

煤炭检验 统计技术

主编 周尊英 李宜轩 陈玉明



 中国标准出版社

煤炭检验 统计技术

—— 煤炭检验统计技术

煤炭检验统计技术

煤炭检验统计技术

主编 周尊英 李宜轩 陈玉明

中国标准出版社

北京

图书在版编目(CIP)数据

煤炭检验统计技术/周尊英,李宜轩,陈玉明主编.
北京:中国标准出版社,2007
ISBN 978-7-5066-4590-4

I. 煤… II. ①周…②李…③陈… III. 统计-应用-煤炭-检验 IV. TD942.63

中国版本图书馆 CIP 数据核字(2007)第 116081 号

中国标准出版社出版发行
北京复兴门外三里河北街 16 号
邮政编码:100045

网址 www.spc.net.cn

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 787×1092 1/16 印张 19.75 字数 476 千字
2007 年 8 月第一版 2007 年 8 月第一次印刷

*

定价 45.00 元

如有印装差错 由本社发行中心调换

版权专有 侵权必究

举报电话:(010)68533533

前 言

煤炭是我国最主要的能源,产量和消费量都稳居世界首位。因此,煤炭在我国的国民经济中占有非常重要的地位。

在煤炭的勘探、生产、加工等诸多环节以及电力、冶金、建材、贸易、运输、环保、科研等诸多领域中,都需要对煤炭进行检验,以煤质的分析结果来评价和指导各项业务的开展。煤质分析必须以煤炭的检验标准为依据,在众多的煤炭检验标准(包括行业标准、国家标准、国际标准和其他国外标准)中,基本上都是以统计技术作为理论基础或直接采用一些统计方法和统计工具的。由于统计技术理论较深、比较抽象,因而在煤炭检验标准的贯彻和执行中,许多煤质分析人员对检验标准感到难以掌握和理解,特别是从事煤炭采样和制样的工作人员对繁多的统计计算更是无所事从。同时在煤炭实验室日常检验中,数据处理、统计分析、质量体系建立运行、过程监控、能力验证、方法确认、仪器核查、科学研究以及科技论文的撰写等也都离不开统计技术的支持。

为了提高广大煤炭检验人员、煤质管理人员的统计技术理论水平和应用技能,在煤质分析工作中正确贯彻、执行、理解和运用煤炭检验的标准,提高煤炭检验的工作质量,确保检验结果的准确度,我们编著了这本《煤炭检验统计技术》,以期帮助大家达到上述目的。

本书共分14章,第1章至第8章结合煤炭检验的一些实例和示例,比较系统地介绍了统计技术的理论和方法;第9章至第14章具体地阐述了统计技术在煤炭实验室质量管理体系、质量控制、煤炭采样、采样设备性能验收、煤炭制样、煤质分析以及科学研究中的一些具体应用。

本书编著人员参阅了大量国内外有关的最新标准和文献,根据煤炭检验的具体实践和科学研究,密切结合统计技术的理论、方法,编著了该书。因此,书中的概念定义标准规范、内容有较大的创新,并很好地解释、纠正和理顺了煤炭检验中存在的一些模糊概念、认识

和不确切做法,使煤炭采样、制样和化验标准更加切合实际,更具操作性。本书一方面考虑了统计技术理论的系统性,另一方面更多的还是结合煤炭检验的实际需要,对煤炭实验室日常检验所需的一些统计方法和工具,以尽量多的实例和示例加以说明,使该书能够深入浅出、通俗易懂。

本书是煤炭、电力、冶金、建材、质检、贸易、环保等行业的煤炭质量管理人员、煤炭检验人员的一本较好的参考读物,对有关大专院校、科研机构的技术人员在该领域也有一定的帮助。虽然不同行业和领域的实验室所检验的物料不同,但所应用的统计技术理论和方法大体上是一致的,因此该书对所有从事实验室理化检验人员也有一定的参考价值。

在本书的编写过程中得到了国家质量监督检验检疫总局、山东出入境检验检疫局、日照出入境检验检疫局各级领导的大力帮助和支持,在此深表感谢!

