

Research on the Application of Mathematical Statistics in Data Analysis

Tingting Duan

School of Statistics, Renmin University of China, Beijing 10000, China

Abstract

Data is an essential element of all production activities. However, since data is generated from a wide range of sources and forms, it must be sorted and counted in order to obtain effective data. Because computers have powerful data processing capabilities, big data technology has been widely used in the production and operation of major enterprises in recent years, and the development of big data technology is based on mathematical statistics and data analysis. The paper is based on the application of mathematical statistics in data analysis.

Keywords

mathematical statistics; data analysis; applied research

数理统计在数据分析中的应用研究

段婷婷

中国人民大学统计学院, 中国·北京 10000

摘要

数据是一切生产活动的必备要素,然而,由于数据产生的来源广,形式多,要获得有效的数据就必须对其整理和统计。由于计算机对数据有着强大的处理能力,近些年大数据技术被广泛应用于各大企业的生产经营,而大数据技术发展的基础就是数理统计和数据分析。论文基于数理统计在数据分析中的应用进行研究。

关键词

数理统计; 数据分析; 应用研究

1 引言

自古以来数据就存在于人们的生活中,无论是人口统计还是社会调查都能看到数理统计的应用。近些年,中国载人航天飞船的研制,各种互联网企业的蓬勃发展,都离不开数据的支撑。随着现代移动互联网的发展,数理统计的方式变得多样化和智能化,数据分析也变得越来越专业化。为了提供更有效的数据分析结果,正确导引生产生活的进行,对于数理统计进行研究是非常有必要的。

【作者简介】段婷婷(1993-),女,中国宁夏中卫人,中国人民大学硕士,金融科技工程师,从事金融科技行业智慧政务领域产品设计与业务分析研究。

2 数理统计的含义

2.1 什么是数理统计

从专业的角度看,数理统计是数学专业中的一个分支,主要有描述统计和推断统计两种方式。它以概率论为基础来研究社会现象和自然现象的规律性,并通过具体的实验活动收集并分析过程中产生的各种数据,为相关的决策者提供行动依据或者建议^[1]。

从具体应用来看,数理统计在自然科学、工程技术、管理科学等多个学科中有着交叉应用,推动着各类科学技术的进步。数理统计的作用必须借助数据分析得到体现,因此数理统计注重数据收集的有效性,要尽可能确保数据没有干扰因素。

2.2 数理统计的具体步骤

从具体过程来看,数理统计分为三个阶段。一是准备阶段的实验设计;二是实施阶段的数据收集;三是验证阶段的统计推断。无论是企业还是科研单位在进行数理统计时,首先要根据统计对象建立好合理的数学模型,确定收集的数据参数,其次要根据建立的模型进行数据收集整理,并将整理好的数据用于分析,最后进行预测或者决策^[2]。

在实验设计过程中,相关人员一定要选用合理的试验方法,确定好实验的影响因素和变量,避免收集到的数据存在巨大误差,这是数理统计正确进行的前提。数据收集阶段可以采用多样化的收集方式,对于人文社科类的调查要利用好互联网设备来进行收集,确保数据的准确性。在数据分析阶段,首先要筛选出不合理的数据源,以免影响最终的分析结果。

2.3 数理统计的特点

数理统计要根据不同的情况采用不同的方式。在中国的人口普查过程中,就需要采用全面调查的方式,尽可能确保数据的完整性和有效性。在社会经济活动中,要确保数理统计不会干预实际活动的进行,可以采用抽样调查的方法。数据收集完成后,要根据数据特征进行分类整理,用图表的形式进行呈现。

数理统计的特点是全面和随机性,如非必要的人工干预,尽可能保证数据来源的随机性和普遍性。在对于数据进行分析时,要考虑到数理统计的方式,从而选用合理的指标来进行分析。

3 数理统计与数据分析在各行业中的应用

3.1 农业领域的应用

由于农作物的生长特性,其生长繁育数据对于改良下一代农作物有着极其重要的作用,因此必须使用数理统计方式对收集到的数据进行分析^[3]。例如,要想获得优质的农作物种子,首先要运用统计学知识正确设计田间试验,根据种苗存活率,农作物产量等多个方面进行统计。要结合相关的线性回归方法,计算好优良种苗的存活度。中国杂交水稻之父袁隆平院士在改良杂交水稻过程中,进行了大量的田间试验,并成功对数万稻穗进行数理统计分析,最终成功培育杂交水稻,解决了中国十几亿人口的生存问题。因此,数理统计在农业领域中的合理运用,能够有效解决实际问题,提升人民生活质量。

3.2 工业领域的应用

随着社会的不断发展,人们对于工业产品有着更高的要求,工业生产中合理运用数理统计的方式能够有效对工业产品进行改良,实现代际淘汰,提升市场竞争力。在新产品试制时,要使用正交设计、多元回归等统计方法,确定批量生产的产品最终样板,要在生产过程中,对其工艺流程所用耗材和能量进行统计分析,优化材料用量和生产工艺。在对于已经生产完毕的商品出库时,要采用抽样调查的方式确保商品的合格性。

在现代工业中,企业可以通过数据的控制来把控生产成本,由于工业产品的大批量生产属性,企业可以通过数据统计来核算商品的生产成本,每个产品节约1分钱在庞大基数下也会给企业带来客观的经济效益。另外,由于数据的独特性,许多大型企业也会将其运用在仓储上,如京东自建仓库,通过对数据进行统计可以设计出合理的商品放置布局,从而减少人员的时间成本。

