面的距离 (m)	74.7	76.2	83.5	87.2	97	109.6	125
距最低点距离 (m)	700	800	900	1000	1100	1200	1300
线缆各点距水面的距离 (m)	143.2	164.2	188.1	214.9	244.6	277.2	312.7

又知该航道的深水槽距岸边 200 米，由以上结果可以画出该架空线缆的水面上通航高度分布图如下：

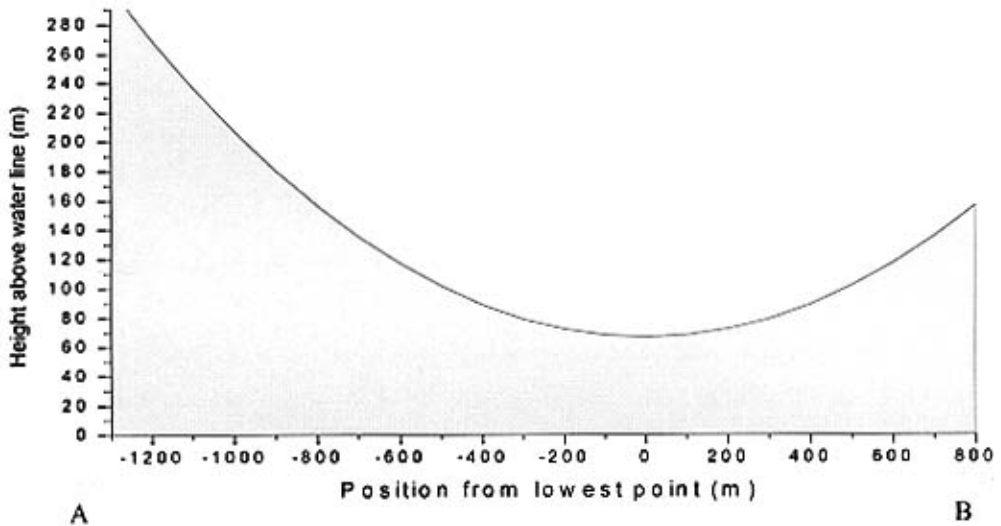


图 2.9 可通航的水面高度的分布图
Fig.2.9 Distributed height of the wire over strait

这样我们就可以很直观的看到，距离最低点水平距离一定的点的通航净空高度（即通过能力）。

第三章 回归分析的一般理论

3.1 回归分析概述

3.1.1 回归方程

回归分析是处理变量 x 与 y 之间关系的一种统计方法和技术。既当给定 x 的值, y 的值不能确定, 只能通过一定的概率分布来描述。于是, 我们称给定 x 时 y 的条件数学期望

$$f(x) = E(y/x) \quad (3.1)$$

为随机变量 y 对 x 的回归函数, 或称为随机变量 y 对 x 的均值回归函数。(3.1) 式从平均意义上刻画了变量 x 与 y 之间的统计规律。

在实际问题中, 我们把 x 称为自变量, y 称为因变量。如果要由 x 预测 y , 就是要利用 x , y 的观察值, 既样本观测值

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (3.2)$$

来建立一个公式, 当给定 x 值后, 就代入此公式算出一个 y 值, 这个值就称为 y 的预测值。如何建立这个公式, 这就要从样本观测值 (x_i, y_i) 出发, 观察 (x_i, y_i) 在平面直角坐标系上的分布情况。假设 (x_i, y_i) 在直角坐标系上的分布情况呈线性关系, 就可以考虑用一个线性函数来描述

$$y = \alpha + \beta x \quad (3.3)$$

方程 (3.3) 式中的参数 α, β 尚不知道, 这就需要由样本数据 (3.2) 式去进行估计。

当我们由样本数据 (3.2) 式估计出 α, β 的值后, 以估计值 $\hat{\alpha}, \hat{\beta}$ 分别代替 (3.3) 式中的 α, β 得方程

$$\hat{y} = \hat{\alpha} + \hat{\beta} x \quad (3.4)$$

(3.4) 式方程就称为回归方程。这里因为因变量 y 与自变量 x 的关系呈线性关系，故我们称 (3.4) 式为 y 对 x 的线性回归方程。又因 (3.4) 式的建立依赖于观察或试验积累的数据 (3.2) 式，所以我们又称 (3.4) 式为经验回归方程。相对这种叫法，我们把 (3.3) 式称为理论回归方程。理论回归方程是设想把所研究问题的总体中每一个体的 (x, y) 值都测量了，利用其全部结果而建立的回归方程 (3.3) 式，这在实际中是办不到的。理论回归方程中的 α 是方程 (3.3) 式所画出的直线在 y 轴上的截距， β 为直线的斜率，他们分别称为回归常数和回归系数。而方程 (3.4) 式中的参数 $\hat{\alpha}, \hat{\beta}$ 被称为经验回归常数和经验回归系数。

3.1.2 回归分析的主要内容及其一般模型

3.1.2.1 回归分析研究的主要内容

回归分析研究的主要对象是客观事物变量间的统计关系，它是建立在对客观事物进行大量试验和观察的基础上，用来寻找隐藏在那些看上去是不确定的现象中的统计规律性的统计方法。回归分析方法是通过对建立统计模型研究变量间相互关系的密切程度、结构状态、模型预测的一种有效的工具。

回归分析方法在生产实践中的广泛应用是它发展和完善的根本动力。如果从 19 世纪初 (1809 年) 高斯 (Gauss) 提出最小二乘法算起，回归分析的历史已有 190 多年。从经典的回归分析方法到近代的回归分析方法，它们所研究的内容已非常丰富。

3.1.2.2 回归模型的一般形式

如果变量 x_1, x_2, \dots, x_p 与随机变量 y 之间存在着相关关系，通常就意味着每当 x_1, x_2, \dots, x_p 取定值后， y 便有相应的概率分布与之对应。随机变量 y 与相关变量 x_1, x_2, \dots, x_p 之间的概率模型为

$$y = f(x_1, x_2, \dots, x_p) + \varepsilon \tag{3.5}$$

其中，随机变量 y 称为被解释变量 (因变量)； x_1, x_2, \dots, x_p 称为解释变量 (自变量)。

$y = f(x_1, x_2, \dots, x_p)$ 为一般变量 x_1, x_2, \dots, x_p 的确定性关系, ε 为随机误差。正是因为随机误差项 ε 的引入, 才将变量之间的关系描述为一个随机方程, 使得我们可以借助随机数学的方法研究 y 与 x_1, x_2, \dots, x_p 的关系。

模型 (3.5) 式清楚地表达了变量 x_1, x_2, \dots, x_p 与随机变量 y 的相关关系, 它由两部分组成: 一部分是确定性函数关系, 由回归函数 $f(x_1, x_2, \dots, x_p)$ 给出; 另一部分是随机误差项 ε 。由此可见模型 (3.5) 式准确地表达了相关关系那种既有联系又不确定的特点。

当概率模型 (3.5) 式中回归函数为线性函数时, 既有

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (3.6)$$

其中, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 为未知参数, 常称它们为回归系数。线性回归模型的“线性”是针对未知参数 $\beta_i (i = 0, 1, 2, \dots, p)$ 而言的。对于回归解释变量是非本质的, 因为解释变量是非线性时, 常可以通过变量的替换把它转化成线性的。

如果 $(x_{i1}, x_{i2}, \dots, x_{ip}; y_i), i = 1, 2, \dots, n$ 是 (3.6) 式中变量 $(x_1, x_2, \dots, x_p; y)$ 的一组观测值, 则线性回归模型可表示为

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, i = 1, 2, \dots, n \quad (3.7)$$

为了估计模型参数的需要, 古典线性回归模型通常应满足以下几个基本假设:

- (1) 解释变量 x_1, x_2, \dots, x_p 是非随机变量, 观测值 $x_{i1}, x_{i2}, \dots, x_{ip}$ 是常数。
- (2) 等方差及不相关的假定条件为

$$\begin{cases} E(\varepsilon_i) = 0, i = 1, 2, \dots, n \\ \text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \end{cases} (i, j = 1, 2, \dots, n) \end{cases}$$

这个条件称为高斯-马尔柯夫 (Gauss-Markov) 条件, 简称 G-M 条件。在此条件下, 便可以得到关于回归系数的最小二乘估计及误差项方差 σ^2 估计的一些重要性质, 如回归系数的最小二乘估计是回归系数的最小方差线性无偏估计等。

- (3) 正态分布的假定条件为

$$\begin{cases} \varepsilon_i \sim N(0, \sigma_2), i=1, 2, \dots, n \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \text{ 相互独立} \end{cases} \quad (3.8)$$

在此条件下便可得到关于回归系数的最小二乘估计及 σ^2 估计的进一步结果，如它们分别是回归系数及 σ^2 的最小方差无偏估计等，并且可以作回归的显著性检验及区间估计。

(4) 通常为了便于数学上的处理，还要求 $n > p$ ，既样本容量的个数多余解释变量的个数。在整个回归分析中，线性回归的统计模型最为重要。一方面因为线性回归的应用最广泛；另一方面只有在回归模型为线性的假定下，才能得到比较深入和一般的结果；再就是有许多非线性的回归模型可以通过适当的转化变为线性回归问题进行处理。

3.1.3 建立实际问题回归模型的过程

(1) 根据研究的目的，设置指标变量

回归分析模型主要是揭示事物间相关变量的数量关系。首先要根据所研究问题的目的设置因变量 y ，然后在选取与 y 有统计关系的一些变量作为自变量。

(2) 收集、整理统计数据

回归模型的建立是基于回归变量的样本统计数据。当确定好回归模型的变量之后，就要对这些变量收集整理统计数据。数据的收集是建立回归模型的重要一环，是一项基础性工作，样本数据的质量如何，对回归模型的水平有至关重要的影响。

(3) 确定理论回归模型的数学形式

当收集到所设置的变量的数据之后，就要确定适当的数学形式来描述这些变量之间的数学关系。绘制变量 y_i 与 x_i ($i=1, 2, \dots, n$) 的样本散点图是选择数学模型形式的重要一环。一般我们把 (x_i, y_i) 所对应的点在平面直角坐标系上画出来，看散点图的分布情况。如果这 n 个样本点大致分布在一条直线的周围，我们可考虑用线性回归模型去拟合这条直线，也既选择线性回归模型。如果 n 个样本点的分布大致在一条指数曲线的周围，我们就可选择指数形式的理论回归模型去描述它。

(4) 模型参数的估计

回归理论模型确定之后，利用收集、整理的样本数据对模型的未知参数给出估计是回归分析的重要内容。未知参数的估计方法最长用的是普通最小二乘法，它是经典的估计方法。对于不满足基本假设的回归问题，人们给出了种种新方法，如岭回归、主成分回归、偏最小二乘估计等，但它们都是以普通最小二乘法为基础。

(5) 模型的检验与修改

当模型的未知参数估计出来后，可以说初步建立了一个回归模型。我们建立回归模型的目的是为了应用它来研究问题，但如果马上就用这个模型去作预测、控制和分析，显然是不够慎重的。因为这个模型是否真正揭示了被解释变量与解释变量之间的关系，必须通过对模型的检验才能决定。对回归模型的检验通常包括对回归模型的显著性检验，以及回归系数的显著性检验，还有拟合优度的检验，随机误差项的序列相关检验，异方差性检验，解释变量的多重共线性检验等。

(6) 回归模型的运用

当一个模型通过了各种统计检验，且模型具有合理的意义时，我们就可以利用所求得的回归模型来进一步研究实际问题了。

3.2 一元线性回归

3.2.1 一元线性回归模型的数学形式

在回归模型中当只考虑两个变量 x , y 之间的关系时，描述 x 与 y 间的线性关系的数学结构式可看作是回归模型 (3.7) 式的特例，既当 (3.7) 式中 $p=1$ 时的情况。亦既

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (3.9)$$

(3.9) 式将实际问题中变量 y 与 x 之间的关系用两个部分描述。一部分是由于 x 的变化引起 y 线性变化的部分，既 $\beta_0 + \beta_1 x$ ；另一部分是由其他一切随机因素引起的，记为 ε 。(3.9) 式确切地表达了变量 x 与 y 之间密切相关，但密切程

度又没有到由 x 惟一确定 y 的地步的这种特殊关系。

(3.9) 式称为变量 y 对 x 的一元线性理论回归模型。一般我们称 y 为被解释变量 (因变量), x 为解释变量 (自变量)。式中 β_0 和 β_1 是未知参数, 称 β_0 回归常数, β_1 为回归系数。 ε 表示其他随机因素的影响。在 (3.9) 式中我们一般假定 ε 是不可观测的随机误差, 它是一个随机变量, 通常假定 ε 满足

$$\begin{cases} E(\varepsilon) = 0 \\ \text{var}(\varepsilon) = \sigma^2 \end{cases} \quad (3.10)$$

这里 $E(\varepsilon)$ 表示 ε 的数学期望, $\text{var}(\varepsilon)$ 表示 ε 的方差。对 (3.9) 式两端求期望, 得

$$E(y) = \beta_0 + \beta_1 x \quad (3.11)$$

称 (3.11) 式为回归方程。

一般情况下, 对我们所研究的某个实际问题, 获得的 n 组样本观测值 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 来说, 如果它们符合模型 (3.9) 式, 则

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.12)$$

由 (3.10) 式有

$$\begin{cases} E(\varepsilon_i) = 0 \\ \text{var}(\varepsilon_i) = \sigma^2 \end{cases} \quad i = 1, 2, \dots, n \quad (3.13)$$

通常我们还假定 n 组数据是独立观测的, 因而 y_1, y_2, \dots, y_n 与 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, 都是相互独立的随机变量。而 $x_i (i = 1, 2, \dots, n)$ 是确定性变量, 其值是可以精确测量和控制的。我们称 (3.12) 式为一元线性样本回归模型。

(3.9) 式的理论回归模型与 (3.12) 式的样本回归模型是等价的, 因而我们常不加区分地将两者统称为一元线性回归模型。

对 (3.12) 式两边分别求数学期望和方差, 得

$$E(y_i) = \beta_0 + \beta_1 x_i, \quad \text{var}(y_i) = \sigma^2, \quad i = 1, 2, \dots, n \quad (3.14)$$

(3.14)式表明随机变量 y_1, y_2, \dots, y_n 的期望不等, 方差相等, 因而 y_1, y_2, \dots, y_n 是独立的随机变量, 但并不同分布。而 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, 独立同分布的随机变量。

$E(y_i) = \beta_0 + \beta_1 x_i$ 从平均意义上表达了变量 y 与 x 的统计规律性。回归分析的主要任务就是通过 n 组样本观测值 (x_i, y_i) , $i=1, 2, \dots, n$, 对 $\hat{\beta}_0, \hat{\beta}_1$ 进行估计。一般用 $\hat{\beta}_0, \hat{\beta}_1$ 分别表示 β_0, β_1 的估计值, 则称

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3.15)$$

为 y 关于 x 的一元线性经验回归方程。

通常 $\hat{\beta}_0$ 表示经验回归直线在纵轴上的截距。如果模型范围里包括 $x=0$, 则 $\hat{\beta}_0$ 是 $x=0$ 时 y 概率分布的均值; 如果不包括 $x=0$, $\hat{\beta}_0$ 只是作为回归方程中的分开项, 没有别的具体意义。 $\hat{\beta}_1$ 表示经验回归方程的斜率, $\hat{\beta}_1$ 在实际应用中表示自变量 x 每增加一个单位时因变量 y 的平均增加数量。

在实际问题的研究中, 为了方便地对参数作区间估计和假设检验, 我们还假定模型 (3.9) 式中的误差项 ε 遵从正态分布, 既

$$\varepsilon \sim N(0, \sigma^2) \quad (3.16)$$

由于 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是 ε 的独立同分布的样本, 因而有

$$\varepsilon_i \sim N(0, \sigma^2) \quad i=1, 2, \dots, n \quad (3.17)$$

在 ε_i 遵从正态分布的假设下, 进一步有随机变量 y_i 也遵从正态分布

$$y_i \sim N(\beta_0 + \beta_1 x_i), i=1, 2, \dots, n \quad (3.18)$$

为了在今后的讨论中充分利用矩阵这个处理线性关系的有力工具, 我们这里将一元线性回归的一般形式 (3.9) 式用矩阵表示。令

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad (3.19)$$

于是模型 (3.9) 式表示为

$$\begin{cases} y = x\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{\varepsilon}) = 0 \\ \text{var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n \end{cases} \quad (3.20)$$

其中 I_n 为 n 阶单位矩阵。

3.2.2 参数 β_0, β_1 的估计

3.2.2.1 普通最小二乘估计

为了由样本数据得到回归参数 β_0 和 β_1 的理想估计值,我们将使用普通最小二乘估计(Ordinary Least Square Estimation, OLSE)。对每一个样本观测值 (x_i, y_i) , 最小二乘法观测值 y_i 与其回归值 $E(y_i) = \beta_0 + \beta_1 x_i$ 的离差越小越好, 综合地考虑 n 个离差值, 定义离差平方和为