由于作者水平有限,书中难免有不当之处,敬请读者批评指正。

编 者

二〇〇七年三月

目 录

第 1 章 统计技术的基础知识	1
1.1 统计技术概述	1
1.2 随机现象及其规律性	1
1.2.1 随机现象	1
1.2.2 随机现象的规律性	2
1.3 统计数据及统计方法的分类	2
1.3.1 统计数据的分类	2
1.3.2 统计方法的分类	2
1.4 随机变量及其分布	3
1.4.1 随机变量	3
1.4.2 随机变量的概率分布	4
1.4.3 离散型随机变量的分布	6
1.5 正态分布	8
1.5.1 正态分布的定义与性质	8
1.5.2 标准正态分布	10
1.5.3 正态分布的概率计算	10
1.5.4 正态分布的特征值	12
1.6 抽样分布	14
1.6.1 χ^2 分布	14
1.6.2 t 分布	15
1.6.3 F 分布	17
1.7 大数定律和中心极限定理	18
1.7.1 大数定律	18
1.7.2 中心极限定理	18
1.8 统计技术在煤炭检验中的应用	19
1.8.1 煤炭实验室质量管理体系	19
1.8.2 煤炭检验标准	20
1.8.3 煤炭的采样和制样	20
1.8.4 煤炭质量特性的检测	20
1.8.5 煤炭检验的科学研究	20
1.8.6 统计技术在其他方面的应用	21
习题一	21
第 2 章 试验误差理论	22
2.1 测试误差	22
2.1.1 真值和约定真值	22
2.1.2 测试误差与误差的表示方式	22
2.2 误差的分类	24
2.2.1 系统误差	24
2.2.2 随机误差	25
2.2.3 过失误差	25
2.3 准确度(正确度与精密度)	26
2.3.1 正确度	26
2.3.2 精密度	27
2.3.3 准确度	27
2.3.4 准确度与正确度、精密度的关系	27
2.3.5 精密度的评价	28
2.3.6 如何评价和提高测试结果的准确度	30
2.3.7 重复性与再现性	37
2.4 精密度的主要表示方法	39
2.4.1 极差	39
2.4.2 标准差	39
2.4.3 方差	40
2.4.4 变异系数	40
2.5 测试不确定度	41
2.5.1 不确定度的有关概念	41
2.5.2 测量误差与不确定度的区别	41
2.5.3 不确定度的评定过程	42
2.5.4 煤炭检验不确定度评定示例	44
习题二	48
第 3 章 试验数据处理	49
3.1 数字修约	49
3.1.1 有效数字	49
3.1.2 有效位数与修约间隔	49
3.1.3 修约规则	51
3.2 运算法则	54
3.2.1 加法与减法的运算	54
3.2.2 乘法和除法的运算	54
3.2.3 四则运算的演算方法	55
3.2.4 乘方、开方以及对数、反对数的运算	56
3.3 记数的一般规则	56
3.4 数字特征	57
3.4.1 随机变量的数字特征	57
3.4.2 样本数字特征	59

3.5 离群值的判断与处理	61	原理	90
3.5.1 离群值的判断	61	5.3.2 常规控制图的种类和选择	92
3.5.2 处理异常值的一般规则	62	5.3.3 分析用控制图和控制用控制 图的判别准则	93
3.5.3 在未知标准差情况下异常值 (≤ 1)的判断和处理	62	5.3.4 制订和使用控制图的步骤	93
3.5.4 在未知标准差情况下判断和 处理异常值(>1)的规则	64	5.3.5 均值-极差控制图($\bar{x}-R$ 图)	94
3.5.5 科克伦最大方差检验法判断 和处理的准则	64	5.3.6 均值-标准差控制图($\bar{x}-s$ 图)	98
3.6 数据的插补	65	5.3.7 质量控制图在实验室中的应用	101
3.6.1 数据插补的应用	65	习题五	101
3.6.2 线性内插法	66		
习题三	67		
第4章 统计推断	69	第6章 方差分析与试验设计	102
4.1 概述	69	6.1 方差分析	102
4.1.1 统计推断的目的及内容	69	6.1.1 方差分析中常用统计术语	102
4.1.2 统计推断的有关概念	69	6.1.2 方差分析的基本思想、方法步骤 及应用条件	103
4.2 参数估计	72	6.1.3 单因素方差分析的程序	103
4.2.1 均值 μ 的估计	72	6.1.4 双因素方差分析的程序	107
4.2.2 总体方差 σ^2 、总体标准差 σ 的 估计	73	6.2 试验设计	111
4.2.3 总体方差 σ^2 、总体标准差 σ 的 区间估计	74	6.2.1 试验设计中的常用术语	112
4.3 显著性检验	75	6.2.2 正交试验法	113
4.3.1 显著性检验中的有关术语	75	6.2.3 有交互作用(交互效应)的正交 试验	118
4.3.2 第一类错误与第二类错误	77	习题六	119
4.3.3 显著性检验的步骤与方法	78		
4.3.4 t 检验	79	第7章 回归分析	120
4.3.5 χ^2 检验	82	7.1 概述	120
4.3.6 F 检验	83	7.2 函数关系与相关关系	120
习题四	84	7.3 散点图	121
		7.4 回归函数、回归方程与回归系数	121
第5章 统计过程控制	85	7.5 最小二乘法	122
5.1 统计过程控制的意义	85	7.5.1 用最小二乘法求回归方程的 方法	122
5.2 过程能力分析	86	7.5.2 使用最小二乘法求一元线性 回归方程的示例	123
5.2.1 过程能力	86	7.6 相关系数	124
5.2.2 过程能力指数	86	7.7 回归方程的相关显著性检验	126
5.2.3 过程能力等级的评定与应用	88	7.7.1 统计推断法	126
5.2.4 过程能力指数与过程不合格 品率的关系	89	7.7.2 相关系数检验法	127
5.2.5 过程能力的应用	89	7.8 利用回归方程进行预测和控制	128
5.3 SPC图(统计过程控制图)	89	7.8.1 关于预测问题	128
5.3.1 常规控制图的基本格式及统计		7.8.2 关于控制问题	129

