|  |
| --- |
| **戈塞特与t分布** |
| . |
| |  | | --- | | Francis Galton  　　戈塞特(William Sealey Gosset)，英国统计学家。出生于英国肯特郡坎特伯雷市，求学于曼彻斯特学院和牛津大学，主要学习化学和数学。1899年，戈塞特进入都柏林的A.吉尼斯父子酿酒厂，在那里可得到一大堆有关酿造方法、原料（大麦等）特性和成品质量之间的关系的统计数据。提高大麦质量的重要性最终促使他研究农田试验计划，并于1904年写成第一篇报告《误差法则应用》。 　　戈塞特是英国现代统计方法发展的先驱，由他导出的统计学T检验广泛运用于小样本平均数之间的差别测试。他曾在伦敦大学K.皮尔逊生物统计学验室从事研究（1906-1907），对统计理论的最显著贡献是《平均数的机误》（1908）。这篇论文阐明，如果是小样本，那么平均数比例对其标准误差的分布不遵循正态曲线。由于吉尼斯酿酒厂的规定禁止戈塞特发表关于酿酒过程变化性的研究成果，因此戈塞特不得不于1908年以“学生”的笔名发表他的论文，导致该统计被称为“学生”的笔名发表他的论文，导致该统计被称为“学生的T检验”。1907-1937年间，戈塞特发表了22篇统计学论文，这些论文于1942年以《“学生”论文集》为书名重新发行。 　　戈塞特是英国现代统计方法发展的先驱，由他导出的统计学T检验广泛运用于小样本平均数之间的差别测试。他曾在伦敦大学K.皮尔逊生物统计学验室从事研究（1906-1907），对统计理论的最显著贡献是《平均数的机误》（1908）。这篇论文阐明，如果是小样本，那么平均数比例对其标准误差的分布不遵循正态曲线。由于吉尼斯酿酒厂的规定禁止戈塞特发表关于酿酒过程变化性的研究成果，因此戈塞特不得不于1908年，Gosset首次以“学生”(Student)为笔名，在《生物计量学》杂志上发表了“平均数的概率误差”。Gosset在文章中使用Z統計量來檢驗常態分配母群的平均數。由于这篇文章提供了“学生t检验”的基础，为此，许多统计学家把1908年看作是统计推断理论发展史上的里程碑。后来，哥塞特又连续发表了“相关系数的概率误差”(1909)、“非随机抽样的样本平均数分布”(1909)、“从无限总体随机抽样平均数的概率估算表”(1917)，等等。他在这些论文中，第一，比较了平均误差与标准误差的两种计算方法；第二，研究了泊松分布应用中的样本误差问题；第三，建立了相关系数的抽样分布；第四，导入了“学生”分布，即t分布。这些论文的完成，为“小样本理论”奠定了基础；同时，也为以后的样本资料的统计分析与解释开创了一条崭新的路子。由于哥塞特开创的理论使统计学开始由大样本向小样本、由描述向推断发展，因此，有人把哥塞特推崇为推断统计学的先驱者。 　　William Sealey Grosset在20世纪前三十余年是统计界的活跃人物，他的成就不限于《均》文.同年他发表了在总体相关系数为0时，二元正态样本相关系数的精确分布，这是关于正态样本相关系数的第1个小样本结。 　　他对回归和试验设计方面也有相当的研究，在与费歇尔的通信中时常讨论到这些问题。费歇尔很尊重他的意见，常把自己工作的抽印本送给William Sealey Grosset请他指教.在当时，能受到费歇尔如此看待的学者为数不多。 　　 William Sealey Grosset的一些思想，对他日后与奈曼合作建立其假设检验理论有着启发性的影响.他说（引自《耐曼&frac34;现代统计学家》）：“我认为现在统计学界中有非常多的成就都应归功于William Sealey Grosset…….” 　　戈塞特是小样本统计理论的开创者.戈塞特在酿酒公司工作中发现，供酿酒的每批麦子质量相差很大，而同一批麦子中能抽样供试验的麦子又很少，每批样本在不同的温度下做实验，其结果相差很大.这样一来，实际上取得的麦子样本，不可能是大样本，只能是小样本.可是，从小样本来分析数据是否可靠？误差有多大？小样本理论就在这样的背景下应运而生。1905年，戈塞特利用酒厂里大量的小样本数据写了第一篇论文《误差法则在酿酒过程中的应用》，在此基础上，1907年戈塞特决心把小样本和大样本之间的差别搞清楚.为此，他试图把一个总体中的所有小样本的平均数的分布刻画出来.做法是，在一个大容器里放了一批纸牌，把它们弄乱，随机地抽若干张，对这一样本做实验记录观察值，然后再把纸牌弄乱，抽出几张，对相应的样本再做实验观察，记录观察值.大量地记录这种随机抽样的小样本观察值，就可借以获得小样本观察值的分布函数.若观察值是平均数，戈塞特把它叫做t分布函数。1908年，戈塞特以“学生（Student）”为笔名在《生物计量学》杂志发表了论文《平均数的规律误差》。这篇论文开创了小样本统计理论的先河，为研究样本分布理论奠定了重要基础.被统计学家誉为统计推断理论发展史上的里程碑.戈塞特这项成果，不仅不再依靠近似计算，而且能用所谓小样本来进行推断，并且还成为使统计学的对象由集团现象转变为随机现象的转机.换句话说，总体应理解为含有未知参数的概率分布（总体分布）所定义的概率空间；要根据样本来推断总体，还必须强调样本要从总体中随机地抽取，也就说，一定要是随机样本.但是，应该指出：戈塞特推导t分布的方法是极不完整的，后来费希尔利用n维几何方法给出了完整的证明。戈塞特在其论著中，引入了均值、方差、方差分析、样本等概率、统计的一些基本概念和术语。1907-1937年间，戈塞特发表了22篇统计学论文，这些论文于1942年以《“学生”论文集》为书名重新发行。 | |