3.3 医学和药学领域的应用

数理统计在医学和药学上已有上百年的历史。由于引起疾病的病原体存在变异的可能性,再加上每个人的免疫能力不同,即使人们得了相同的病也可能会有不同的病症特征^[4]。因此,在医学和药学中必须对采用大样本的数理统计方式,从而得出病症的共性和特异性。例如,“抽烟有害健康”,这是医学工作者从数理统计和数据分析的角度论证了吸烟对肺部发生病变有着强大的促进作用,是用数据对抽烟的有害性进行了证明。另外,近些年肆虐的新冠肺炎,国家采取的防控措施就是基于其传播特征,通过阻断传播途径进行疫情防控。通过对新冠肺炎的发病形式和传染特征进行分析,确定了密切接触者的具有极大地可能被传染,再通过网络平台向全社会公布发病人数和密切接触者,都是数理统计和数据分析的应用

3.4 自然科学领域的运用

由于科学理论具有前瞻性,必须对其进行试验和论证。人们在验证一个学说或者假说是否正确时,通常会采用设计试验来进行验证。在试验设计环节,要根据其具体假说的内容设计数理统计实验,设计对比试验、对照试验等形式。通过对实验现象中物理数据或者化学数据进行测量和记录,形成原始的数据文件。最后通过分析得出该学说是否正确。在

孟德尔的豌豆杂交实验中,孟德尔对豌豆花朵颜色进行了猜想,并假设豌豆体内有一种物质决定了其花的颜色。通过多次试验,孟德尔利用数据统计的方法,分别采用正交和反交的方式进行试验,确定了显性性状和隐性性状的存在。孟德尔的杂交实验产生了大量数据,孟德尔对试验数据的分析推动了生物学界的迈向了新的阶段。

3.5 经济领域的应用

由于经济与社会发展有着极其紧密的联系,无论是集体和个人对自身的经济状况要进行充分的掌握。因此,数理统计和数据分析对于集体和个人而言,有着至关重要的影响^[5]。由于经济风险的客观存在,国家财政部门需要保证社会经济的平稳运行,才能保证生产生活的顺利进行。

数据是国家制定经济政策的理论依据,由于国家人口多,社会结构复杂,如果没有充分的数理统计方式,制定的政策将会不具有可行性。例如,津巴布韦为了对抗国内经济整体下滑的趋势,没有以国内经济数据分析为政策导向,而是采取疯狂印钞的财政政策,导致津巴布韦的钞票不具有实际价值,并带来了一系列的社会问题。国家经济政策是对国民生产生活的有利调控,必须充分以全面科学的经济数据为导向。对于企业而言,经济上的数理统计是日常工作,为了确保企业财务的良性循环,企业要根据相关财务制度确定经济管理方式,防止由于资金链断裂影响企业的正常经营。此外,企业要充分对市场数据进行分析,用市场反馈的销售数据、客户满意度等加大竞争优势,弥补竞争劣势,以确保经济效益最大化。对于个人而言,财务收支是客观存在的,一个人要确保自己的生活状况不出现问题,就需要对自己的财务状况进行统计和分析,确保个人经济状况健康增长。

3.6 企业经营管理过程中的运用

现代企业的发展必须以数据为导向。

从市场层面出发,企业必须对市场环境,消费者购买力等多个数据源进行分析,确定企业能够在当前市场下有良好的竞争优势,相关的产品能够满足市场需要。

从生产层面出发,企业要充分以市场的数据为导向,进行产品换代和更新,对于不符合生产需要的工艺流程,要进行革新和改善,减少企业运营成本。在人员管理上,企业更要充分发挥以KPI为导向的数据考核方式,形成有数据可依的权威性管理方式。另外,针对电商企业,要着重收集用户

流量来源,分析消费者的产品喜好,以此来调整企业的市场定位。

4 如何确保数理统计和数据分析的有效性

第一,要确保数理统计和数据分析的有效性,在建模过程就要引起重视,确保建立的理论数据模型尽可能贴合现实需要。

第二,在试验过程中,要尽可能减少干扰因素,确保试验不受不相干的因素影响。

第三,对采取到的数据要进行合理整理,采用科学的数学统计方法对数据做好单独分析和综合分析。在分析过程中,要合理利用SPSS等软件工具对数据进行高效处理,充分考虑由于事件概率、样本误差等情况引起的数据误差,对数据进行合理取舍。

第四,要多次反复核验数据的准确性,对于社会调查或者科学实验,可以采用重复实验的方式提高数据的准确性。

5 结语

数理统计是建立在数学理论上精准数据分析方式,现代的数理统计有更强的理论指导和实践意义,它被广泛运用于日常生活中的各行各业。无论是国家还是个人,都需要采用科学的数理统计方式对自身的发展进行审视,从而有针对性地对存在问题的环节进行改善。对于不同的领域,数理统计的方式也存在差异,要兼顾方便和准确的原则进行数据分析。现代化的数理统计要尽可能结合计算机软件进行分析,提高数据分析的精确性。

参考文献

- [1] 钟明.探索数理统计在数据分析中的应用[J].中国新通信,2020,22(15):113.
- [2] 韩静.浅谈数理统计在数据分析中的应用[J].发明与创新(职业教育),2020(6):65.
- [3] 李秀昌,孙健.数理统计在数据分析中的应用[J].教育教学论坛,2020(8):152-153.
- [4] 谢逢彬.数理统计在数据分析中的应用探讨[J].中国新通信,2019,21(9):110.
- [5] 邢书源.数理统计在数据分析中的应用研究[J].学周刊,2019(6):185-186.