7.9 煤炭检验中灰分与发热量一元 回归方程的建立	130	第 10 章 煤炭采样统计技术应用 ...	178
7.10 煤炭检验中灰分、挥发分与发热量 二元回归方程的建立	132	10.1 煤炭采样概述	178
7.10.1 二元线性回归方程的建立步骤	132	10.1.1 煤炭采样	178
7.10.2 灰分、挥发分与发热量二元回归 方程的建立	133	10.1.2 煤炭采样的重要性	178
7.11 煤炭检验中常用的一些相关关系	134	10.1.3 煤炭采样的理论依据	178
7.11.1 煤的工业分析与元素分析结果 间的关系	134	10.1.4 煤炭采样原理	179
7.11.2 煤的元素分析与各指标间的 关系	135	10.1.5 煤炭采样术语及定义	180
7.11.3 发热量的有关回归方程式	136	10.2 采样方案	181
7.11.4 灰熔融性与煤灰成分分析结 果的相关关系	137	10.2.1 煤炭采样精密度	181
7.11.5 其他一些相关关系式	137	10.2.2 煤的变异性的确定	184
习题七	138	10.2.3 采样单元数和每一采样单元 的子样数	191
第 8 章 描述性统计	139	10.2.4 样品质量	194
8.1 调查表	139	10.3 煤炭采样的方法	199
8.2 流程图	140	10.3.1 煤炭采样的一般原则	199
8.3 因果图	141	10.3.2 时间基采样	200
8.4 条形图	143	10.3.3 质量基采样	200
8.5 饼分图	144	10.3.4 分层随机采样	201
8.6 直方图	145	10.3.5 静止批中采样	201
8.7 排列图	148	10.4 采样精密度的核​​验	203
8.8 散点图	150	10.4.1 概述	203
8.9 趋势图	151	10.4.2 双倍子样数的双份采样方法 ...	203
习题八	153	10.4.3 在常规采样中的双份采样 ...	205
第 9 章 煤炭实验室质量管理体系与 质量控制	154	10.4.4 精密度调整程序	205
9.1 实验室质量管理体系	154	10.4.5 特定批煤精密度的测定 (多份采样法)	205
9.1.1 统计技术是质量管理体系的 基础	154	习题十	206
9.1.2 六西格玛管理模式	159	第 11 章 机械采样设备的性能试验	208
9.2 实验室质量控制	162	11.1 水分损失率	208
9.2.1 能力验证	162	11.2 破碎率试验	210
9.2.2 质量控制中标准煤样的应用 ...	171	11.3 采样量(子样质量)检查	210
9.2.3 利用 QC 样品进行质量控制 ...	173	11.3.1 概述	210
习题九	176	11.3.2 子样量的变异系数	210
		11.3.3 子样量和流速的相关性	212
		11.4 精密度核​​验	213
		11.4.1 概述	213
		11.4.2 精密度的核​​验方法	214
		11.5 偏倚试验	215
		11.5.1 概述	215
		11.5.2 统计分析及说明	215
		11.5.3 偏倚试验计算示例	221
		习题十一	227