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^n (y_i - E(y_i))^2 \\ &= \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned} \quad (3.21)$$

所谓最小二乘法, 就是寻找参数 β_0, β_1 的估计值 $\hat{\beta}_0, \hat{\beta}_1$, 使 (3.21) 使定义的离差平方和达到极小, 即寻找 $\hat{\beta}_0, \hat{\beta}_1$, 满足

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^n (y_i - E(y_i))^2 \\ &= \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned} \quad (3.22)$$

依照 (3.22) 式求出的 $\hat{\beta}_0, \hat{\beta}_1$, 就称为回归参数 β_0, β_1 的最小二乘估计。称

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (3.23)$$

为 $y_i (i = 1, 2, \dots, n)$ 的回归拟合值，简称回归值或拟合值。称

$$e_i = y_i - \hat{y}_i \quad (3.24)$$

为 $y_i (i = 1, 2, \dots, n)$ 的残差。

从几何关系上看，用一元线性回归方程拟合 n 个样本观测点 $(x_i, y_i), i = 1, 2, \dots, n$ ，就是要求回归直线 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 位于这 n 个样本点中间，或者说这 n 个样本点能最靠近这条回归直线。由图 3.1 可以直观地看到这种思想。

残差平方和

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (3.25)$$

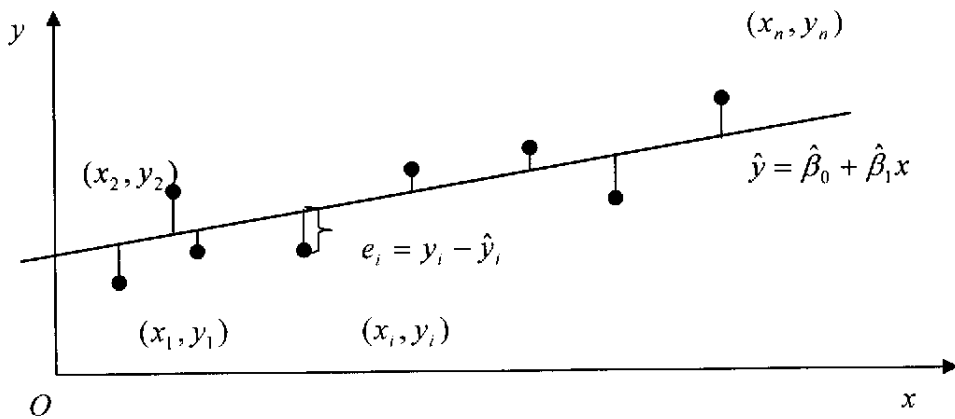


图 3.1 回归分析图

Fig.3.1 Map of regression analysis

从整体上刻画了 n 个样本观测点 $(x_i, y_i), i = 1, 2, \dots, n$ ，到回归直线 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 距离的大小。

从 (3.22) 式中求出 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是一个求极值问题。由于 Q 是关于 $\hat{\beta}_0, \hat{\beta}_1$ 的非负二次函数，因而它的最小值总是存在的。根据微积分中求极值的原理， $\hat{\beta}_0, \hat{\beta}_1$ 应满足下列方程组

$$\begin{cases} \left. \frac{\partial Q}{\partial \beta_0} \right|_{\beta_0 = \hat{\beta}_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \left. \frac{\partial Q}{\partial \beta_1} \right|_{\beta_1 = \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{cases} \quad (3.26)$$

经整理后，得正规方程组

$$\begin{aligned} n\hat{\beta}_0 + \left(\sum_{i=1}^n x_i\right)\hat{\beta}_1 &= \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i\right)\hat{\beta}_0 + \left(\sum_{i=1}^n x_i^2\right)\hat{\beta}_1 &= \sum_{i=1}^n x_i y_i \end{aligned} \quad (3.27)$$

求解以上正规方程组得 $\hat{\beta}_0, \hat{\beta}_1$ 的最小二乘估计 (OLSE) 为

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{cases} \quad (3.28)$$

其中

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

记

$$L_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n(\bar{x})^2 \quad (3.29)$$

$$L_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \quad (3.30)$$

则 (3.28) 式可简写为

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = L_{xy} / L_{xx} \end{cases} \quad (3.31)$$

易知， $\hat{\beta}_1$ 可以等价地表示为

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.32)$$

或

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \quad (3.33)$$

由 $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ 可知

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad (3.34)$$

可见回归直线 $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ 是通过点 (\bar{x}, \bar{y}) 的, 这对回归直线的作图很有帮助。

从物理学的角度看, (\bar{x}, \bar{y}) 是 n 个样本值 (x_i, y_i) 的重心, 也就是说回归直线通过样本的重心。

由 (3.24) 式可以得到残差的一个有用的性质

$$\begin{cases} \sum_{i=1}^n e_i = 0 \\ \sum_{i=1}^n x_i e_i = 0 \end{cases} \quad (3.35)$$

既残差的平均值为 0, 残差以自变量 x 的加权平均值为 0。

我们要确定的回归直线就是想使它与所有样本数据点都比较靠近, 为了刻画这种靠近程度, 人们曾设想用绝对残差和, 既

$$\sum_{i=1}^n |e_i| = \sum_{i=1}^n |y_i - \hat{y}_i| \quad \sum |e_i| \quad (3.36)$$

来度量观测值与回归直线的接近程度。显然, 绝对残差和越小, 回归直线就与所有数据点越近。然而, 绝对残差和 $\sum |e_i|$ 在数学处理上比较麻烦, 所以在经典的回归分析中, 都用残差平方和 (3.25) 式来描述因变量观测值 $y_i (i = 1, 2, \dots, n)$ 与回归直线的偏离程度。

3.2.2.2 最大似然估计

除了上述的最小二乘估计外，最大似然估计 (Maximum Likelihood Estimation, MLE) 方法也可以作为回归参数的估计方法。最大似然估计是利用总体的分布密度或概率分布的表达式及其样本所提供信息建立起来求未知参数估计量的一种方法。

最大似然估计的直观想法可用下面的例子说明：设有一事件 A，已知其发生的概率 p 只可能是 0.01 或 0.1。若在一次试验中事件 A 就发生了，自然应认为事件 A 发生的概率 p 是 0.1 而不是 0.01。把这种考虑问题的方法一般化就得到最大似然准则。

当总体 X 为连续型分布时，设其分布密度族为 $\{f(x, \theta), \theta \in \Theta\}$ ，假设总体 X 的一个独立同分布的样本为 x_1, x_2, \dots, x_n 。其似然函数为

$$L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) \quad (3.37)$$

最大似然估计应在一切 θ 中选取使随机样本 (X_1, X_2, \dots, X_n) 落在点 (x_1, x_2, \dots, x_n)

附近，概率最大的 $\hat{\theta}$ 为未知参数 θ 真值的估计值。既选取 $\hat{\theta}$ 满足

$$L(\hat{\theta}; x_1, x_2, \dots, x_n) = \max_{\theta} L(\theta; x_1, x_2, \dots, x_n) \quad (3.38)$$

对连续型随机变量，似然函数就是样本的联合分布密度函数，对离散型随机变量，似然函数就是样本的联合概率函数。似然函数的概念并不局限于独立同分布的样本，只要样本的联合密度的形式是已知的，就可以应用极大似然估计。

对于一元线性回归模型参数的最大似然估计，我们如果已经得到样本观测值 $(x_i, y_i), i=1, 2, \dots, n$ ，其中， x_i 为非随机变量； y_1, y_2, \dots, y_n 为随机样本。那么在假设 $\varepsilon_i \sim N(0, \sigma^2)$ 时，由 (3.18) 式可知 y_i 服从如小正态分布

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \quad (3.39)$$

y_i 的分布密度为

$$f_i(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} [y_i - (\beta_0 + \beta_1 x_i)]^2\right\}, i=1, 2, \dots, n \quad (3.40)$$

于是 y_1, y_2, \dots, y_n 的似然函数为

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n f_i(y_i) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \end{aligned} \quad (3.41)$$

由于 L 的极大化与 $\ln(L)$ 的极大化是等价的, 所以去对数似然函数为

$$\ln(L) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (3.42)$$

求 (3.42) 式的极大值, 等价于对 $\sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$ 求极小值, 到此又与最小

二乘原理相同。因而 $\hat{\beta}_0, \hat{\beta}_1$ 的极大似然估计就是 (3.28) 式的最小二乘估计。另

外, 由极大似然估计还可以得到 σ^2 的估计值为

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \end{aligned} \quad (3.43)$$

这个估计量是 σ^2 的有偏估计。在实际应用中, 常用无偏估计量

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \frac{1}{n-2} \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \end{aligned} \quad (3.44)$$

作为 σ^2 的估计量。

在此需要注意的是, 以上极大似然估计是在 $\varepsilon_i \sim N(0, \sigma^2)$ 的正态分布假设下求得的, 而最小二乘估计则对分布假设没有要求。另外, y_1, y_2, \dots, y_n 是独立的正态分布样本, 但并不是同分布的, 期望值 $E(y_i) = \beta_0 + \beta_1 x_i$ 不相等, 但这并不妨碍极大似然方法的应用。

3.2.3 最小二乘估计的性质

1. 线性

所谓线性就是估计量 $\hat{\beta}_0, \hat{\beta}_1$ 为随机变量 y_i 的线性函数。由 (3.32) 式得

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} y_i \quad (3.45)$$

其中, $\frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 是 y_i 的常数, 所以 $\hat{\beta}_1$ 是 y_i 的线性组合。同理可以证明 $\hat{\beta}_0$ 是 y_i 的

线性组合。

因为 y_i 为随机变量, 所以作为 y_i 的线性组合, $\hat{\beta}_0, \hat{\beta}_1$ 亦为随机变量, 因此各有其概率分布、均值、方差、标准差及两者的协方差。

2. 无偏性

下面我们讨论 $\hat{\beta}_0, \hat{\beta}_1$ 的无偏性。由于 x_i 是非随机变量, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$,

$E(\varepsilon_i) = 0$ 因而有

$$E(y_i) = \beta_0 + \beta_1 x_i \quad (3.46)$$

再由 (3.45) 式可得

$$\begin{aligned} E(\hat{\beta}_1) &= \sum_{i=1}^n \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_j - \bar{x})^2} E(y_i) \\ &= \sum_{i=1}^n \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_j - \bar{x})^2} (\beta_0 + \beta_1 x_i) = \beta_1 \quad (3.47) \end{aligned}$$

得 $\hat{\beta}_1$ 是 β_1 的无偏估计, 其中用到 $\sum (x_i - \bar{x}) = 0, \sum (x_i - \bar{x}) x_i = \sum (x_i - \bar{x})^2$ 。

同理可证 $\hat{\beta}_0$ 是 β_0 的无偏估计。

无偏估计的意义是, 如果屡次变更数据, 反复求 β_0, β_1 估计值, 这两个估计

量没有高估或低估的系统趋向，它们的平均值将趋于 β_0, β_1 。

进一步有

$$\begin{aligned} E(\hat{y}) &= E(\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= \beta_0 + \beta_1 x_i \\ &= E(y) \end{aligned} \quad (3.48)$$

这表明回归值 \hat{y} 是 $E(y)$ 的无偏估计，也说明 \hat{y} 与真实值 y 的平均值是相同的。

3. $\hat{\beta}_0, \hat{\beta}_1$ 的方差

一个估计量是无偏的，只揭示了估计量优良性的一个方面。我们通常还关心估计量本身的波动状况，这就需进一步研究它的方差。

由 y_1, y_2, \dots, y_n 相互独立 $\text{var}(y_i) = \sigma^2$ 及 (3.45) 式，得

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= \sum_{i=1}^n \left[\frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} \right]^2 \text{var}(y_i) \\ &= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \end{aligned} \quad (3.49)$$

我们知道，方差的大小表示随机变量取值波动的大小，因而 $\text{var}(\hat{\beta}_1)$ 反映了估计量 $\hat{\beta}_1$ 的波动大小。假设我们反复抽取容量为 n 的样本建立回归方程，每次计算的 $\hat{\beta}_1$ 的值是不相同的， $\text{var}(\hat{\beta}_1)$ 正是反映了这些 $\hat{\beta}_1$ 的差异程度。

由 $\text{var}(\hat{\beta}_1)$ 的表达式我们能得到对实际应用有指导意义的思想。从 (3.49) 式中看到，回归系数 $\hat{\beta}_1$ 不仅与随机误差的方差 σ^2 有关，而且还与自变量 x 的取值波动程度有关。如果 x 的取值比较分散，即 x 的波动较大，则 $\hat{\beta}_1$ 的波动就小， β_1 的估计值 $\hat{\beta}_1$ 就比较稳定。反之，如果原始数据 x 是在一个较小的范围内取值，则 β_1 的估计值稳定性就差，当然也就很难说精确了。这一点显然对我们收集原始数据

有重要的指导意义。类似地，有

$$\text{var}(\hat{\beta}_0) = \left[\frac{1}{n} + \frac{(\bar{x})^2}{\sum (x_i + \bar{x})^2} \right] \sigma^2 \quad (3.50)$$

由(3.50)式可知，回归常数 $\hat{\beta}_0$ 方差也不仅与随机误差的方差 σ^2 和自变量 x 的取值波动程度有关，而且还同样本数据的个数 n 有关。显然数据 n 越大时， $\text{var}(\hat{\beta}_0)$ 越小。

总之，由(3.49)式和(3.50)式我们可以看到，要想使 β_0, β_1 的估计值 $\hat{\beta}_0, \hat{\beta}_1$ 更稳定，在收集数据时，就应该考虑 x 的取值尽可能分散一些，不要挤在一块，样本量也应尽可能大一些，样本量 n 太小时估计量的稳定性肯定不会太好。

由前边 $\hat{\beta}_0, \hat{\beta}_1$ 线性的讨论我们知道 $\hat{\beta}_0, \hat{\beta}_1$ 都是 n 个独立正态随机变量 y_1, y_2, \dots, y_n 的线性组合，因而 $\hat{\beta}_0, \hat{\beta}_1$ 也遵从正态分布。由上边 $\hat{\beta}_0, \hat{\beta}_1$ 的均值和方差的结果，有

$$\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{(\bar{x})^2}{L_{xx}}\right)\sigma^2\right) \quad (3.51)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{L_{xx}}\right) \quad (3.52)$$

另外，还可得到 $\hat{\beta}_0, \hat{\beta}_1$ 的协方差

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}}{L_{xx}} \sigma^2 \quad (3.53)$$

(3.53)式说明，在 $\bar{x} = 0$ 时， $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 不相关，在正态假定下独立；在 $\bar{x} \neq 0$ 时不独立。它揭示了回归系数之间的关系状况。

在前边我们曾给出回归模型随机误差项 ε_i 等方差及不相关的假定条件，这个条件称为高斯-马尔柯夫(Gauss-Markov)条件，即

$$\begin{cases} E(\varepsilon_i) = 0, i = 1, 2, \dots, n \\ \text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \end{cases} (i, j = 1, 2, \dots, n) \end{cases} \quad (3.54)$$

在此条件下可以证明, $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 分别是 β_0 与 β_1 的最佳线性无偏估计 (Best Linear Unbiased Estimator, BLUE) 也称为最小方差线性无偏估计。BLUE 即指在 β_0 和 β_1 的一切线性无偏估计中, 它们的方差最小。

进一步知, 对固定的 x_0 来讲

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (3.55)$$

也是 y_1, y_2, \dots, y_n 的线性组合, 且

$$\hat{y}_0 \sim N\left(\beta_0 + \beta_1 x_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{L_{xx}}\right)\sigma^2\right) \quad (3.56)$$

由此可见, \hat{y}_0 是 $E(y_0)$ 的无偏估计, 且 \hat{y}_0 的方差随给定 x_0 的值与 \bar{x} 的距离 $|x_0 - \bar{x}|$ 的增大而增大。既当给定的 x_0 与 x 的样本平均值 \bar{x} 相差较大时, \hat{y}_0 的估计值波动就增大。这说明在实际应用回归方程进行控制和预测时, 给定的 x_0 值不能偏离样本平均值太大。

3.2.4 回归方程的显著性检验

当我们得到一个实际问题的经验回归方程 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 后, 还不能马上就用它去作分析和预测, 因为 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 是否真正描述了变量 x 与 y 之间的统计规律性, 还需运用统计方法对回归方程进行检验。在对回归方程进行检验时, 通常需要正态性假设 $\varepsilon_i \sim N(0, \sigma^2)$, 以下的检验内容若无特别声明, 都是在此正态性假设下进行的。下面我们介绍几种检验方法。