第 12 章 统计技术在煤炭制样中的应用

12.1 煤炭制样的原理和工序	228
12.1.1 制样原理	228
12.1.2 制样工序	229
12.2 煤炭制样精密度	229
12.2.1 概述	229
12.2.2 公式的推导与应用	230
12.3 制样和化验方差	230
12.3.1 概述	230
12.3.2 制样和化验方差的直接测定	231
12.4 制样和化验精密度的核验	232
12.4.1 整体的核验程序	232
12.4.2 分阶段核验	234
12.4.3 举例	239
12.5 制样程序(设备)偏倚试验	242
12.5.1 试验方法	242
12.5.2 参比样品	242
习题十二	242

第 13 章 统计技术在煤质分析中的应用

13.1 煤质分析的一般规定	243
13.1.1 测试次数的规定及理论依据	243
13.1.2 煤炭的基与基的换算	244
13.1.3 结果报告	247
13.1.4 方法精密度	247
13.2 准确度试验	247
13.2.1 概述	247
13.2.2 正确度估计(偏倚试验)	248
13.2.3 精密度估计	251
13.2.4 示例	252
13.3 实验室检验结果准确度的检查	257
13.3.1 使用密码标准煤样	257
13.3.2 实验室检验结果的内部审核	258
13.3.3 使用存查样品	258
13.4 煤质分析结果的审查	258
13.4.1 明显错误数据的检查	258
13.4.2 利用回归分析进行试验结果	

的审核 260

习题十三 260

第 14 章 统计技术在煤炭检验科研中的应用

14.1 煤炭检验标准的制定	261
14.1.1 基本模型与参数估计	262
14.1.2 精密度试验的统计分析	262
14.1.3 精密度数值的表示	263
14.1.4 测试方法精密度的确定	264
14.1.5 正确度和精密度数值的实际应用	266
14.1.6 检验方法的确认	267
14.2 标准样品的制备	268
14.2.1 定值的一般原则	269
14.2.2 一个实验室的试验定值	269
14.2.3 多个实验室协作试验定值	270
14.2.4 均匀性检验	272
14.2.5 稳定性检验	274
14.3 煤炭检验仪器设备的研制	274
14.3.1 准确度试验	275
14.3.2 型式检验的抽样和判定规则	278
习题十四	283

附录 习题参考答案 284

附表 291

附表 1 正态分布函数表 291

附表 2 计量值控制图系数表 292

附表 3 χ^2 分布的分位数表 293

附表 4 t 分布的分位数表 294

附表 5 F 分布分位数表 295

附表 6 格拉布斯检验法的临界值表 298

附表 7(a) 狄克逊检验法的临界值表 299

附表 7(b) 双侧狄克逊检验法的临界值表 300

附表 8 科克伦检验法的临界值表(单侧检验) 301

附表 9 常用的正交试验表 302

附表 10 相关系数检验表 303

附表 11 随机数表 304

参考文献 306

第 1 章 统计技术的基础知识

1.1 统计技术概述

统计就是有目的地收集数据,整理数据,并使用相应的方法制图、列表与分析数据的过程。对于煤炭检验而言,必须以数据为依据,作出对煤炭质量的评判。所以就需收集数据、整理数据以及分析数据,因此,煤炭检验也包括了大量的统计工作。

统计技术是以概率论为理论基础,研究随机现象中确定的数学规律的学科。是应用数学的一个分支。统计技术是一个大的概念,是指整个学科而言,指的是一门技术的总概括。

煤炭检验的本身就是抽取较小的样本来推断整批煤质量特性的过程,由于真值永远是未知的,所以只能以一定的概率来评估质量特性的真值。因此,煤炭检验也必须使用以概率论为理论基础的统计技术。

统计方法是指统计技术中的具体方法。如排列图是统计技术中的一种方法,显著性检验是统计技术中的一种方法等等。原则上不应称排列图、显著性检验等为统计技术。

统计工具是指简化的统计方法。统计工具的开发是日本质量管理专家在质量管理工作中的一大贡献。统计技术的理论基础是概率论,但概率论这一理论对于基层的工人甚至对于一般的技术人员来说也是难以理解的,因此也妨碍了统计技术的推广应用。为此,针对基层工人和初级技术人员的特点,上世纪 60 年代日本质量管理专家开发了:因果图、排列图、调查表、散布图、控制图和分层法,称之为质量管理的常用七种工具。随着质量管理的不断深化,上世纪 70 年代日本质量管理专家又开发出系统图、关联图、矩阵图、矢线图、KJ(亲和图)法、PDPC(过程决策程序图)法和矩阵数据解析法,称之为质量管理的新七种工具。这些方法简便、易行、易懂、直观、质量改进的效果明显,故亦称质量改进工具和技术或称作描述性统计。

所谓的工具,指不讲统计方法的原理和设计,也不讲对统计结果的分析,只讲操作步骤。好比工人在生产过程中操作各种工具,如榔头、扳手、螺丝刀等,并不要求工人掌握这些工具的设计原理和制造工艺,只要能在适用的条件下得心应手地使用工具就可以了。

因此,在一些国际标准如 ISO 9000 族标准中把这些统计工具作为质量改进的工具和技术,帮助质量改进项目和活动的成功。

1.2 随机现象及其规律性

1.2.1 随机现象

所谓的现象是指事物的表现,任何现象都不会无缘故地发生,在一定条件下导致结果。

根据条件与结果的关系,现象可分为确定现象和随机现象两大类。

当给定一组所需的条件后,所发生的结果是不确定的(随机的)现象称为随机现象。

当给定一组所需要的条件后,所发生的结果是必然的(确定的)现象称为确定现象(必然现象)。如水在一个标准大气压、 0°C 时,必然结冰。煤样在马弗炉中升温至 815°C 时一定形成灰分。这些都属于确定现象。

而随机现象结果是不确定的,例如煤样在 900°C 的马弗炉中,隔绝空气加热 7 min 所得的挥发分结果无法预测。这是一些不确定随机因素影响的结果,是一种随机现象。

1.2.2 随机现象的规律性

检验结果虽然是一种随机现象,但并不是漫无边际的随机,而是在一定范围内符合一定数学规律的随机。

检验结果的随机性造成质量特性(试验数据)的波动,而表现为质量特性值(试验数据)的分布。

统计技术就是研究随机现象中确定的数学规律的学科。煤炭检验是以抽样推断总体(批)的质量,其检验结果随煤的质量变异性、试验误差的随机性而波动,并服从一定的分布。因此煤炭的检验结果属于一种随机现象。

1.3 统计数据及统计方法的分类

1.3.1 统计数据的分类

数据是统计的对象,人们习惯上把由数字组成的数字数据称为数据。根据数据的性质,数据又可分为计量值数据和计数值数据。计数值数据又可分为计件值数据和计点值数据。

(1) 计量值数据

计量值数据是指可以连续取值,在有限的区间内可以无限取值的数据。如长度、面积、体积、质量、密度、电压、电流、强度、纯度等,煤炭检验中大部分质量特性的数值都属于计量值数据。计量值数据服从正态分布。

(2) 计数值数据

计数值数据是指只能间断取值,在有限的区间内只能取有限数值的数据。计数值数据就是通常所讲的数数,如数一数运煤列车的车皮数,载煤汽车的数目,不合格产品的件数等。所以计数值数据,是以正整数(自然数)的方式表现。计数值数据分为以下两类:

① 计件值数据:数产品(或其他物件)的件数而得到的数值。计件值数据服从二项分布。

② 计点值数据:数产品缺陷数而得到的数值。计点值数据服从泊松分布。

1.3.2 统计方法的分类

统计数据是统计方法应用中的统计对象,根据统计数据类别不同,统计方法可分为统计型方法(适用于数字数据的方法)和情理型方法(适用于非数字数据的方法)。表 1-1 总结了两类不同方法的特点。

表 1-1 两类统计方法的比较

方法类别 比较项目	统计型(数字数据)方法	情理型(非数字数据)方法
收集对象	收集大量定量描述的数字数据,收集方法:取样、测试、计算、记录	收集大量定性描述的非数字数据。收集方法:调查、研究
处理方式	对数字数据进行统计计算,取得反映客观规律的特性值	对非数字数据(语言资料)进行分类、归纳、整理,得到有条理的思路
功能	实施统计推断(预测)及统计控制	作为决策依据
方法类型	常见统计方法中大部分属于统计型方法,如控制图、散布图、直方图、试验设计、方差分析、回归分析等	常见统计方法中少数为情理型方法如因果图、分层图、流程图、树图、水平对比法、头脑风暴法等

(1) 统计型方法(适于数字数据的方法)

① 收集的对象:统计型方法收集的是大量可以定量描述的数字数据。收集数据是通过取样、测试、计算和记录而实现的。

② 特征值:统计型方法应用一定的数学公式,对取得的数字数据进行统计计算,得到反映客观规律的特征值。数字数据分布的特征值反映了产品的质量波动状况,所以统计型方法常用于过程的统计控制和统计推断。

③ 用途:在常用的统计方法中统计型方法占有较大的比重。如描述性统计中的排列图、直方图、散布图、控制图、趋势图等以及显著性检验、方差分析、试验设计、过程能力分析等。在煤炭检验中我们所用的统计方法大都是统计型方法。