(1) t 检验

t 检验是统计推断中常用的一种检验方法。在回归分析中, t 检验用于检验回归系数的显著性。检验的原假设

$$H_0: \beta_1 = 0 \quad (3.57)$$

对立假设是

$$H_0: \beta_1 \neq 0 \quad (3.58)$$

回归系数的显著性检验就是要检验因变量 y 对自变量 x 的影响程度是否显著。如果原假设 H_0 成立，则因变量 y 与自变量 x 之间并没有真正的线性关系，也就是说自变量 x 的变化对因变量 y 并没有影响。由 (3.52) 式知， $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{L_{xx}})$ ，

因而当原假设 $H_0: \beta_1 = 0$ 成立时有

$$\hat{\beta}_1 \sim N(0, \frac{\sigma^2}{L_{xx}}) \quad (3.59)$$

此时 $\hat{\beta}_1$ 在零附近波动，构造 t 统计量

$$t = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/L_{xx}}} = \frac{\hat{\beta}_1 \sqrt{L_{xx}}}{\hat{\sigma}} \quad (3.60)$$

其中

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.61)$$

是 σ^2 的无偏估计，称 $\hat{\sigma}^2$ 为回归标准差。

当原假设 $H_0: \beta_1 = 0$ 成立时，(3.60) 式构造的 t 统计量服从自由度为 $n-2$ 的 t 分布。给定显著性水平 α ，双侧检验的临界值为 $t_{\alpha/2}$ 。当 $|t| \geq t_{\alpha/2}$ 时拒绝原假设 $H_0: \beta_1 = 0$ ，认为 β_1 显著不为零，因变量 y 对自变量 x 的一元线性回归成立；当 $|t| < t_{\alpha/2}$ 时接受原假设 $H_0: \beta_1 = 0$ ，认为 β_1 为零，因变量 y 对自变量 x 的一元线性回归不成立。

(2) F 检验

对线性回归方程显著性的另外一种检验是 F 检验，F 检验是根据平方和分解式，直接从回归效果检验回归方程的显著性。平方和分解式是

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.62)$$

其中, $\sum_{i=1}^n (y_i - \bar{y})^2$ 称为总平方和, 简记为 SST 或 $S_{\text{总}}$ 或 L_{yy} , SST 表示 Sum of Squares for Total。

$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 称为回归平方和, 简记为 SSR 或 $S_{\text{回}}$, R 表示 Regression。

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 称为残差平方和, 简记为 SSE 或 $S_{\text{残}}$, E 表示 Error。

因而平方和分解式可以简写为

$$\text{SST} = \text{SSR} + \text{SSE}$$

总平方和反映因变量 y 的波动程度或称不确定性, 在建立了 y 对 x 的线性回归后, 总平方和 SST, 就分解成回归平方和 SSR 与残差平方和 SSE 这两个组成部分, 其中 SSR 是由回归方程确定的, 也就是由自变量 x 的波动引起的, SSE 是不能用自变量解释的波动, 是由 x 之外的未加控制的因素引起的。这样, 总平方和 SSE 中, 能够由自变量解释的部分为 SSR, 不能由自变量解释的部分为 SSE。这样, 回归平方和 SSR 越大, 回归的效果就越好, 可以据此构造 F 检验统计量如下

$$F = \frac{\text{SSR}/1}{\text{SSE}/(n-2)} \quad (3.63)$$

在正态假设下, 当原假设 $H_0: \beta_1 = 0$ 成立时, F 服从自由度为 (1, $n-2$) 的 F 分布。当 F 值大于临界值 $F_{\alpha}(1, n-2)$ 时, 拒绝 H_0 , 说明回归方程显著, x 与 y 有显著的线性关系。

(3) 相关系数的显著性检验

由于一元线性回归方程讨论的是变量 x 与变量 y 之间的线性关系, 所以我们可以用变量 x 与 y 之间的相关系数来检验回归方程的显著性。设 $(x_i, y_i), i = 1, 2, \dots, n$, 是 (x, y) 的 n 组样本观测值, 我们称

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$= \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} \quad (3.64)$$

为 x 与 y 的简单相关系数，简称相关系数。其中， L_{xy} ， L_{xx} ， L_{yy} 与前边定义相同。

相关系数 r 表示 x 和 y 的线性关系的密切程度。相关系数的取值范围为 $|r| \leq 1$ 。

相关系数的直观意义如图 3.2 所示。

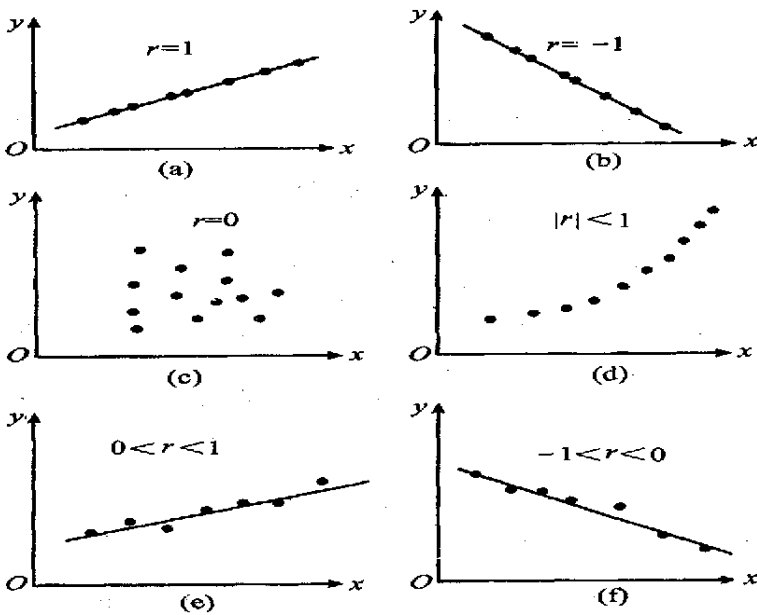


图 3.2 系数直观图

Fig.3.2 The sketch map of coefficient

图 3.2 中的 (a)、(b) 和 (c)、(d) 是四种极端情况，即当 x 与 y 有精确的线性关系时， $r=1$ 或 $r=-1$ 。 $r=1$ 表示 x 与 y 之间完全正相关，所有的对应点都在一条直线上； $r=-1$ 表示 x 与 y 之间完全负相关，对应点也都在一条直线上。这实际上就是一种确定的线性函数关系。图中 (c) 这种极端情况，说明所有的样本点分布杂乱无章，变量 x 与 y 之间没有相关关系，即 $r=0$ 。在实际中 $r=0$ 的情况

很少，往往我们拿来毫不相干的两个变量序列，计算相关系数绝对值都会大于零。图中 (d) 这种情况，表明 x 与 y 有确定的非线性函数关系，或称曲线函数关系。此时 $|r| < 1$ ，并不等于 1，这是因为简单相关系数只是反映两个变量间的线性关系，并不能反映变量间的非线性关系。因而，即使 $r = 0$ 也并不能说明 x 与 y 无任何关系。

当变量 x 与 y 之间有线性统计关系时， $0 < |r| < 1$ ，如图 3.2 中 (e)、(f) 所示。(e) 图表示 x 与 y 是正的线性相关，(f) 图表示 x 与 y 是负的线性相关。我们在实际问题中经常碰到的是这两种情况。

由 (3.64) 式和回归系数 $\hat{\beta}_1$ 的表达式可得

$$r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \hat{\beta}_1 \sqrt{\frac{L_{xx}}{L_{yy}}} \quad (3.65)$$

由上式可以看到一个很有用的结论，即一元线性回归的回归系数 $\hat{\beta}_1$ 的符号与相关系数 r 的符号相同。

这里需要指出的是，相关系数有个明显的缺点，就是它接近于 1 的程度与数据组数 n 有关，这样容易给人一种假象。因为，当 n 较小时，相关系数的绝对值容易接近于 1；当 n 较大时，相关系数的绝对值容易偏小。特别是当 $n=2$ 时，相关系数的绝对值总为 1。因此在样本容量 n 较小时，我们仅凭相关系数较大就说变量 x 与 y 之间有密切的线性关系，就显得匆忙。

(4) 样本决定系数

由回归平方和与残差平方和的意义我们知道，如果在总的离差平方和中回归平方和所占的比重越大，则线性回归效果就越好，这说明回归直线与样本观测值拟合优度就越好；如果残差平方和所占的比重大，则回归直线与样本观测值拟合得就不理想。这里把回归平方和与总离差平方和之比定义为样本决定系数，记为 r^2 即

$$r^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.66)$$

由关系式

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.67)$$

可以证明 (3.66) 式的 r^2 正好是 (3.65) 式中相关系数 r 的平方。即

$$r^2 = \frac{SSR}{SST} = \frac{L^2_{xy}}{L_{xx}L_{yy}} = (r)^2 \quad (3.68)$$

决定系数 r^2 是一个回归直线与样本观测值拟合优度的相对指标，反映了因变量的波动中能用自变量解释的比例。 r^2 的值总是在 0 和 1 之间，也可以用百分数表示。一个线性回归模型如果充分利用了 x 的信息，因变量不确定性的绝大部分能由回归方程解释，则 r^2 越接近于 1，拟合优度就越好。反之，如 r^2 不大，说明从模型中给出的 x 对 y 的信息还不充分，回归方程的效果不好，应进行修改，使 x 与 y 的信息得到充分利用。

3.2.5 残差分析

一个线性回归方程通过了 t 检验或 F 检验，只是表明变量 x 与 y 之间的线性关系是显著的，或者说线性回归方程是有效的，但不能保证数据拟合得很好，也不能排除由于意外原因而导致的数据不完全可靠，比如有异常值出现、周期性因素干扰等。只有当与模型中的残差项有关的假定满足时，我们才能放心地运用回归模型。因此，在利用回归方程作分析和预测之前，应该用残差图帮助我们诊断回归效果与样本数据的质量，检查模型是否满足基本假定，以便对模型作进一步的修改。

(1) 残差概念与残差图

残差 $e_i = y_i - \hat{y}_i$ 的定义已由 (3.24) 式给出， n 对数据产生 n 个残差值。残差

e_i 是实际观测值 y 与通过回归方程给出的回归值之差，残差 e_i 可以看做是误差项 ε_i 的估计值。残差 $e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ ，误差项 $\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$ ，比较两个表达式可以正确区分残差 e_i 与误差项 ε_i 的异同。

以自变量 x 作横轴（或以因变量回归值 \hat{y} 作横轴），以残差作纵轴，将相应的残差点画在直角坐标系上，就可得到残差图，残差图可以帮助我们对数据质量作一些分析。图 3.3 给出常见的一些残差图，这些残差图各不相同，它们分别说明样本数据的不同表现情况。

一般认为，如果一个回归模型满足所给出的基本假定，所有残差应是在 $e=0$ 附近随机变化，并在变化幅度不大的一条带子内。见图 3.3 中 (a) 的情况。如果残差都落在变化幅度不大的一条带子内，也就可以说明回归模型满足基本假设。

(b) 的情况表明 y 的观测值的方差并不相同，而是随着 x 的增加而增加的。(c) 的情况表明 y 和 x 之间的关系并非线性关系，而是曲线关系这就需要考虑用另外的曲线方程去拟合样本观测值 y 。另外一种可能性是 y 存在自相关。(d) 的情形称为蛛网现象，表明 y 具有自相关。

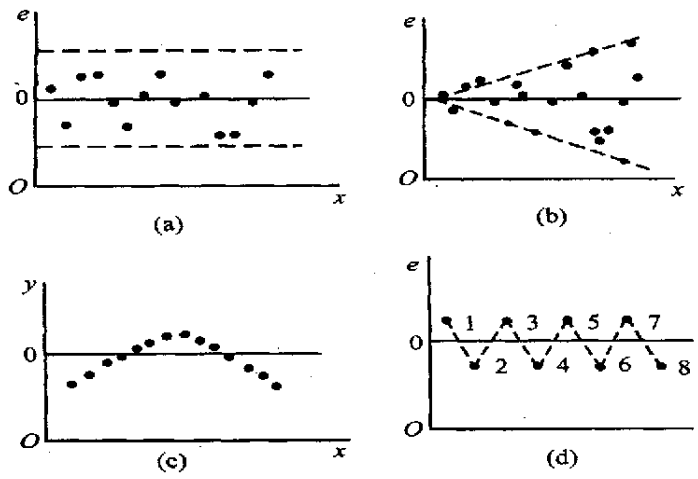


图 3.3 平方差直观图

Fig.3.3 The sketch map of square error

(2) 有关残差的性质

性质 1 $E(e_i) = 0$

证明: $E(e_i) = E(y_i) - E(\hat{y}_i)$

$$= (\beta_0 + \beta_1 x_i) - (\beta_0 + \beta_1 x_i) = 0$$

性质 2

$$\text{var}(e_i) = \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{L_{xx}} \right] \sigma^2$$
$$= (1 - h_{ii}) \sigma^2 \quad (3.69)$$

其中 $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{L_{xx}}$ 称为杠杆值。 $0 < h_{ii} < 1$ 当 x_i 靠近 \bar{x} 时, h_{ii} 的值接近 0,

相应的残差方差就大。当 x_i 远离 \bar{x} 时, h_{ii} 的值接近 1, 相应的残差方差就小。也就是说靠近 \bar{x} 附近的点相应的残差方差较大, 远离 \bar{x} 附近的点相应的残差方差较小, 这条性质可能令读者感到意外。实际上, 远离 \bar{x} 的点数目必然较少, 回归线容易“照顾”到这样的少数点, 使得回归线接近这些点, 因而远离 \bar{x} 附近的 x_i 相应的残差方差较小。

性质 3 残差满足约束条件: $\sum_{i=1}^n e_i = 0, \sum_{i=1}^n x_i e_i = 0$, 此关系式已在 (3.35) 式中给出。

这表明残差 e_1, e_2, \dots, e_n 是相关的, 不是独立的。

(3) 改进的残差

在残差分析中, 一般认为超过 $\pm 2\hat{\sigma}$ 或 $\pm 3\hat{\sigma}$ 的残差为异常值, 考虑到普通 e_1, e_2, \dots, e_n 的方差不等, 用 e_i 作判断和比较会带来一定的麻烦, 人们引入标准化残差和学生化残差的概念, 以改进普通残差的性质。分别定义如下:

标准化残差

$$ZRE_i = \frac{e_i}{\hat{\sigma}} \quad (3.70)$$

学生化残差

$$SRE_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}} \quad (3.71)$$

标准化残差使残差具有可比性， $|ZRE_i| > 3$ 的相应观测值即判定为异常值，这简化了判定工作，但是没有解决方差不等的问题。学生化残差则进一步解决了方差不等的问题，因而在寻找异常值时，用学生化残差优于用普通残差，认为 $|SRE_i| > 3$ 的相应观测值为异常值。学生化残差的构造公式类似于 t 检验公式，而 t 分布则是 Studen（学生）分布的简称，因而把（3.71）式称为学生化残差。

3.2.6 回归系数的区间估计

当我们用最小二乘法得到 β_0 和 β_1 的点估计后，在实际应用中往往还希望给出回归系数的估计精度，即给出其置信水平为 $1-\alpha$ 的置信区间。换句话说，就是分别给出以 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 为中心的一个区间，这个区间以 $1-\alpha$ 的概率包含参数 β_0 和 β_1 。置信区间的长度越短，说明估计值 $\hat{\beta}_0$ ， $\hat{\beta}_1$ 与 β_0 ， β_1 接近的程度越好，估计值就越精确；置信区间的长度越长，说明估计值 $\hat{\beta}_0$ ， $\hat{\beta}_1$ 与 β_0 ， β_1 接近的程度越差，估计值就越不精确。

在实际应用中，我们主要关心回归系数在 $\hat{\beta}_1$ 的精度，因而这里只推导 $\hat{\beta}_1$ 的置信区间。根据（3.52）式 $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{L_{xx}})$ 可得

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}_2/L_{xx}}} = \frac{(\hat{\beta}_1 - \beta_1)\sqrt{L_{xx}}}{\hat{\sigma}} \quad (3.72)$$

服从自由度为 $n-2$ 的 t 分布。因而

$$P\left[\left|\frac{(\hat{\beta}_1 - \beta_1)\sqrt{L_{xx}}}{\hat{\sigma}}\right| < t_{\alpha/2}(n-2)\right] = 1 - \alpha \quad (3.73)$$

上式等价于

$$P\left(\hat{\beta}_1 - t_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{L_{xx}}} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{L_{xx}}}\right) = 1 - \alpha \quad (3.74)$$

即得 β_1 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\hat{\beta}_1 - t_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{L_{xx}}}, \hat{\beta}_1 + t_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{L_{xx}}} \right) \quad (3.75)$$