(2) 情理型方法(适于非数字数据的方法)

① 收集的对象:情理型方法收集的是大量定性描述的非数字数据,非数字数据需通过广泛、深入的调查研究和观察获得。

② 数据的处理:非数字数据的处理又称为资料处理。即把复杂事物的多方面信息收集起来,利用其相互内在的联系,加以整理,通过分层、分类、归纳、总结,得到解决问题的思路。

③ 实现的功能:通过情理型方法的应用,所得到的有条理的思路,可作为决策过程的输入(决策依据)。情理型方法多应用于描述性统计方面,如因果图、流程图、分层图、关联图、水平对比法、头脑风暴法、KJ法、PDPC法等。在煤炭检验中除描述性统计外,其他较少应用。

1.4 随机变量及其分布

1.4.1 随机变量

量可分为常量和变量。常量即是取固定数值的量,如圆周率 $\pi=3.141\ 6$,自然常数 $e=2.718$ 等。变量是数值可以变化的量,它是相对于常量而言的。

随机变量是指取值随试验结果而定,且有一定的概率分布的变量。

随机变量是变量中的一种类型,随机变量的数值是由随机因素的作用而发生的。随机变量具有以下特点:

① 随机性:即在试验前只知道它的可能的取值范围,而不能预知它具体取什么值。

② 统计规律性:由于试验的各个结果的出现有一定概率,于是随机变量取值也有一定的概率。

③ 互斥性:随机变量所取的各个可能值是互斥的,取某个数值后,不能再取其他值。

最常见的随机变量有两大类,即离散型随机变量和连续型随机变量。如果一个随机变量只可能取数轴上有限个或可数个孤立的值,并且对应于这些有确定的概率,则称这个随机变量为离散型随机变量。离散型随机变量的取值是计数值数据。如果一个随机变量所有可能取的值是充满某一区间,甚至整个数轴,就称为连续型随机变量。连续型随机变量的取值是计量值数据。

1.4.2 随机变量的概率分布

1.4.2.1 概率分布的有关概念

(1) 概率

度量一随机事件发生可能性大小的实数,其值介于0与1之间。

注:随机事件即是在一定的条件下,可能发生也可能不发生的事件称为随机事件,简称事件。

例如,投钱币时,蒲丰(Buffon J. L.)投过4 040次,得2 048次正面,出现正面频率为 $2\,048/4\,040=0.506\,9$;皮尔逊(Pearson K.)投过24 000次,得到12 012次正面,出现的频率为 $12\,012/24\,000=0.500\,5$,频率稳定到常数0.5,故出现的正面概率为 $0.5=1/2$ 。

事件越容易发生,其出现的概率越大。事件越不易发生,其出现的概率越小。必然事件出现的概率为1,不可能事件出现的概率为0。

(2) 随机变量的概率分布

给出一个随机变量取任何给定值或取值于任何给定集合的概率的函数。

注:随机变量在全部集合取值的概率为1。概率分布简称分布。

(3) 分布函数

随机变量 X 小于或等于实数 x 的概率,它是 x 的函数:

$$F(x) = P(X \leq x)$$

分布函数的意义就是: X 是一个随机变量, x 是任意实数,那么“ X 取小于(或小于或等于) x 之值”是这个事件的概率。

(4) 连续随机变量及[概率]密度函数

如果随机变量的分布函数 $F(x)$ 可表示为一非负函数 $f(x)$ 的积分:

$$F(x) = \int_{-\infty}^x f(t) dt$$

则称该随机变量为连续随机变量。 $f(x)$ 称为它的[概率]密度函数。

任何密度函数 $f(x)$ 必满足

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

注:连续随机变量的概率分布称为连续分布。

(5) 独立

若两组随机变量有联合概率分布,其任何一组的条件分布都不随另一组的取值而变化,则称它们是独立的。否则称它们是相依的。

在煤炭检验中,广泛应用独立这一术语。

两次测试独立(两次重复测定)的意思是它们不相依,即它们不相互依赖,一次测试结果的取值与另一次测试结果取值无关。

独立的对立术语是相依,相依为不独立,在煤炭检验中如统计推断、回归分析、方差分析、正交试验、协同试验、能力验证、不确定度评定等统计技术应用时,都要求测试必须相互独立,否则以上的试验结果有效性难以令人信服。如若在能力验证或协同试验中,一个实验室的结果,参考了另一个实验室的结果,本来应为相互独立的试验变成了相依的试验,这样就不可能体现随机变量的概率分布的实际状况,因此试验评定的结果必然是不准确或完全是错误的。

(6) 期望

对以概率 P_i 取值 x_i 的随机变量 X ,其期望的定义为:

$$E(X) = \sum P_i x_i$$

式中 $\sum$ 是对 X 的所有可能的值 x_i 求和。

对密度函数为 $f(x)$ 的连续随机变量 X ,其期望的定义为:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

随机变量的期望也称为它的概率分布的期望。同义词:均值(mean)、期望值或数学期望。

期望是随机变量的最重要的位置特征,它表示随机变量的平均大小和位置,期望越大,随机变量的平均愈大,位置愈靠数轴的右边。它与算术平均值在概念上有严格的区别,期望是对随机变量或分布或总体而言,而算术平均值是对样本而言。样本算术平均值对期望来说是一个无偏、一致、有效的估计值。

(7) 方差与标准差

方差:随机变量 X 的方差定义为:

$$V(X) = E[(X - E(X))^2]$$

它亦称为 X 的分布方差。

标准差:方差的正平方根 $\sqrt{V(X)}$ 。

标准差是重要的分散度量,它是方差的正平方根。标准差也称标准偏差,又称均方根,因为它是偏差 $X - E(X)$ 的平方平均再开方。

偏差 $X - E(X)$ 有正有负,但平方后恒正,再平均开方后取正值,它是各偏差的均方根值,代表了各个偏差而不计其正负,且能将绝对值大的偏差显露出来。

标准差愈大,随机变量愈分散,标准差愈小,随机变量愈集中。

(8) 分位数

对随机变量 X ,满足条件 $P(X < x_p) \leq p$ 和 $P(X \leq x_p) \geq p$ 的实数 x_p 称 X 的或其分布的 p 分位数。

注:① p 分位数可以不唯一。

② $x_{0.75}$ 与 $x_{0.25}$ 分别称为上、下四分位数。

分位数中比较重要的有以下两个:

中位数 $x_{0.5}$:常用作集中性指标。

$x_{0.75}, x_{0.25}$: 在 $[x_{0.75}, x_{0.25}]$ 区间内有这组数据的一半, 而且 $x_{0.75} - x_{0.25}$ 是 50% 数据的极差, 所以也是一个离散性度量的指标。

为做假设检验, 常需临界值, 将它和统计量比较, 以作出假设为真或非真的判断, 临界值就是分位数。取定一个小概率 α (如 $\alpha=0.05, 0.01$ 等), x_α 或 $x_{1-\alpha}$ 常用作单侧假设检验的临界值, 在生物、医疗中常使用单侧检验。 $x_{\alpha/2}$ 或 $x_{1-\alpha/2}$ 常用作双侧假设检验的临界值, 在煤炭检验中较多的使用双侧检验, 但在 ISO 13909《硬煤和焦炭——机械采样》中的机械设备偏倚试验双侧和单侧检验都要用到。

对不同随机变量, x_p 常列于统计技术表或附于专业书刊后, 这样由 p 可查出 x_p 值。本书的后面也附有一些我们日常常用的分位数表。

(9) 统计技术应用中的两类错误

既然统计技术是以概率论为理论基础, 事件的发生都有一定的概率, 都有一定的随机性, 特别是在煤炭检验中, 都是以样本来推断总体, 因此统计方法在应用过程中, 都会犯两类不同性质的错误, 第一类错误称之为弃真的错误, 即以真当假; 第二类错误称之为取伪的错误, 即以假当真。任何统计方法的应用都不能幸免。这两类错误在后面的统计推断一章中还要详细地论述, 在此不再重复。

1.4.2.2 随机变量的概率分布

离散型随机变量的概率分布主要有二点分布(0-1 分布), 二项分布、负二项分布、泊松分布、超几何分布、多项分布等。在煤炭检验中应用最多的是二项分布和泊松分布。

连续型随机变量的概率分布主要有均匀分布、三角分布、正态分布、标准正态分布、B 分布、对数正态分布、指数分布、 Γ 分布、贡贝尔分布(I 型极值分布)、弗莱契分布(II 型极值分布)、威布尔分布(III 型极值分布)、二维正态分布等, 在不同的行业、不同的领域中, 随机变量有不同的出现形式, 它们用不同的概率分布来描述, 故有多种分布及应用, 如在不确定度中常用均匀分布和三角分布; 计数抽样中常用二项分布和泊松分布; 产品分析常用超几何分布; 放射性测定常用泊松分布; 可靠性寿命试验用指数分布; 极值分析中用极值分布; B 分布、 Γ 分布常用于寿命研究及准确度研究等。它们在概率分析中都很重要。