3.2.7 预测和控制

建立回归模型的目的是为了应用，而预测和控制是回归模型最重要的应用。下面我们专门讨论回归模型在预测和控制方面的应用。

3.2.7.1 单值预测

单值预测就是用单个值作为因变量新值的预测值。比如我们研究某地区小麦亩产量 y 与施肥量 x 的关系时，在 n 块面积为一亩的地块上各施肥 x_i (kg)，最后测得相应的产量 y_i ，建立回归方程 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 。某农户在一亩地块上施肥 $x = x_0$ 时，该地块预期的小麦产量为

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

此即因变量新值 $y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0$ 的单值预测。这里预测目标 y_0 是一个随机变量，因而这个预测不能用普通的无偏性来衡量。根据 (3.48) 式 $E(\hat{y}_0) = E(y_0) = \beta_0 + \beta_1 x_0$ 说明预测值 \hat{y}_0 与目标值 y_0 有相同的均值。

3.2.7.2 区间预测

以上的单值预测 \hat{y}_0 只是这个地块小麦产量的大概值。仅知道这一点意义并不大，对于预测问题，除了知道预测值外，还希望知道预测的精度，这就需要作区间预测，也就是给出小麦产量的一个预测值范围。给一个预测值范围比只给出单个值 \hat{y}_0 更可信，这个问题也就是对于给定的显著性水平 α ，找一个区间 (T_1, T_2) ，使对应于某特定的 x_0 的实际值 y_0 以 $1 - \alpha$ 的概率被区间 (T_1, T_2) 所包含，用式子表示，就是

$$P(T_1 < y_0 < T_2) = 1 - \alpha \quad (3.76)$$

对因变量的区间预测又分为两种情况，一种是对因变量新值的区间预测，另一种是对因变量新值的平均值的区间预测。

(1) 因变量新值的区间预测

为了给出新值 y_0 的置信区间, 需要首先求出其估计值 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 的分布。

由于 β_0 与 β_1 都是 y_1, y_2, \dots, y_n 的线性组合, 因而 y_1, y_2, \dots, y_n 也是 y_1, y_2, \dots, y_n 的线性组合, 在正态假定下 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 服从正态分布, 其期望值为 $E(\hat{y}_0) = \beta_0 + \beta_1 x_0$, 以下计算其方差, 首先

$$\begin{aligned}\hat{y}_0 &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\ &= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0 \\ &= \sum_{i=1}^n \left[\frac{1}{n} + \frac{(x_i - \bar{x})(x_0 - \bar{x})}{L_{xx}} \right] y_i\end{aligned}\quad (3.77)$$

因而有

$$\begin{aligned}\text{var}(\hat{y}_0) &= \sum_{i=1}^n \left[\frac{1}{n} + \frac{(x_i - \bar{x})(x_0 - \bar{x})}{L_{xx}} \right]^2 \text{var}(y_i) \\ &= \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{L_{xx}} \right]^2\end{aligned}\quad (3.78)$$

从而得

$$\hat{y}_0 \sim N \left\{ \beta_0 + \beta_1 x_0, \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{L_{xx}} \right] \sigma^2 \right\}\quad (3.79)$$

记

$$h_{00} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{L_{xx}}\quad (3.80)$$

为新值 x_0 的杠杆值, 则上式简写为

$$\hat{y}_0 \sim N(\beta_0 + \beta_1 x_0, h_{00} \sigma^2)\quad (3.81)$$

y_0 是先前独立观测到的随机变量 y_1, y_2, \dots, y_n 的线性组合, 现在小麦产量的新值 y_0 与先前的观测值是独立的, 所以 y_0 与 \hat{y}_0 是独立的。因而

$$\text{var}(y_0 - \hat{y}_0) = \text{var}(y_0) + \text{var}(\hat{y}_0)$$

$$= \sigma^2 + h_{00}\sigma^2 \quad (3.82)$$

再由 (3.48) 式知 $E(y_0 - \hat{y}_0) = 0$ ，于是有

$$y_0 - \hat{y}_0 \sim N(0, (1 + h_{00})\sigma^2) \quad (3.83)$$

进而可知统计量

$$t = \frac{y_0 - \hat{y}_0}{\sqrt{1 + h_{00}}\hat{\sigma}} \sim t(n-2) \quad (3.84)$$

可得

$$P\left[\left|\frac{y_0 - \hat{y}_0}{\sqrt{1 + h_{00}}\hat{\sigma}}\right| \leq t_{\alpha/2}(n-2)\right] = 1 - \alpha \quad (3.85)$$

由此我们可以求得 y_0 的置信概率为 $1 - \alpha$ 的置信区间为

$$\hat{y}_0 \pm t_{\alpha/2}(n-2)\sqrt{1 + h_{00}}\hat{\sigma} \quad (3.86)$$

当样本容量 n 较大， $|x_0 - \bar{x}|$ 较小时， h_{00} 接近零， y_0 的置信度为 95% 的置信区间近似为

$$\hat{y}_0 \pm 2\hat{\sigma} \quad (3.87)$$

由 (3.86) 式可看到，对给定的显著性水平 α ，样本容量 n 越大，

$L_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ 越大， x_0 越靠近 \bar{x}_0 ，则置信区间长度越短，此时的预测精度就

高。所以，为了提高预测精度，样本量 n 应越大越好，采集数据 x_1, x_2, \dots, x_n 不能太集中。在进行预测时，所给定的 x_0 不能偏离 \bar{x} 太大，太大时，预测结果肯定不好；如果给定值 $x_0 = \bar{x}$ 时，置信区间长度最短，这时的预测结果最好。因此，如果在自变量观测值之外的范围作预测，精度就较差。这种情况进一步说明当 x 的取值发生较大变化时，即 $|x_0 - \bar{x}|$ 很大时，预测就不准。所以在作预测时一定要看 x_0 与 \bar{x} 相差多大，相差太大时，效果肯定不好。

(2) 因变量新值的平均值的区间估计

(3.86) 式给出的是因变量单个新值的置信区间，我们关心的另外一种情况

是因变量新值的平均值的区间估计。对于前面提出的小麦产量问题，如果该地区的一大片麦地每亩施肥量同为 x_0 ，那么这一大片地小麦的平均亩产如何估计呢？这个问题就是要估计平均值 $E(y_0)$ 根据 (3.48) 式， $E(y_0)$ 的点估计仍为 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ ，但是其区间估计却与因变量单个新值 y_0 的置信区间 (3.86) 式有所不同，由于 $E(y_0) = \beta_0 + \beta_1 x_0$ 是常数，由 (3.79) 式知

$$\hat{y}_0 - E(y_0) \sim N\left(0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{L_{xx}}\right) \sigma^2\right) \quad (3.88)$$

进而可得置信水平为 $1 - \alpha$ 的置信区间为

$$\hat{y}_0 \pm t_{\alpha/2}(n-2) \sqrt{h_{00}} \hat{\sigma} \quad (3.89)$$

3.3 可化为线性回归的非线性回归

实际问题中，有许多回归模型的被解释变量 y 与解释变量 x 之间的关系都不是线性的，其中一些回归模型通过对自变量或因变量的函数变换可以转化为线性关系，利用线性回归求解未知参数，并作回归诊断。如有下列模型

$$y = \beta_0 + \beta_1 e^x + \varepsilon \quad (3.90)$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon \quad (3.91)$$

$$y = a e^{bx} e^\varepsilon \quad (3.92)$$

$$y = a e^{b^x} + \varepsilon \quad (3.93)$$

对于 (3.90) 式，只需令 $x' = e^x$ 即可化为 y 对 x' 是线性的形式

$$y = \beta_0 + \beta_1 x' + \varepsilon$$

需要指出的是，新引进的自变量只能依赖于原始变量，而不能与未知参数有关。对于 (3.91) 式，我们可以令 $x_1 = x, x_2 = x^2, \cdots, x_p = x^p$ 于是得到 y 关于 x_1, x_2, \cdots, x_p 的线性表达式

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (3.94)$$

(3.91) 式本来只有一个自变量 x ，是一元 p 次多项式回归，在线性化后，变为 p 元线性回归。

对于 (3.92) 式，我们对等式两边同时取自然对数，得

$$\ln y = \ln a + bx + \varepsilon$$

令 $y' = \ln y, \beta_0 = \ln a, \beta_1 = b$ 于是得到 y' 关于 x 的一元线性回归模型

$$y' = \beta_0 + \beta_1 x + \varepsilon$$

对于 (3.93) 式，当 b 未知时，我们不能通过对等式两边同时取自然对数的方法将回归模型线性化，只能用非线性最小二乘方法求解。回归模型 (3.92) 式可以线性化，而回归模型 (3.93) 式不可以线性化，两个回归模型有相同的回归函数 ae^{bx} ，只是误差项 ε 的形式不同。(3.92) 式的误差项称为乘性误差项，(3.93) 式的误差项称为加性误差项。因而一个非线性回归模型是否可以线性化，不仅与回归函数的形式有关，而且与误差项的形式有关，误差项的形式还可以有其他多种形式。

在对非线性回归模型线性化时，总是假定误差项的形式就是能够使回归模型线性化的形式，为了方便，常常省去误差项，仅写出回归函数的形式。例如把回归模型 (3.92) 式简写为 $y = ae^{bx}$

3.4 几种实用的回归模型

3.4.1 线性回归模型

线性回归函数可表示为

$$g(x) = c_1 x + c_2 \quad (3.95)$$

其中 c_1, c_2 为未知系数。由于数据节点数大于 2，直线不可能经过每一个点，但是直线与数据的偏差一定要达到最小。直线与点的偏离程度（既残差）定义为：

$$r_i = y_i - g(x_i) = y_i - (c_1 x_i + c_2), i = 1, 2, \dots, L \quad (3.96)$$

其中 L 是总的节点数， c_1, c_2 是待定系数。

残差的平方和为：

$$R = \sum_{i=1}^L r_i^2 = \sum_{i=1}^L (\gamma_i - c_1 x_i - c_2)^2 \quad (3.97)$$

当R关于 c_1, c_2 的偏导数都为零，既：

$$\begin{cases} \frac{\partial R}{\partial c_1} = -2 \sum_{i=1}^L x_i (\gamma_i - c_1 x_i - c_2) = 0 \\ \frac{\partial R}{\partial c_2} = -2 \sum_{i=1}^L (\gamma_i - c_1 x_i - c_2) = 0 \end{cases} \quad (3.98)$$

时R达到最小。上式也可以写为：

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad (3.99)$$

其中：

$$\begin{aligned} a_{1,1} &= \sum_{i=1}^L x_i^2 \\ a_{1,2} &= a_{2,1} = \sum_{i=1}^L x_i \\ a_{2,2} &= \sum_{i=1}^L 1 = L \\ z_1 &= \sum_{i=1}^L x_i \gamma_i \\ z_2 &= \sum_{i=1}^L \gamma_i \end{aligned}$$

式 (3.99) 的解为：

$$\begin{aligned} c_1 &= (a_{2,2} z_1 - a_{1,2} z_2) / (a_{1,1} a_{2,2} - a_{1,2} a_{2,1}) \\ c_2 &= (a_{1,1} z_2 - a_{2,1} z_1) / (a_{1,1} a_{2,2} - a_{1,2} a_{2,1}) \end{aligned} \quad (3.100)$$

确定系数的另一种方法是考虑超定线性方程，如果每一个数据满足式 (3.95)，既：

$$c_1 x_i + c_2 = y_i, \quad i = 1, 2, \dots, L$$

或

$$Ac = y \quad (3.101)$$

由于方程个数L大于未知数个数，式 (3.101) 是一个不可解的超定线性方程。

为使方程可解，可将式 (3.101) 两端左乘A：

$$A'A_c = A'y \quad (3.102)$$

由于 $A'A$ 是一个 2×2 的方阵， $A'y$ 是一个长度为 2 的向量，上式就变成一个常规的 2×2 问题，其解为：

$$c = (A' * A) \setminus (A' * y)$$

与式 (3.99) 的解相等。

3.4.2 非线性曲线回归模型：幂函数回归模型

对某一类型的数据，通常做拟合函数为：

$$g(x) = \beta x^\alpha \quad (3.103)$$

其中 α, β 为待定的系数。为确定系数，首先将上式两边取对数得：

$$\log(g) = \alpha \log(x) + \log(\beta) \quad (3.104)$$

定义：

$$G = \log(g)$$

$$C_1 = \alpha$$

$$C_2 = \log(\beta)$$

$$X = \log(x)$$

式 (3.104) 变为：

$$G = c_1 X + c_2 \quad (3.105)$$

则问题简化为直线拟合问题，其拟合数据点为： $(\log(y_i), \log(x_i))$ 。式 (3.104) 也称为对数回归模型。

3.4.3 高次多项式曲线模型

最小二乘的思想可以推广到高次多项式拟合。 n 次多项式的形式为：

$$g(x) = c_1 x^n + c_2 x^{n-1} + \dots + c_{n+1} \quad (3.106)$$

曲线与数据点的残差为：

$$r_i = y_i - g(x_i), i = 1, 2, \dots, L \quad (3.107)$$

其中L是数据点数。残差的平方和为：

$$R = \sum_{i=1}^L r_i^2 \quad (3.108)$$

为使其最小化，可令R关于 c_j 的偏倒数为零，既：

$$\frac{\partial R}{\partial c_j} = 0, \quad j = 1, 2, \dots, n+1 \quad (3.109)$$

或

$$\sum_{j=1}^{n+1} \left(\sum_{i=1}^L x_i^{2n+2-j-k} \right) c_j = \sum_{i=1}^L x_i^{n+1-k} y_i, k = 1, 2, \dots, n+1 \quad (3.110)$$

另一种推导方法是将其写成超定方程：

$$Ac = y \quad (3.111)$$

其中：

$$A = \begin{bmatrix} x_1^n & x_1^{n-1} & \bullet & 1 \\ x_2^n & \bullet & \bullet & 1 \\ \bullet & \bullet & \bullet & \bullet \\ x_L^n & \bullet & \bullet & 1 \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \bullet \\ c_{n+1} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \bullet \\ y_L \end{bmatrix}$$

当 $L > n+1$ 时，方程就为超定的。因为方程个数大于未知数个数，所以不可解。

为使其可解，两边同时乘以 A' 得：

$$A'Ac = A'y \quad (3.112)$$

解得：

$$c = (A' * A) \setminus (A' * y)$$

第四章 回归模型的选择及其误差分析

4.1 回归模型的选择

本文使用两个重要的回归统计量来衡量模型的优良程度，并选择其中最优的一个作为我们的回归模型，它们就是 F 统计量^[12]和决定系数 r^2 ^[12]。

4.1.1 F 统计量函数

$$\begin{aligned} F &= \frac{SSR/1}{\hat{\sigma}^2} \\ &= \frac{(SST - SSE)/1}{\frac{SSE}{n-2}} \\ &= \frac{SSR/1}{SSE/(n-2)} \end{aligned}$$

其中：

SSR：表示回归平方和， $SSR = SST - SSE$

SST：表示 y_i 的校正平方和，既 $SST = \sum (y_i - \bar{y})^2$

SSE：表示回归残差平方和

在正态假设下，当原假设 $H_0: \beta_1 = 0$ 成立时，F 服从自由度为 $(1, n-2)$ 的 F 分布。当 F 值大于临界值 $F_\alpha(1, n-2)$ 时，拒绝 H_0 ，说明回归方程显著， x 与 y 有显著的线性关系。

4.1.2 决定系数 R^2

$$r^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

决定系数 r^2 是一个回归直线与样本观测值拟合优度的相对指标，反映了因变量的波动中能用自变量解释的比例。 r^2 的值总是在 0 和 1 之间，也可以用百分数表示。一个线性回归模型如果充分利用了 x 的信息，因变量不确定性的绝大部分

能由回归方程解释，则 r^2 越接近于 1，拟合优度就越好。反之，如 r^2 不大，说明从模型中给出的 x 对 y 的信息还不充分，回归方程的效果不好，应进行修改，使 x 与 y 的信息得到充分利用。

在实际的回归模型中，较大的 F 值和较大的 r^2 ，说明线性回归方程拟合的效果较好。因此，在模型选择中，我们选择模型的一个重要依据就是使模型的 F 统计量和决定系数 r^2 达到最大值。

4.2 误差分析

在误差分析中，我们采用平均相对误差 Z 作为回归模型的误差，下面我们具体分析一下平均相对误差 Z 的数学原理。

设回归模型为 $y = \beta_0 + \beta_1 x$ ，则对每一个观测值 (x_i, y_i) 有回归值 $\hat{y}_i = \beta_0 + \beta_1 x_i$ ，称 $e_i = y_i - \hat{y}_i$ 为 y_i 的残差。

设 $P_i = \frac{e_i}{y_i} \times 100\%$ 为残差与观测值 y_i 比值的百分数表示。 $P'_i = |P_i|$ ， S_p 为 P_i 的

标准差

$$S_p = \sqrt{\frac{\sum (P'_i - \bar{P})^2}{n-1}}$$

其中 $\bar{P} = \frac{\sum P'_i}{n}$ 为 P_i 的平均值。

定义变量 Z ，用它来描述观测值分布的情况，表示一个变量与该变量的平均值之差是标准差的多少倍，其计算公式为 $Z_i = \frac{X_i - \bar{X}}{S}$ 。其中 X_i 为该变量的观测

值； S 为标准差； \bar{X} 为该变量的平均数。则 $Z_{p_i} = \frac{P'_i - \bar{P}}{S_p}$ ，则 $Z = \frac{\sum Z_{p_i}}{n}$ 即为回归

模型 $y = \beta_0 + \beta_1 x$ 的平均相对误差。

平均相对误差 Z 反应了观测值 y_i 与回归值 \hat{y}_i 之间差的平均值

第五章数学模型的确定及数据处理结果

5.1 数据处理及回归模型确定

5.1.1 数据处理

为了确定船舶空载水线面以上至最高固定点高度随船舶载重吨的分布规律，本文收集了大量的实船基本数据，首先按照不同船型进行分类整理，依据一些基本的理论、方法和规范，利用相关软件对船舶数据做了统一处理。

(1) 对数据的分类、分组

由于近年来航运发展的特点，大型化的集装箱船、油船和散货船占据了航运的主要市场，本文将收集到的实船按船舶的种类进行分组，分为杂货船、散货船、油船、集装箱船四组进行数据处理。

(2) 奇异值的处理

为了提高回归方程的显著性水平确保其真实性，需要把影响力较大的奇异观测值删除掉。本论文收集的船舶数据由于录入的误差以及特殊船型的需要等，也会产生奇异值，因此有必要对其进行处理。因此本文利用了 SAS 计量分析法^[27]对奇异值进行了处理。

(3) SAS 计量分析法

① 奇异观测值的判别

K 元线性回归模型表示为

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon \quad (5.1)$$

为了判别 k 元线性回归模型公式 (5.1) 中独立变量 X_i 的奇异观测值，需要测量第 i 个观测值 $(X_i, X_{1i}, X_{2i}, \cdots, X_{ki})$ 与所有观测值的平均值 $\bar{X}(\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k)$ 间的距离 (离散程度)。为测量第 i 个观测值 X_i 与所有平均观测值 \bar{X} 间的离散程度可用奇异观测值检验统计量 h_{ii} ， h_{ii} 又称为观测值 X_i 的扭转力矩。如果 h_{ii} 越大，则第 i 个观测值 X_i 离 k 个独立变量观测值中心的距离越远，观测值 X_i 为奇异观测值的可能性越大。奇异观测值的检验统计量 h_{ii} 的性质和判别奇异观测值的方法如

下:

$$\sum_{i=1}^n = k+1 \quad 0 \leq h_{ii} \leq 1 \quad (5.2)$$

如果第 i 个观测值 X_i 的扭转力矩 $h_{ii} > \frac{2(k+1)}{n}$ (n 为观测值数), 则 X_i 为奇异观测值。

② 奇异值处理的 SAS 分析程序

```
PROC REG  
MODEL Y=X1-XK;  
OUTPUT OUT=TWO  
H=HAT RSTUDENT =SDR DFFITS=DF COOKD=D;  
RUN;
```

对船舶基本数据进行了以上处理后,基本上能够满足本论文的要求,即可对数据进行分析处理。

5.1.2 回归模型的确定

(1) 可供选择的回归模型

在回归模型的选择中有多种模型可供选择,其中比较常见的回归模型有以下几种:

① 线性回归模型 $Y = b_0 + b_1 t$ (5.3)

② 二次回归模型 $Y = b_0 + b_1 t + b_2 t^2$ (5.4)

③ 复合回归模型 $Y = b_0 \cdot b_1 t$ (5.5)

④ 对数回归模型 $Y = b_0 + b_1 \ln t$ (5.6)

⑤ 三次回归模型 $Y = b_0 + b_1 t + b_2 t^2 + b_3 t^3$ (5.7)

⑥ 幂函数回归模型 $Y = b_0 t^{b_1}$ (5.8)

⑦ 指数回归模型 $Y = b_0 e^{b_1 t}$ (5.9)

在以上回归模型中: t 表示自变量; b_0 表示常数项; b_n 表示回归系数; e 表示自然对数的底; \ln 表示以 e 为底的自然对数。

(2) 回归模型的选择

在选择回归模型过程中，我们采用第四章的回归模型选择方法。使用 F 统计量和决定系数 r^2 这两个重要的回归统计量来衡量模型的优良程度，并选择其中最优的一个作为我们的回归模型。利用相关软件（如 SPSS，Excel 等）对船舶数据分别按照以上 7 种回归模型进行处理。并比较其 F 值和 r^2 值，则 F 值和 r^2 值都比较大的模型就是我们要选择的模型。通过分析比较，我们选择了对数回归模型 (5.6)。

5.2 数据处理

5.2.1 散货船实船数据处理结果

船舶空载水线面以上最大高度 H_{\max} 与船舶载重吨 T 之间的关系式为：

$$H_{\max} = 7.169 \ln(T) - 34.093 \quad (5.10)$$

$$R^2 = 0.8644 \quad \text{标准误差 } \sigma \quad 2.8$$

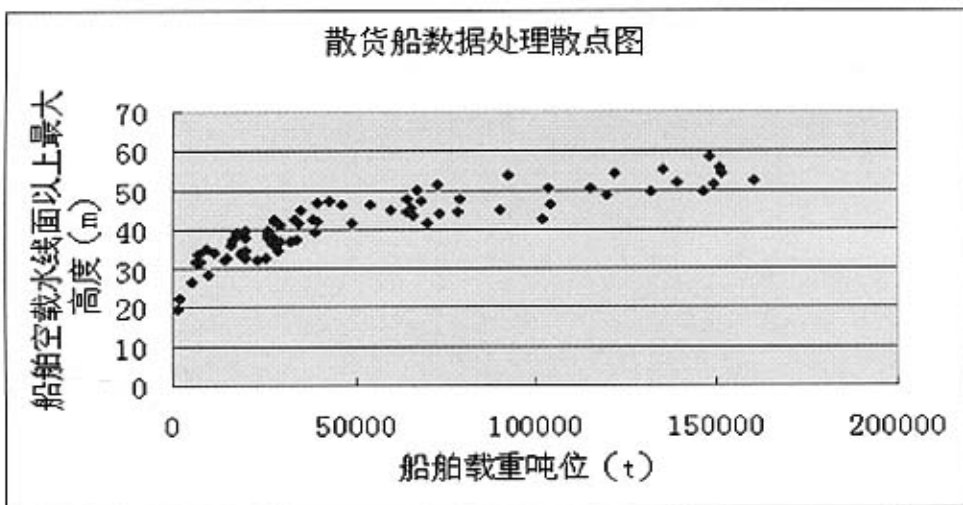


图 5.1 散货船数据处理散点图

Fig.5.1 Dots scattered of the worked data of the bulk carrier

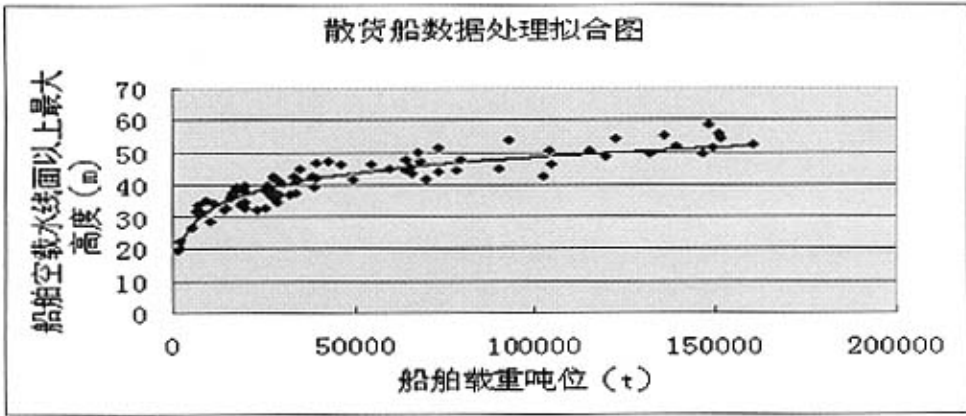


图 5.2 散货船数据处理拟合图

Fig.5.2 Regression curve of the worked data of the bulk carrier

表 5.1 散货船各吨位级船舶的 Hmax 实船数据平均值和回归公式预测值

Tab.5.1 The Hmax of the bulk carrier of different size

吨位级 (t)	5000	10000	30000	50000	100000	150000
Hmax (m)						
实船数据平均值	28.62	33.53	37.60	45.82	48.01	52.78
回归预测值	26.97	31.97	39.81	43.47	48.44	51.35
2σ 最大包容高度	32.57	36.57	45.41	49.07	54.04	56.95

5.2.2 杂货船实船数据处理结果

船舶空载水线面以上最大高度 H_{max} 与船舶载重吨之间的关系式为：

$$H_{max} = 8.004 \ln(T) - 37.895 \quad (5.11)$$

$$R^2 = 0.8922$$

$$\text{标准误差 } \sigma \quad 3.10$$

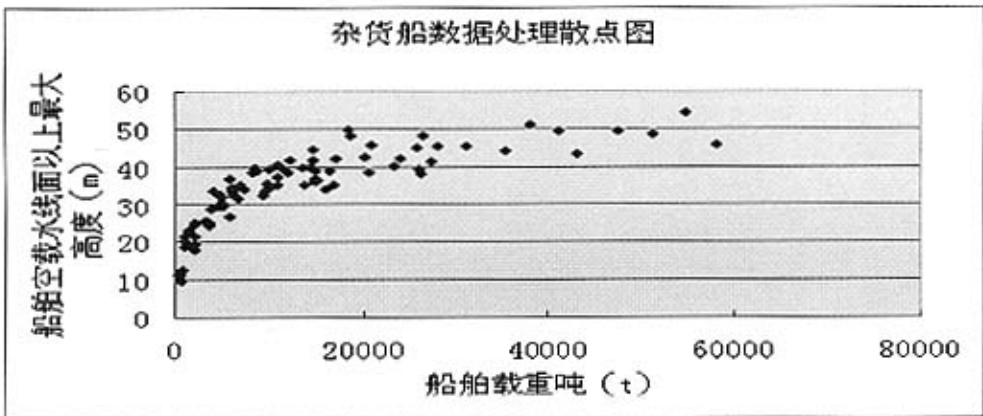


图 5.3 杂货船数据处理散点图

Fig.5.3 Dots scattered of the worked data of the dry-cargo carrier

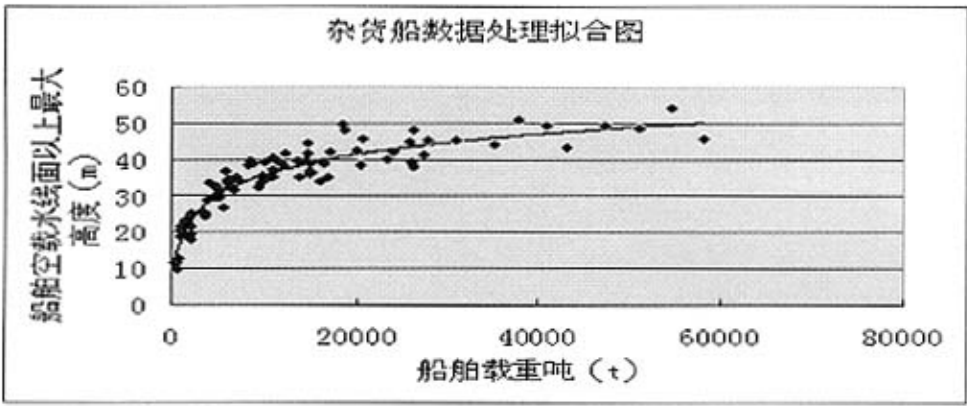


图 5.4 杂货船数据处理拟合图

Fig.5.4 Regression curve of the worked data of the dry-cargo carrier

表 5.2 杂货船各吨位级船舶的 Hmax 实船数据平均值和预测值

Tab.5.2 The Hmax of the dry-cargo carrier of different size

吨位级 (t) \ Hmax (m)	1000	5000	10000	30000	50000	60000
实船数据平均值	20.32	32.70	36.81	43.90	46.65	51.53
回归预测值	17.40	30.27	35.82	44.62	48.70	50.17
2σ 最大包容高度	23.60	26.47	42.02	50.82	54.90	56.37

5.2.3 油船实船数据处理结果

船舶空载水线面以上最大高度 H_{max} 与船舶载重吨之间的关系式为:

$$H_{max} = 6.431 \ln(T) - 27.018 \quad (5.12)$$

$$R^2 = 0.8751$$

$$\text{标准误差 } \sigma \ 3.21$$

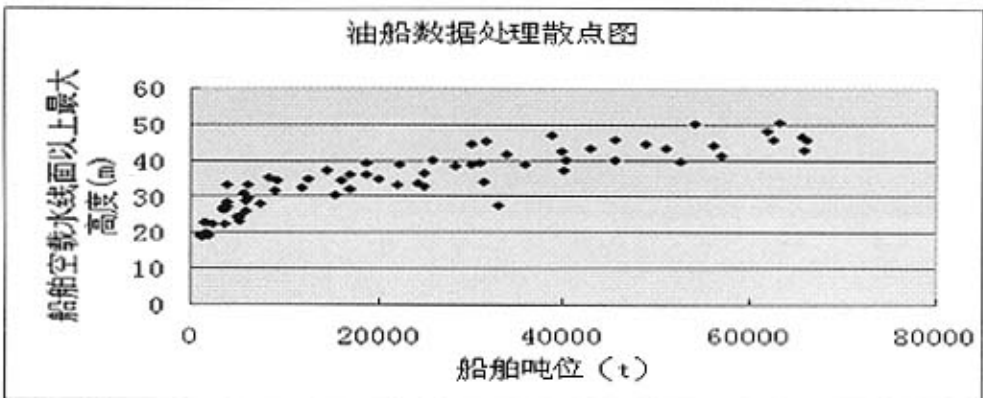


图 5.5 油船数据处理散点图

Fig.5.5 Dots scattered of the worked data of the tanker

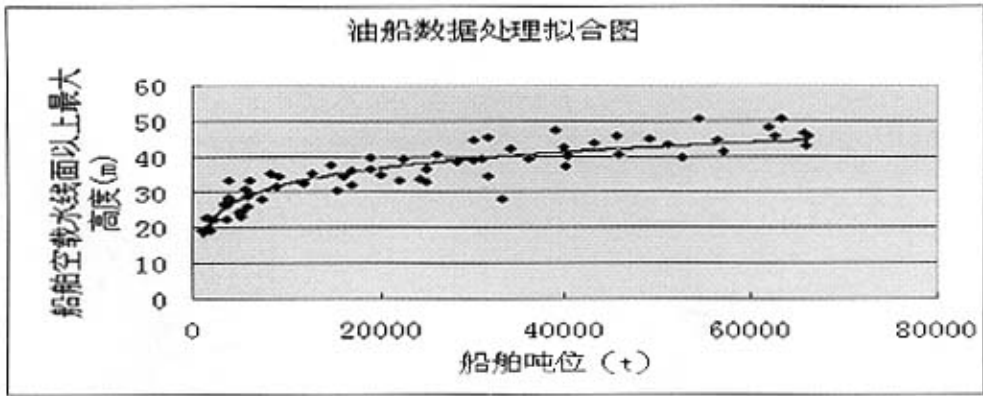


图 5.6 油船数据处理拟合图

Fig.5.6 Regression curve of the worked data of the tanker

表 5.3 油船各吨位级船舶的 Hmax 实船和预测值

Tab.5.3 The Hmax of the tanker of different size

吨位级 (t) \ Hmax (m)	5000	10000	30000	50000	70000
实船数据平均值	27.65	33.85	40.64	43.86	45.67
回归预测值	27.76	32.21	39.27	42.56	44.73
2σ 最大包容高度	24.16	38.61	45.27	48.96	51.33

5.2.4 集装箱船实船数据处理结果

船舶空载水线面以上最大高度 H_{max} 与船舶载重吨之间的关系式为:

$$H_{max} = 9.0687 \ln(T) - 46.571 \quad (5.13)$$

$$R^2 = 0.8922$$

$$\text{标准误差 } \sigma = 3.21$$

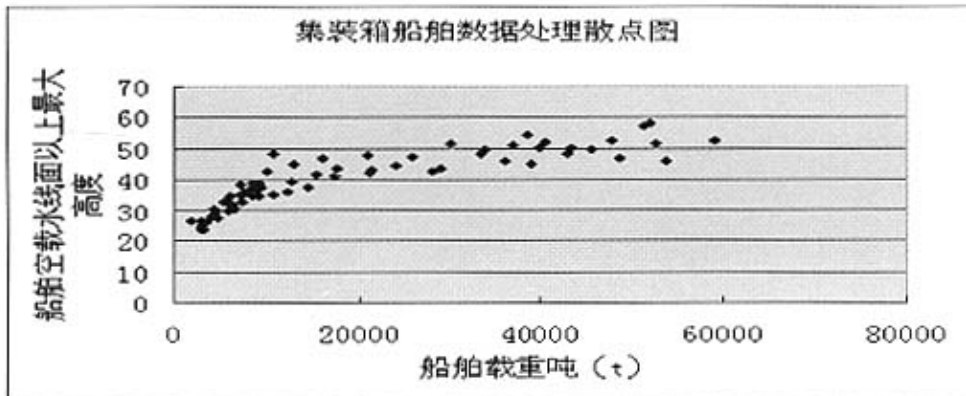


图 5.7 集装箱船舶数据处理散点图

Fig.5.7 Dots scattered of the worked data of the container

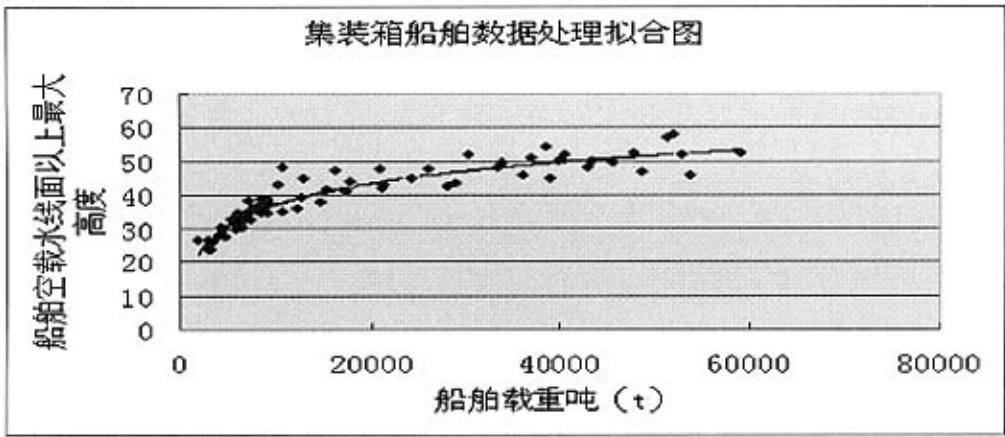


图 5.8 集装箱船舶数据拟合图

Fig.5.8 Regression curve of the worked data of the container

表 5.4 集装箱船各吨位级船舶的 Hmax 实船和预测值

Tab.5.4 The Hmax of the of container different size

吨位级 (t) \ Hmax (m)	5000	10000	30000	50000	60000
实船数据平均值	32.15	37.20	48.20	52.44	53.65
回归预测值	30.67	36.95	46.92	51.55	53.20
2σ 最大包容高度	36.71	42.99	52.96	57.59	59.24

5.2.5 所有种类的船舶综合数据处理结果

船舶空载水线面以上最大高度 H_{max} 与船舶载重吨之间的关系式为：

$$H_{max} = 6.8405 \ln(T) - 28.664 \quad (5.14)$$

$$R^2 = 0.8847$$

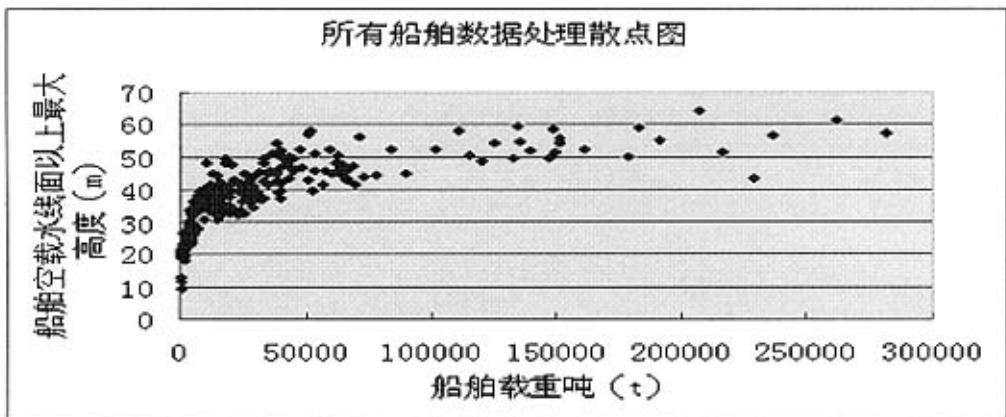


图 5.9 所有船舶数据处理散点图

Fig.5.9 Dots scattered of the worked data of all ships

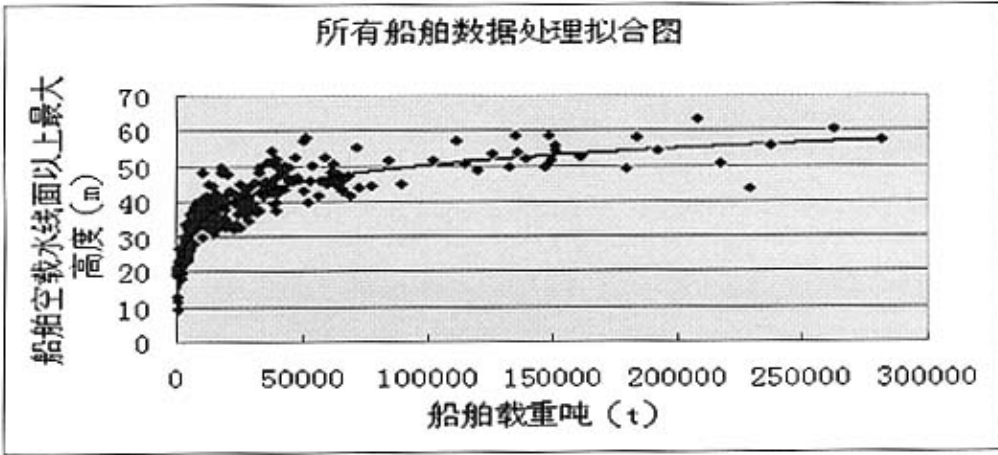


图 5.10 所有船舶数据处理拟合图

Fig.5.10 Regression curve of the worked data of all ships

以上各表内船舶吨级是按《海港总平面设计规范》(JTJ211—99)附录中设计船型尺度表中的分类方法来分的,所有的船舶没有统一的分类方法,这里不计算。

5.3 数据处理结果分析

由以上数据图表和计算结果可知船舶的空载水线面以上至最高固定点高度和船舶吨位之间存在着一定的数量关系,各个船型相应吨位级回归预测值与实船数据观测平均值较接近。由随机误差分布概律知船舶空载水线面以上至最高固定点的高度实船数据出现在 2σ 范围内的概率为 95.4%,实船数据平均值均包含在 2σ 包容最大高度内。由于我们在确定净空高度时考虑的是最大高度,所以我们可以将回归预测公式得到的预测值加上 2σ 作为设计跨海大桥和架空线缆通航净空高时的参考值。

第六章富裕高度的确定

6.1 影响富裕高度的因素

为了确保船舶能安全地通过桥梁和架空线缆，如何准确的计算富裕高度的大小，对跨航道桥梁和架空线缆净空高度的研究无疑也是十分重要的环节。安全而合理的富裕高度的数值无论对船舶的航行安全、大桥和架空线缆的经济价值都很重要。如果富裕高度的数值取得偏小，就会限制航道附近航运事业的发展，使大桥和架空线缆的经济效益降低。如果数值取得偏大，就会加大桥梁和架空线缆的造价，造成浪费，同时也会降低它们的使用年限。通过对《通航海轮桥梁通航标准》的研究和对相关因素的分析，笔者认为影响富裕高度的因素主要有以下几个方面（如下图所示）。

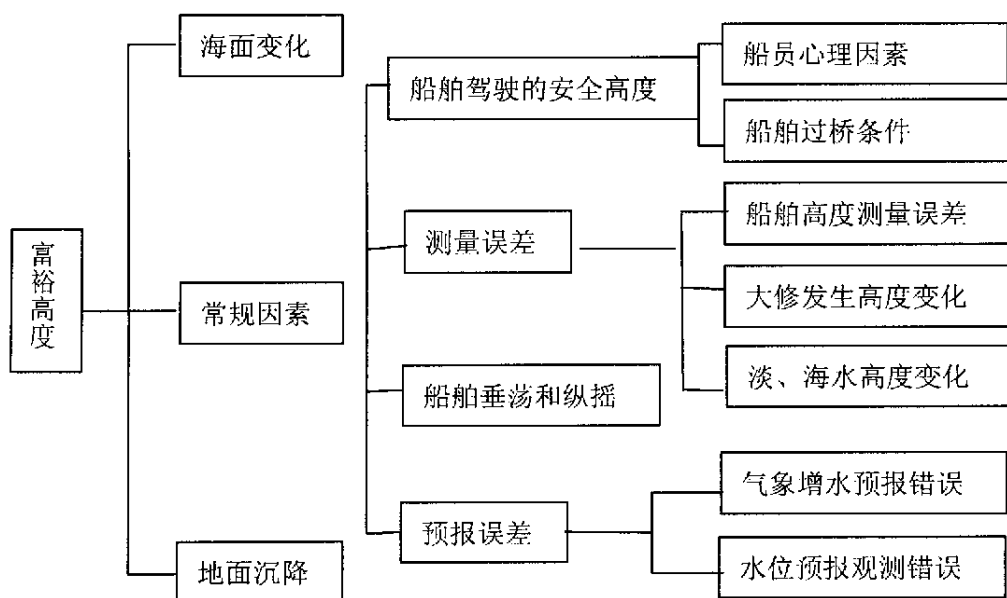


图 6.1 影响富裕高度的因素

Fig.6.1 Factors affecting on the safe height

6.2 对影响富裕高度相关因素的分析 and 计算

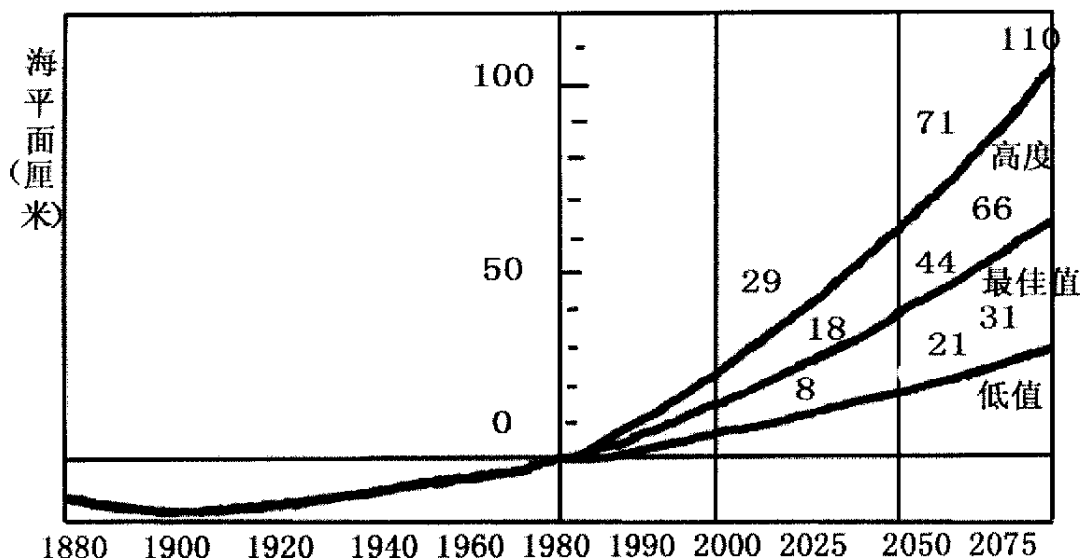
6.2.1 海平面上升引起的富裕高度变化量

(1) 全球海平面变化趋势

全球海平面变化是指全球平均海平面的升降值。引起海平面变化的主要因素

有以下四个方面。第一，大洋盆地容积的变化，主要是构造作用引起的。第二，大洋水体积的变化，主要是冰川的推进与收缩作用引起的。第三，大洋物质分布的变化引起的洋面的变化。第四，动力作用，如气象、水文、引力的变化引起的海面变化。5000 万年前由于印度、非洲板块与欧亚板块碰撞，陆地面积减小，洋面扩大，致使海洋平面下降了约 20 米。18000-6000 年前海平面呈持续上升趋势。6000 年以来全球海平面整体稳定，但局部有变化。近 100 年来，有实际的观测值作为全球海平面变化研究的依据。但是由于采用的统计方法不同，结果也不相同。根据联合国教科文组织发表的公告，全球海平面近百年的年平均上升值为 1.0-1.5mm/年。各个方面都对下世纪的全球海平面上升值做出了估计。都是过去 100 年的平均增长率的基础上考虑到下一世纪的全球变暖的趋势。全球变暖将使南极冰盖碎裂速度加快。格陵兰气候较南极为暖，故会逐渐减小冰盖。关于全球气温上升较前 100 年要高的根据主要是人类活动，特别是工业中产生的温室气体，其中 60% 为二氧化碳的排放量造成的。但是研究表明，人们所产生的 CO₂ 至少有三分之一被海洋吸收。海洋与大气之间的碳交换过程尚不清楚。有人认为，气温随大气中 CO₂ 含量的升高是有一定限度的。过此限度 CO₂ 再增加，温度不一定再上升。因此，南极冰盖的碎裂速度不会比现在快很多。同时，在过去 20 次的任何间冰期内南极的冰盖都未曾完全融化。因此，那种认为在未来 100 年内海平面会上升 1 米至几米的说法是不切实际的。下图是政府间气候变化委员会 (IPCC) 于 1992 年对下一世纪全球海平面上升值的估计曲线。其中的最佳值为，2030 年较现在上升 18 厘米 (cm)，2070 年为 44cm，2100 年为 66cm。1990-2030 年的年升率为 4.5mm/年，2031-2070 年为 6.5mm/年，2071-2100 年为 7.3mm/年。

历史海平面上升值

图 6.2 海平面变化趋势图^[2]Fig.6.2 Trendline of height of sea level^[2]

(2) 中国海平面变化趋势

全球的海平面变化量并不反映世界某一地的实际海平面的升降值。世界某一地点的海平面变化量等于全球海平面的上升（或下降）值加上当地陆地上升（或下降）值之和，这就是该地区的相对海平面变化。相对海平面变化对人类社会的影影响更有实际意义。世界一些大三角洲，包括长江和老黄河三角洲的地面沉降率均在 6-10mm/年，是目前全球海平面上升率的 10 倍以上。因此，研究某地的海平面变化必须包括海面 and 陆面的整体变化。

由于各种自然的和人为的因素不同，各处的相对海平面变化差异很大。我国的大地构造造成了海岸带的特点是三角洲皆在沉降带内。这些地区的相对海平面变化较大。而在非三角洲地区海平面上升率较小。我国沿海地区的地面沉降主要是由于过量抽取地下水所致，各处的沉降量也不相同。老黄河三角洲内的天津新港港区，根据 1985 年-1986 年 10 月的实测地面沉降图，全港区 80% 地区的地面沉降量在 20mm 左右。长江三角洲一些地区，1985-1990 年沉降量在 12-16mm。这是一些观测点上的沉降值，全区的平均沉降值要小于此数。关于我国沿海未来的相对海平面上升情况用任美锷院士给出的数据作为参考，如下表

表 6.1 我国沿海未来的相对海平面上升情况^[2]Tab.6.1 sea level of our coutray in the future^[2]

海域	2030 年海平面上升 估计值 (cm)			2050 年海平面上升 估计值(cm)		
	低	中等	高	低	中等	高
辽东半岛东部	12	19	31	22	37	59
辽宁西部和河北北部沿海	-3	4	16	-1	14	36
山东半岛东部~南坡沿海	4	11	23	10	25	47
江苏~长江口北部沿海	5	13	24	12	27	49
长江口沿海	23	30	42	38	53	75
浙江~广东东部沿海	-1	6	18	2	17	39
珠江口附近沿海	4	11	23	9	24	46
广东西部	3	10	22	9	24	46
平均值	6	13	25	13	28	50

目前一些跨越通航海轮航道的桥梁大多数是大型桥梁，使用期限一般都要超过 50 年，因此预测平均海平面上升幅度的年限不少于 50 年。从上面表格中可以看到我国沿海海平面的变化因地区的不同而差别很大，所以在考虑富裕高度的取值时应加以考虑，认真分析。

6.2.2 船舶垂荡值的计算

目前预报船舶在风浪中的船舶垂荡值(升沉)一般采用莫尔(Moore)的方法，莫尔法是莫尔基于模型试验结果回归的得到的。莫尔法具有推算方便，准确度高的优点。他的试验模型包括中速和快速的杂货船、油船和散货船等(它们的方形系数在 0.55~0.88 之间)，有很强的代表性。预报值计算公式如下：

$$(2Z_a)_{1/3} = A_0 + A_1 C_W + A_2 C_B + A_3 L/B + A_4 (L/d)^{-1} + A_5 L_{CB} + A_6 K_{yy} / L + A_7 V / L^{1/2}$$

式中：

C_W 、 C_B 、 L/B 、 L/d 分别代表船满载水线面系数、方形系数、长宽比和长度吃水比； L_{CB} 是浮心在横中剖面前的纵向位置，以船长的百分数表示出； K_{yy} 是实船的纵向质重惯性半径，也以船长的百分数表示； $V/L^{1/2}$ 是速长比，其中 V 用

节给出，而 L 的单位是 ft。

式中的各个系数 $A_0 \sim A_7$ 是随蒲氏风级、船长而变化的系数。可以用莫尔给出的 A_i 数值表中内插得到。

在计算船舶的垂荡值时首先要获取有关大桥和架空线缆所在水域的风浪高度，尤其应注意取得大桥和架空线缆下或其附近的风浪高度，根据波浪的特点以及历史资料，结合大桥和架空线的使用年限，进行综合分析，最终确定核算波高值。通常取 50 年一遇和 100 年一遇的大波值进行比较，由于大桥桥墩本身可能会对周围水域产生一定的影响，因此最终取值建议应考虑以下几个方面：

(1) 大桥下的海浪强度是否会大大降低。

(2) 运用近岸海浪破碎指标进行判断，大波是否会发生破碎。也就是说，如果特大波在还没有到达主通航孔之前已经破碎，波高、波能已大大衰减，则应考虑取主通航孔处的极限波高。

莫尔公式中要求的风浪用有效波高表示。以下是风级和有效波高对应表

表 6.2 风级和有效波高对应表

Tab.6.2 The valid height of wave according to different wind power

蒲氏风级		5	6	7	8
有效波高	Ft	9	13.9	18.9	24.5
	m	2.74	4.24	5.73	7.47

下面以一万吨级的船舶为例，计算出其在 5~8 级风中的垂荡值。莫尔系数根据莫尔给出的表值用内插法得到，列在表 6.3 中。垂荡值计算结果列在表 6.4 中。

表 6.3 莫尔法计算 1 万吨级船舶在蒲氏 5~8 级风中垂荡值的有关系数（内插值）

Tab.6.3 The coefficient given by Moore of ships at different wind power

蒲氏风级	浪高 m	船长 m	A0	A1	A2	A3	A4	A5	A6	A7
5	2.74	140	0.47	-12.84	4.02	0.145	70.6	0.204	20.9	3.02
5	2.74	120	-0.24	-13.34	4.67	0.096	79.3	0.205	23.7	3.90
6	4.24	140	0.91	-19.59	6.88	0.134	112.8	0.289	33.4	5.66
6	4.24	120	1.61	-18.06	6.49	0.062	107.6	0.260	31.3	5.50
7	5.73	140	1.09	-20.96	7.60	0.069	123.4	0.305	32.6	6.97
7	5.73	120	5.20	-23.60	8.42	0.113	137.7	0.353	38.92	7.49

8	7.47	140	8.01	-23.99	8.39	0.059	134.1	0.412	33.03	7.90
8	7.47	120	11.02	-26.75	9.20	0.090	151.8	0.429	28.44	8.71

表 6.4 莫尔法计算 1 万吨级船舶在蒲氏 5~8 级风中垂荡值

Tab.6.4 The change of ship's draft at different wind power

浪高 m	船长 m	C _w	C _b	L/B	L/d	L _{CB}	K _{YY} /L	V/L ^{1/2}	垂荡 值 m
2.74	140	0.80	0.70	7.0	18	0.5	0.25	0.37	0.67
2.74	120	0.82	0.75	6.5	16	0.5	0.25	0.40	0.83
4.24	140	0.80	0.70	7.0	18	0.5	0.25	0.37	1.16
4.24	120	0.82	0.75	6.5	16	0.5	0.25	0.40	1.38
5.73	140	0.80	0.70	7.0	18	0.5	0.25	0.37	1.26
5.73	120	0.82	0.75	6.5	16	0.5	0.25	0.40	4.19
7.47	140	0.80	0.70	7.0	18	0.5	0.25	0.37	4.25
7.47	120	0.82	0.75	6.5	16	0.5	0.25	0.40	5.14

万吨级船舶在不同浪高中的垂荡值

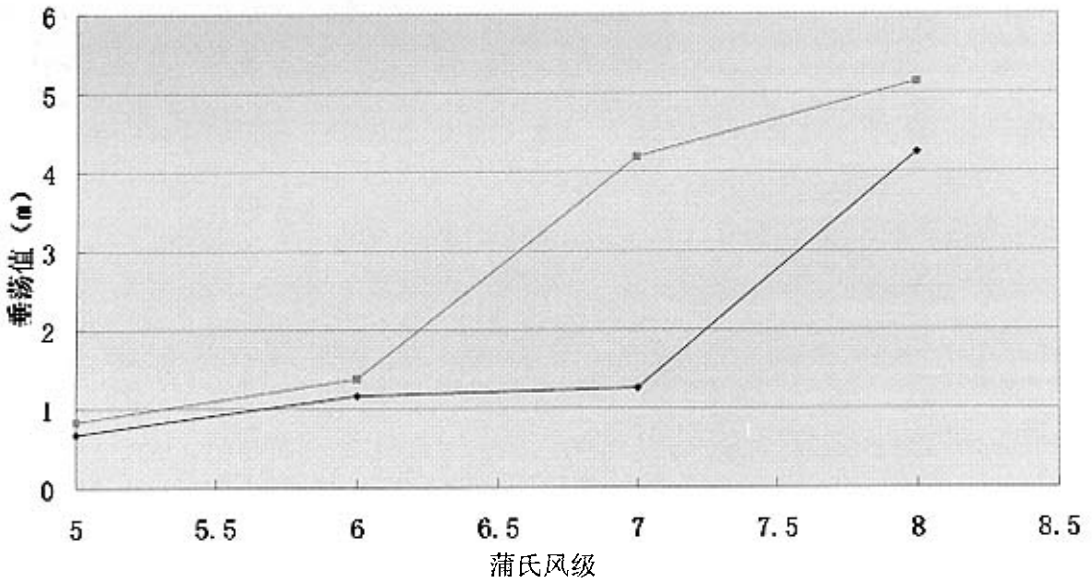


图 6.3 船长不同的万吨级船舶在各级风级中的垂荡值折线图

Fig. 6.3 The change of ship's draft at different wind power at different L_{BP}

根据表中计算结果，1 万 t 载重吨级长度为 140 m 的船舶在蒲氏 5 级风浪（有效波高 2.74 m）和蒲氏 6 级风浪（有效波高 4.24 m）中的垂荡幅值分别是 0.67 m 和 1.16 m，1 万 t 载重吨级长度为 120 m 的船舶在蒲氏 5 级风浪和蒲氏 6 级风浪中的垂荡幅值分别是 0.83 m 和 1.38 m。当有效波高为介于它们之间时，采用内插法可以得出垂荡幅值。

下面本文以各个船型不同级别的实船为例，分别计算出不同船型不同吨位级的船舶在蒲氏 5 级、6 级时的垂荡值。由于船舶在过跨海桥梁和架空线缆时的航速不可能很快，这里计算时船速都取 8 节。

表 6.5 散货船各吨位级的船舶的莫尔系数值表

Tab.6.5 The coefficient given by Moore of dry-cargo ships at different wind power

DWT (t)	船长 m	蒲氏风级	A0	A1	A2	A3	A4	A5	A6	A7
9914	112	5	-0.55	-13.55	4.94	0.075	82.89	0.21	24.9	4.26
		6	1.91	-18.37	6.33	0.040	105.42	0.24	30.49	5.94
22869	136	5	0.34	-12.95	4.16	0.13	72.53	0.20	21.52	3.22
		6	1.06	-19.63	6.79	0.102	111.63	0.28	32.92	5.56
27212	168	5	1.52	-11.07	2.80	0.18	53.32	0.18	16.33	1.86
		6	1.10	-19.58	6.22	0.19	106.22	0.30	31.88	4.48
52450	199.8	5	2.15	-8.45	1.63	0.178	33.97	0.14	11.63	0.91
		6	2.10	-16.94	4.55	0.236	83.64	0.27	26.08	3.031
74044	216	5	2.23	-7.29	1.26	0.17	26.48	0.12	9.74	0.58
		6	2.49	-15.26	3.76	0.24	70.87	0.25	22.99	2.39
109448	249	5	2.36	-5.07	0.58	0.14	12.47	0.079	6.15	-0.02
		6	3.16	-11.98	2.29	0.24	46.69	0.19	17.09	1.19
157998	271	5	2.29	-4.12	0.37	0.12	7.67434	0.06	4.67	-0.21
		6	3.14	-10.37	1.	0.22	37.31	0.16	14.49	0.78
179836	285	5	2.24	-3.52	0.23	0.11	4.62	0.05	3.73	-0.32
		6	3.14	-9.34	1.53	0.21	31.34	0.14	12.84	0.52
260783	311.9	5	2.15	-2.36	-0.03	0.09	-1.26	0.029	1.92	-0.55
		6	3.11	-7.37	0.95	0.19	19.83	0.11	9.65	0.02

表 6.6 散货船各吨位级的船舶在蒲氏风级 5 级和 6 级中的垂荡值表

Tab.6.6 The change of dry-cargo ship's draft at different wind power

船舶吨位	9914	22869	27212	52450	74044	109448	157998	179836	260783
垂荡值 (5 级风)	2.04	1.82	1.27	0.94	0.75	0.51	0.38	0.39	0.28
垂荡值 (6 级风)	3.02	2.998	2.30	1.85	1.55	1.11	0.74	0.69	0.55

船舶吨位与垂荡值得关系图

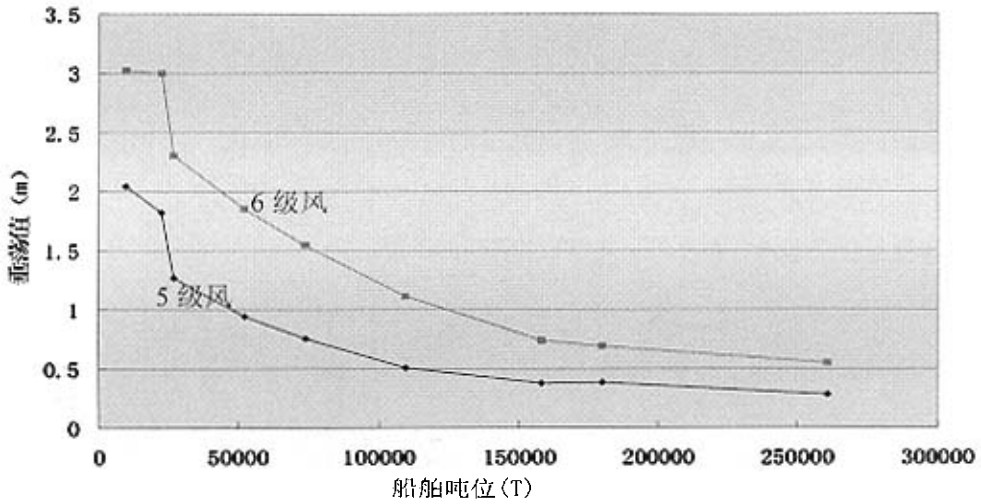


图 6.4 散货船船舶吨位与垂荡值关系图

Fig.6.4 The change of dry-cargo ship's draft at different wind power at different L_{BP}

表 6.7 油船各吨位级的船舶的莫尔系数插值表

Tab.6.7 The coefficient given by Moore of tankers at different wind power

DWT (t)	船长 m	蒲氏风级	A0	A1	A2	A3	A4	A5	A6	A7
12002	125	5	-0.07	-13.22	4.521	0.11	77.28	0.2	23.7	3.69
		6	1.45	-19.06	6.58	0.07	109.79	0.27	31.81	5.74
28750	165	5	1.41	-11.35	2.96	0.18	55.65	0.19	16.89	1.98
		6	0.99	-19.76	6.4	0.18	108.09	0.3	32.41	4.65
50746	237	5	2.33	-6	0.85	0.15	18.15	0.1	7.65	0.22
		6	2.93	-13.39	2.88	0.24	56.67	0.22	19.57	1.68
70060	266	5	2.34	-5.78	0.78	0.15	16.77	0.09	7.3	0.16
		6	3	-13.07	2.74	0.24	54.31	0.21	19.01	1.56
10106 2	234	5	2.308	-4.35	0.42	0.13	8.8	0.066	5.02	-0.16
		6	3.15	-10.75	1.94	0.23	39.54	0.17	15.11	0.88
15480	278	5	2.268	-3.82	0.3	0.12	6.15	0.056	4.2	-0.26
		6	3.14	-9.86	1.68	0.22	34.34	0.16	13.67	0.65
211469	288	5	2.36	-3.4	0.21	0.11	4.02	0.048	3.55	-0.35
		6	3.13	-9.15	1.47	0.21	30.18	0.14	12.52	0.47
25473 5	320	5	2.16	-2.59	0.03	0.1	0.28	0.035	2.4	-0.49
		6	3.12	-7.9	1.11	0.2	22.9	0.12	10.5	0.15

表 6.8 油船各吨位级的船舶在蒲氏风级 5 级和 6 级中的垂荡值表

Tab.6.8 The change of tankers draft at different wind power

船舶吨位	12002	28750	50746	70006	101062	145680	211469	254735
垂荡值 (5 级风)	1.92	1.39	0.76	0.66	0.43	0.429	0.424	0.35
垂荡值 (6 级风)	2.92	2.53	1.51	1.27	0.964	0.952	0.95	0.831

油船吨位与垂荡值关系

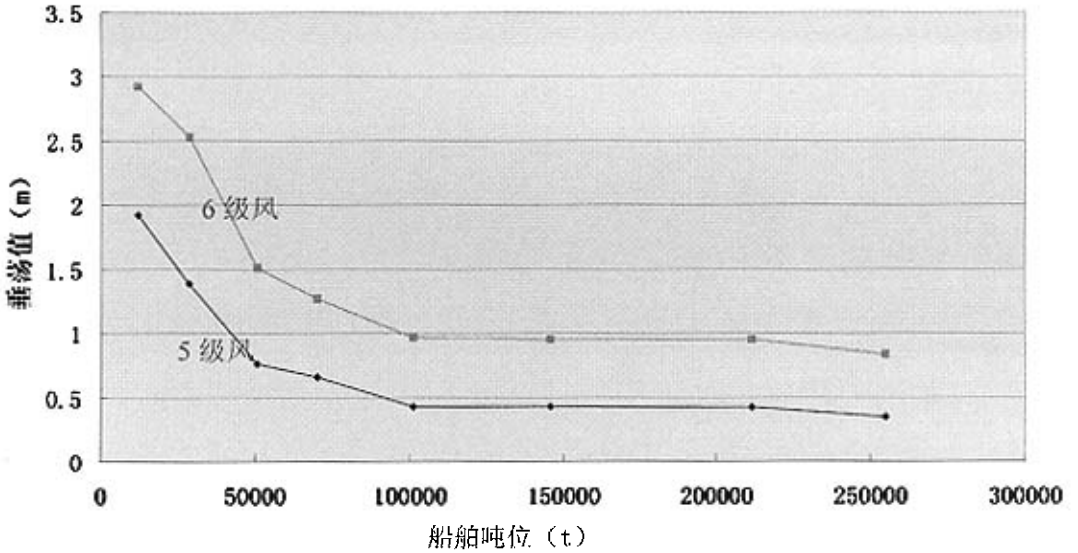


图 6.5 油船船舶吨位与垂荡值关系图

Fig.6.5 The change of tanker draft at different wind power at different L_{BP}

表 6.9 集装箱船各吨位级的船舶的莫尔系数插值表

Tab.6.9 The coefficient given by Moore of containers at different wind power

DWT (t)	船长 m	蒲氏风级	A0	A1	A2	A3	A4	A5	A6	A7
16750	152	5	-0.18	-13.3	4.62	0.10	78.6	0.205	23.5	3.83
		6	1.56	-18.19	6.52	0.066	108.0	0.262	31.5	5.79
23650	170	5	1.59	-10.87	2.69	0.184	51.78	0.180	15.96	1.78
		6	1.18	-19.46	6.11	0.19	104.9	0.30	31.54	4.38
34671	184	5	2.06	-9.59	1.99	0.19	41.37	0.16	13.49	1.23
		6	1.71	-18.61	5.33	0.23	96.27	0.29	29.12	3.66
43200	259.9	5	2.33	-4.59	0.47	0.13	10.07	0.07	5.41	-0.12
		6	3.15	-11.17	2.05	0.23	42.01	0.18	15.79	0.98

47425	261	5	2.32	-4.56	0.46	0.13	9.88	0.06	5.35	-0.12
		6	3.15	-11.11	2.03	0.23	41.62	0.17	15.68	0.97
72238	287	5	2.23	-3.43	0.21	0.11	4.18	0.05	3.59	-0.3
		6	3.13	-9.20	1.48	0.21	30.49	0.15	12.60	0.48
10886	331	5	2.09	-1.54	-0.2	0.08	-5.41	0.01	0.63	-0.70
		6	3.10	-5.98	0.56	0.18	11.73	0.09	7.40	-0.34

表 6.10 集装箱船各吨位级的船舶在蒲氏风级 5 级和 6 级中的垂荡值

Tab.6.10 The change of containers draft at different wind power

船舶吨位	16750	23650	34671	43200	47425	72238	10886
垂荡值 (5 级风)	1.50	1.32	1.28	0.83	0.76	0.59	0.40
垂荡值 (6 级风)	2.58	2.37	2.44	1.70	1.54	1.25	0.93

集装箱船舶吨位与垂荡值的关系

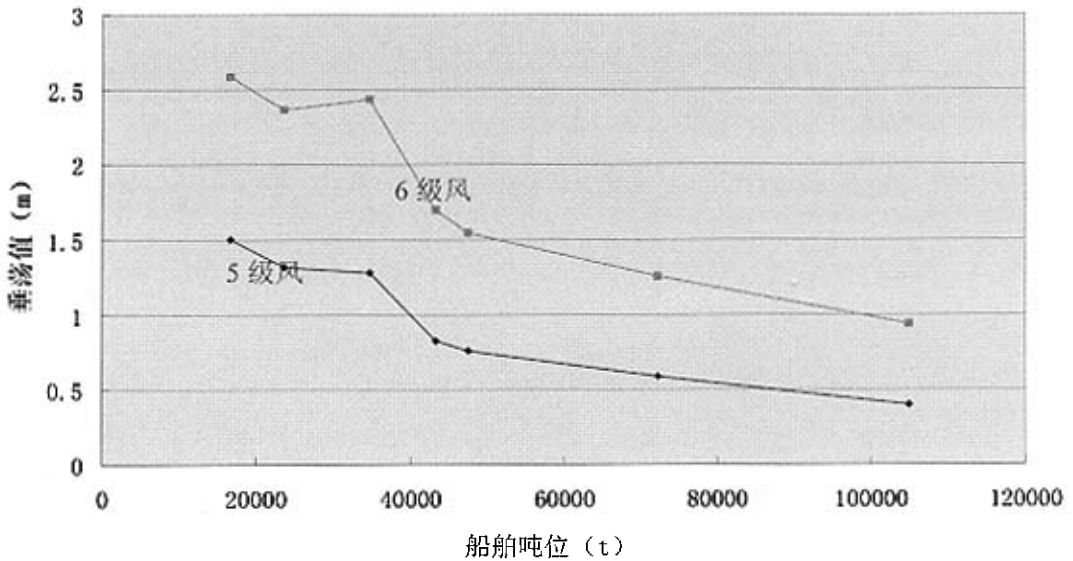


图 6.6 集装箱船舶吨位与垂荡值关系图

Fig.6.6 The change of container draft at different wind power at different L_{BP}

由于近年来航运发展的特点，大型化的集装箱船、油船和散货船占据了航运的主要市场，本文对其三种类型的船舶吨位和垂荡值之间的关系进行了统计分析。

由以上计算结果和关系图分析得出以下结论：

(1) 船舶吨位级别相同，船型相同时，有效浪高的最大值对船舶的垂荡值的影响很大，万吨级船舶在蒲氏 5~8 级风中的差值很大，所以如何准确的确定跨海大

桥和架空线缆附近海域的波浪最大值非常重要。

(2) 不同船型，相同吨位级别的船舶在同样风力的风浪中的垂荡值也不一样，在同样的条件下，散货船和油船的垂荡值较集装箱船舶的垂荡值稍大。

(3) 船舶在风浪中的垂荡值与船舶的方形系数、吃水、航速、船长、水线面系数和船宽等都有关系，在考虑船舶的垂荡值时，应考虑到船舶的所有类型以及吨位，相同吨位级别的船舶当船长不同时，垂荡值的差别也不小，要经过认真的计算和比较，取其较大者作为标准。同时要确定该海区的准确地波浪高度，根据波浪的特点和历史资料，结合大桥和架空线缆的使用年限，最终确定核算最大波高值，这样得出的船舶垂荡值才有意义。

6.2.3 其他因素分析

(1) 船舶驾驶安全高度

船舶驾驶安全高度是一个综合性的因素，其中包括驾驶员的心理素质。船舶越大，波浪越高，船舶过桥时航行条件越差，要求的安全高度也越大。这样当跨越大桥或架空线缆时，驾驶员应谨慎驾驶。

(2) 船舶水线面以上之最高固定点高度的计量误差有以下情况造成：不同的载重量，船舶大修后发生的高度变化，船舶高度测时的误差，船舶在淡水、海水中的吃水差等。

(3) 目前的技术水平在水位观测预报和气象增水的预报方面存在着一定的误差，但其量值不大，可以忽略不计。

6.2.4 桥梁结构挠度和基础沉降对富裕高度的影响分析

由于桥梁结构挠度和基础沉降的计算在公路工程建设标准中有相应的计算方法，并且桥梁设计时已考虑，因此富裕高度不包含着部分因素。

6.2.5 对于架空线缆应特别考虑的对富裕高度有影响的因素分析

(1) 电气安全距离：为了防止意外的人和车辆等接近带电体及防止电气的短路和放电，规定带电体与别的设备和设施之间，带电体相互之间均需保持一定的安全距离。计算跨海线缆的净空高度时应考虑此项，有关规定 1 万伏电的安全距离是 0.7 米，6.6 万伏的安全距离是 1.5 米，22 万伏的安全距离是 3 米，50 万伏的

安全距离是 5 米。

(2) 由于随着导线负荷的增加，导线的温度会增大，温度的变化会对架空线缆的垂度有较大的影响。一般在电力部门设计中，导线垂度是按导线温度 100℃ 时的状况计算的，所以导线在日常工作中 (50℃)，净空高度一般小于设计时的垂度。为了检验温度对架空线缆垂度的影响，本文还以上面的架空线缆为例，对该架空线缆在各种温度下的垂度进行计算，结果如下表：

表 6.11 各种温度下的线缆的垂度

Tab.6.11 The f of wires at different position at different temperature

温度	垂度
50	260.4
60	261.65
70	262.94
80	264.23
90	265.52
100	266.8

利用上面给出的计算架空线缆各点距离水准面的垂直距离的方法，将该架空线缆在 100℃ 和 50℃ 时各点的高度变化计算出来，结果如下表：

表 6.12 线缆各点在 100℃ 和 50℃ 时距离差

Tab.6.12 The vertical distance of wires at different position at temperature 100℃ and 50℃

距离最低点距离 (m)	0	100	200	300	400	500	600
50 度时各点的高度	6.4	7.76	11.84	18.65	28.19	40.46	55.49
100 度时各点的高度	0	1.39	5.53	12.55	22.31	34.89	50.27
100 度和 50 度时该点的距离差	6.4	6.37	6.31	6.10	5.88	5.57	5.22
距离最低点距离 (m)	700	800	900	1000	1100	1200	1300
50 度时各点的高度	73.27	93.82	117.16	143.29	172.26	204.06	238.74
100 度时各点的高度	68.48	89.53	113.4	140.21	169.9	202.49	238.03
100 度和 50 度时该点的距离差	4.97	4.29	3.73	3.08	2.36	1.57	0.71

从计算结果可以看出，温度对跨海线缆的垂度的影响较大，所以在考虑富裕高度是应引起设计者的高度重视。

第七章结论

7.1 结论

(1) 本文通过对大量的实船数据进行分类、整理, 利用 SPSS 等软件对数据进行非线性回归分析, 找到了不同船型的船舶空载水线面以上至最高固定点高度值随船舶吨位的分布规律, 得到了有一定参考价值的船舶空载水线面以上至最高固定点高度值的预测公式。公式如下:

表 7.1 回归模型公式

Tab. 7.1 expressions of regression curve

船型	回归模型预测公式	回归模型的决定系数 R^2
散货船	$H_{\max} = 7.169\ln(T) - 34.093$	0.8644
杂货船	$H_{\max} = 8.004\ln(T) - 37.895$	0.8746
油船	$H_{\max} = 6.431\ln(T) - 27.018$	0.8751
集装箱船	$H_{\max} = 9.0687\ln(T) - 46.571$	0.8927

(2) 通过对本文的研究得出一种新的计算跨海桥梁和架空线缆通航净空高度的公式:

跨海桥梁净空高度值=有本文回归公式得出的船舶空载水线面以上最大高度预测值+ 2σ +海平面预测变化量+经计算得出的船舶垂荡值+其他因素对富裕高度的影响值。

跨海架空线缆通航净空高度值=有本文回归公式得出的船舶空载水线面以上最大高度预测值+ 2σ +海平面预测变化量+经计算得出的船舶垂荡值+其他因素对富裕高度的影响值+电气安全距离+温度变化引起的架空线缆垂度变化值。

7.2 本文所作的研究工作

(1) 收集了大量的实船数据并对其分类、处理, 得出了一定参考价值的船舶空载水线面以上至最高固定点高度值的回归预测公式。

(2) 利用莫尔插值法对相同船型相同吨位级别的船舶在不同风级的海浪中的

垂荡值和不同船型相同吨位级的船舶在同等风级的海浪中的垂荡值进行了计算和分析比较，得出了相应的结论。

(3) 对架空线缆各点的通过能力进行了计算，这样做是为了便于船舶驾驶在驾驶船舶过跨海大桥和架空线缆时参考。

7.3 展望

影响富裕高度因素很多，同时很多因素的不确定性太大，所以研究跨海桥梁和架空线缆的通航净空高度是一个复杂的问题。限于文章的篇幅及作者目前的研究方向和能力，还有收集到的船舶数据没有达到足够多，以至于文中提到的用非线性回归的方法来处理船舶空载水线面以上至最高固定点与载重吨之间的关系时，考虑的因素不够全面，在回归时也没有按吨位级别分别对不同船型进行回归分析，只是给出了基础的回归公式。希望能起到抛砖引玉的作用，能引起更多的航海界的人士进行此方面的研究，为祖国的航运事业作出自己应有的贡献。

参考文献

- [1] 中交水运规划设计院, 通航海轮桥梁通航标准, 北京: 人民交通出版社, 1998
- [2] 张锦文, 王喜亭, 未来中国沿海海平面上升趋势估计, 测绘通报, 2001 No.4 P.4-5
- [3] 杨佐昌等, 跨海大桥净空高度的确定方法, 大连海事大学学报, Vol.30.No.4
Nov..2004
- [4] 邵旭东, 桥梁工程, 北京: 人民交通出版社, 2004
- [5] 邱大洪, 工程水文学, 北京: 人民交通出版社, 1999
- [6] 李积德, 船舶耐波性, 哈尔滨: 哈尔滨船舶工程学院出版社, 1992
- [7] Douglsa M.Bates, 韦博成等译, 非线性回归及其应用, 北京: 中国统计出版社, 1997年9月
- [8] 李庆扬, 数值分析, 武汉: 华中科技大学出版社, 1986年12月
- [9] D · A · Ratkowsky 著, 洪再吉等译, 非线性回归模型统一的实用方法[P], 南京: 南京大学出版社, 1986年
- [10] 岳兴旺, 大型船舶尺度相关性研究, 大连: 大连海事大学, 2005
- [11] 秦士元, 程天柱, 水运系统分析和船型研究, 上海: 上海交通大学出版社, 1997
- [12] 中交水运规划设计院, 海港总平面设计规范, 北京: 人民交通出版社, 2003
- [13] 中交第一航务工程勘察设计院, <<海港总平面设计规范>>局部修订, 北京: 人民交通出版社, 2003
- [14] He xiaoqun, Multiple variable Statistical Analysis of the Causes of National Income Growth in China, Hong Kong: IS MAA, 1992 版社, 1997
- [15] Frank L.E. and Friendman J.H. A Statistical View of Some Chemometrics Refression Tools Technometrics, 1993(35)
- [16] 戴思锐, 计量经济学, 北京: 中国农业出版社, 2003
- [17] 于洪彦, Excel 统计分析与决策, 北京: 高等教育出版社, 2001
- [18] 顾敏童, 船舶设计原理, 上海: 上海交通大学出版社, 2001
- [19] 黄霖生, 回归分析在经济中的应用, 陕西: 陕西经济统计出版社, 1990
- [20] 何晓群, 回归分析与经济数据建模, 北京: 中国人民大学出版社, 1997
- [21] 周纪芾, 回归分析, 上海: 华东师范大学出版社, 1993

- [22] 郑会尧, 数值计算方法, 武汉: 武汉大学出版社, 2002
- [23] 马春庭, 掌握和精通 SPSS10, 北京: 机械工业出版社, 2001
- [24] 李子奈, 计量经济学——方法和应用, 北京: 清华大学出版社, 1992
- [25] 俞宝均, 大型集装箱船的技术现状和发展趋势, 上海: 上海造船, 1998
- [26] 李树范, 船舶设计原理, 大连: 大连理工大学出版社, 1998
- [27] 张德洪, 运输船舶船型技术经济论证方法 (第二版), 北京: 国防工业出版社, 1989
- [28] 周超骏, 计算机辅助船体线型设计, 上海: 上海交通大学出版社, 1992
- [29] Shoichiro Nakamura, 科学计算引论, 梁恒 刘晓艳, 北京: 电子工业出版社, 2002
- [30] 拉姆·拉玛纳山, 应用计量经济学, 薛菁睿, 北京: 机械工业出版社, 2003
- [31] 陆志才, 船舶操纵, 大连: 大连海事大学出版社, 2000
- [32] 周东平, 船舶操纵, 北京: 人民交通出版社, 1996
- [33] 陈华清, 船舶操纵, 北京: 人民交通出版社, 1992
- [34] 于汝民, 港口规划与建设, 北京: 人民交通出版社, 2003
- [35] 高素英, 奇异观测值的判别及其影响力的分析, 河北: 河北工业大学学报, 2002
- [36] 专题报道, 世界经济继续复苏, 瞭望新闻周刊, 2004 年第 14 期
- [37] 蒋维清, 船舶原理, 大连: 大连海事大学出
- [38] 陈英同, 计算方法, 厦门: 厦门大学出版社, 2001
- [39] 郑文振, 陈宗镛 1 中国东部沿海地区与海平面与陆地垂直运动[M]1 北京: 海洋出版社, 19991
- [40] 赵明才, 马继瑞, 等 1 由海平面变化推求中国沿海地区地壳垂直运动[J]1 海洋测绘, 2001, (3): 9~131
- [41] 管铮 1 中国沿海及太平洋西部海区深水测深改正问题[J]1 海洋测绘专辑, 1986, (4): 42~501

攻读学位期间公开发表论文

1] 施行永、戴冉、朱金善、白宇明， 跨航道桥梁和架空线缆净空高度的研究，
航海技术与航海教育论文集，大连海事大学出版社，2006年3月

致 谢

光阴荏苒，两年半时光很快过去了，在短短两年半的时间里，戴老师在做人和做学问上都给了我很大的帮助。戴老师敏捷的才思、渊博的知识和严谨的治学作风永远是我学习的榜样。

本文从论文选题到调研，再到撰写，最后定稿，恩师戴冉教授都给了我精心的指导，在此向我的导师戴冉教授表示衷心的感谢！

感谢航海学院朱金善老师、王少青老师、岳兴旺老师，感谢他们在论文选题与写作过程中给予的指点与关怀。

感谢航海楼 212 实验室的全体师兄师弟，感谢他们在本文收集材料过程中给予了很大的帮助和支持，感谢他们给我了一个舒适的写作环境。

感谢我的父母及家人，是他们的关心、理解与支持给了我学习的动力，使我能够安心于学习，也使本文的工作得以顺利的完成。

感谢所有曾经关心和帮助过我的老师、同学和朋友。

研究生履历

姓 名	施行永
性 别	男
出生日期	1979年2月22日
获学士学位专业及门类	工学
获学士学位单位	大连海事大学
获硕士学位专业及门类	工学
获硕士学位单位	大连海事大学
通信地址	辽宁省大连市凌海路1号
邮政编码	116026
电子邮箱	ayongshi@sina.com