在日常煤炭检验中, 最常用的则是正态分布和由正态总体导出的统计量分布也称抽样分布(正态分布随机变量的样本函数)的 χ^2 分布、 t 分布、 F 分布。它们在煤炭测试中有着广泛的应用, 并且是概率分布的基础。

1.4.3 离散型随机变量的分布

这里主要介绍在煤炭检验中比较常用的二项分布和泊松分布。

1.4.3.1 二项分布

二项分布是一种离散概率分布, 其概率函数为

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, x=0, 1, \dots, n$$

$$\text{式中 } 0 < P < 1, \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\text{此式也可写作: } P(X=x) = C_n^x p^x (1-p)^{n-x}, x=0, 1, \dots, n$$

在质量检验中, 常常遇到这样的一些情况, 其检验所得的两种结果是互斥的。比如产品的合格与不合格; 细菌检验中出现的阳性与阴性反应等等。类似这种互斥事件发生的概率

分布便遵循二项分布。二项分布大量地用在抽样检验中,特别是计件值数据的抽样中,如在煤炭检验仪器的研制时,需进行抽取样机进行型式检验,其所采用的 GB/T 2828.1 计数检验标准中常用的概率分布就是二项分布。而在抽样验收中连续批的 OC 曲线也都是按二项分布进行设计的。如 GB/T 13262 的抽样表全部是根据二项分布设计的。

下面举例说明二项分布在抽样中的应用。若产品的不合格率为 p ,那么它的合格品率则为 $1-p$,从中随机抽取 n 件样品,那么 x 件不合格品的概率 $P(X=x)$ 为多少?由于 n 件样品中有 x 件不合格品,总共有 C_n^x 种可能的组合,而由于每件组合的概率都是 $p^x(1-p)^{n-x}$,所以 n 件样品中出现 x 件不合格品的概率就可以用以上公式方便地计算出,请看下面的具体示例。

示例 1-1:已知某工厂生产的仪器的二等品率为 40%,若从中随机抽取 15 台仪器进行测试,求 10 台仪器为二等品的概率是多少?

解: $P(X=10) = C_{15}^{10} \times 0.4^{10} \times (1-0.4)^{15-10} = 0.024\ 5$

故随机抽取的 15 台仪器中有 10 台仪器是二等品的概率为 2.45%。

示例 1-2:某工厂生产的一批产品中不合格品率为 15%,重复抽取样品检查,共抽取 4 个样品,试求 4 个样品中不合格品数的概率分布?

解:4 个样品中全部是合格品的概率:

$$P(X=0) = C_4^0 (0.15)^0 \cdot (1-0.15)^4 = 0.522\ 0$$

4 个样品中有 1 个不合格品的概率:

$$P(X=1) = C_4^1 (0.15)^1 \cdot (1-0.15)^3 = 0.368\ 5$$

4 个样品中有 2 个不合格品的概率:

$$P(X=2) = C_4^2 (0.15)^2 \cdot (1-0.15)^2 = 0.097\ 5$$

4 个样品中有 3 个不合格品的概率:

$$P(X=3) = C_4^3 (0.15)^3 \cdot (1-0.15)^1 = 0.011\ 5$$

4 个样品中全是不合格品的概率:

$$P(X=4) = C_4^4 (0.15)^4 \cdot (1-0.15)^0 = 0.000\ 5$$

以上的计算还可以通过二项分布数值表直接查出,二项分布表可查阅 GB/T 4086.5 《统计分布数值表 二项分布》。注意示例 1-2 中的重复抽样检查,意思就是有放回的抽样。

1.4.3.2 泊松分布

泊松分布 一种离散概率分布,其概率函数为

$$P(X=x) = e^{-\lambda} \frac{\lambda^x}{x!}, x=0,1,2,\dots$$

式中: $\lambda > 0$ 。

泊松分布的期望和方差都是 λ 。

当 n 大 p 小,且 $np=\lambda$ 时,二项分布可用泊松分布近似。

其中 $e=2.718\ 28\dots$ 是自然对数的底数。

泊松分布与二项分布在产品质量验收及质量控制的实际应用中,都是非常重要的离散随机变量的概率分布。尤其是泊松分布更为重要,特别适于计点值数据的抽样。在 GB/T 2828.1 等一些计数抽样标准中的数学模式,其中就有泊松分布。这是因为在二项分布中,当 n 很大时: