

# 方差分析与回归模型结合在多变量研究中的应用方法

马保忠

昆明城市学院 云南昆明 650000

**摘要:** 本研究旨在探讨方差分析与回归模型结合在多变量研究中的应用价值。传统方法往往难以同时揭示分类变量和连续变量的综合作用,而联合模型能够在同一框架下实现组间差异检验与连续变量效应量化。本文以某医院糖尿病科30例临床试验数据为案例,比较三种降糖药物(阿卡波糖、二甲双胍、利拉鲁肽)的疗效,并考察患者体重指数(BMI)的调节作用。结果表明,药物类型对血糖降低幅度存在显著差异,其中利拉鲁肽疗效最优,二甲双胍在高BMI患者中效果提升更为显著;同时,药物与BMI之间存在交互作用。联合模型的拟合度显著优于单一方法,能够更全面地揭示复杂变量关系。研究表明,方差分析与回归模型的结合不仅具有较强的解释力和预测力,也为医学和社会科学等领域的个体化研究提供了科学工具。

**关键词:** 方差分析; 回归模型; 多变量研究; 药物疗效; 体重指数(BMI)

## 引言

在多变量研究中,方差分析和回归分析是常用的两类统计方法。前者强调比较不同组别之间的均值差异,后者侧重揭示变量之间的定量关系。然而,在实际应用中,单一方法往往难以全面反映复杂因素的综合作用。例如,在医学领域,研究者既需要比较不同药物疗效的差异,又必须考察患者个体特征对疗效的影响,仅依赖方差分析或回归分析都存在局限。

近年来,学界开始关注将方差分析与回归模型结合的思路。该方法通过在同一框架下引入分类变量和连续变量,不仅能够检验组间差异,还能量化连续变量及其交互效应的影响,从而显著提升模型的解释力与预测力。

基于这一背景,本文以某医院糖尿病科临床试验数据为案例,探讨方差分析与回归模型结合的应用方法。通过分析3种降糖药物的疗效差异,并考察患者体重指数(BMI)的调节作用,验证了该方法在揭示复杂变量关系和服务个体化治疗方面的价值。本文的研究不仅丰富了多变量统计方法的应用路径,也为医学及相关领域的实证研究提供了新的工具和参考。

## 1 方差分析(ANOVA)概述

方差分析(Analysis of Variance, ANOVA)是一种用于分析不同组之间均值差异是否显著的统计方法。其主要目标是通过比较不同组之间的变异与组内变异,判断不同因素对实验结果的影响是否显著。在多变量研究中,方差分析能够帮助研究者识别哪些因素对研究对象的影响最为重要,并提供一种有效的统计手段来检验不同组之间的差异性。方差分析广泛应用于各

类实验设计中,如药物效果、教育评估、市场研究等领域。

## 2 回归分析概述

回归分析是一种用于研究自变量(解释变量)与因变量(响应变量)之间定量关系的统计方法。回归分析可以帮助研究者建立数学模型,用于预测因变量的值,或者揭示自变量对因变量的影响程度。在多变量研究中,回归分析不仅可以用于单一自变量的分析,还可以处理多个自变量的影响,构建多元回归模型。回归分析的广泛应用涵盖了经济学、社会学、医学等领域,尤其在政策评估、市场趋势预测以及健康数据分析中发挥着重要作用。

回归分析的基本思想是通过数据拟合找到最佳的模型,使得模型能够尽可能精确地描述变量之间的关系。线性回归分析是回归分析中最为基础和常见的方法,它假设因变量与自变量之间的关系是线性的,即可以用一条直线来表示。

## 3 方差分析与回归模型结合的理论基础

### 3.1 方差分析与回归模型的区别与联系

方差分析和回归模型是统计学中常用的两类方法,但它们在分析思路和应用目标上有所不同。方差分析主要用于检验不同组间均值差异是否显著,而回归模型更关注变量之间的定量关系和预测能力。从表面上看,这两种方法似乎各自独立,但实际上,它们有着紧密的内在联系。方差分析可以被视为回归分析的特例,在某些情况下,方差分析的计算可以通过回归分析来实现。例如,在单因素方差分析中,将因变量与组别变量(分类变量)之间的关系建模,可以

用一个线性回归模型表示，而组别变量通过虚拟变量（Dummy Variables）引入到模型中。通过这种方式，回归模型可以实现对组间差异的检验。

这种联系表明，在实际研究中，方差分析和回归模型可以相互补充。方差分析提供了组间差异的整体评价，而回归模型能够进一步量化因素的影响大小，甚至预测未来的结果。在多变量研究中，结合两种方法能够更全面地解析复杂数据结构，为研究者提供更深层次的见解。

### 3.2 联合模型的构建与优势

将方差分析与回归模型结合的联合模型是一种既能够检验组间差异，又能够量化和预测变量间关系的综合性统计方法。这种模型的核心思路是将分类变量和连续变量同时纳入分析框架，通过引入交互项或协变量，既分析主效应，也揭示复杂的交互效应。例如，在多因素实验中，不仅需要检验组间差异，还需要控制混杂变量或探讨多个自变量的交互作用，这时联合模型能够发挥重要作用。

构建联合模型的一般步骤如下：

1) 确定因变量和自变量，其中因变量通常是连续型变量，自变量既可以是分类变量（对应方差分析）也可以是连续变量（对应回归分析）。

2) 将分类变量转化为虚拟变量，以便与回归模型兼容。例如，对于一个三组实验，可以用两个虚拟变量来表示组别信息。

3) 在回归模型中加入连续变量和虚拟变量，并通过交互项表达复杂的因果关系。

4) 使用联合模型检验主效应、交互效应和协变量对因变量的影响，具体包括拟合度检验（如  $R^2$ ）、残差分析和模型显著性检验（如 F 检验和 t 检验）。

这种联合模型的优势在于，它克服了单一方法的局限性。首先，与单独的方差分析相比，联合模型能够进一步量化组间差异的大小，而不仅仅停留在是否显著的层面；其次，与传统的回归模型相比，它能够处理多组比较和分类变量。此外，联合模型还提供了灵活性，例如可以在模型中引入协变量以控制混杂因素，从而使结果更加可靠。

## 4 方差分析与回归模型结合的应用案例分析

### 4.1 案例背景与数据介绍

本研究基于某医院糖尿病科的临床试验数据，分析三种降糖药物（阿卡波糖、二甲双胍、利拉鲁肽）对糖尿病患者血糖改善效果的影响，同时探讨患者体重指数（BMI）对疗效的调节作用。研究目标是评估

3 种药物的疗效差异，并明确 BMI 如何影响药物的降糖效果。

数据包含 30 名患者的治疗信息，记录了药物类型、BMI 以及治疗后空腹血糖（FPG）的降低百分比（相较于治疗前）。

### 4.2 方差分析与回归模型的应用实施

在分析过程中，首先采用单因素方差分析检验不同药物对血糖降低幅度的影响。结果表明，三组药物间疗效差异显著，其中阿卡波糖组平均降幅最低，二甲双胍组居中，而利拉鲁肽组表现最佳。为了进一步探讨体重指数（BMI）的作用及其与药物类型的交互效应，研究构建了联合回归模型，将药物类别转化为虚拟变量，并引入 BMI 及其交互项。模型结果显示，药物类型主效应显著，利拉鲁肽和二甲双胍的整体疗效均明显优于阿卡波糖；BMI 对疗效具有正向影响，即患者 BMI 越高，血糖降低幅度越大。更为重要的是，药物与 BMI 之间存在交互作用：二甲双胍疗效随 BMI 增加而提升更明显，而利拉鲁肽虽然整体疗效最高，但其受 BMI 变化的影响相对较小。该发现说明，在临床实践中，二甲双胍可能更适合 BMI 较高的患者，而利拉鲁肽则在不同 BMI 水平下均保持较优的疗效。整体来看，联合模型较单一方差分析或回归分析更具解释力，不仅揭示了药物之间的差异，还阐明了连续变量对疗效的调节效应，为个体化治疗提供了更科学的依据。

### 4.3 案例结果分析与讨论

通过方差分析和回归模型的联合应用，本研究揭示了药物类型和 BMI 对血糖降低效果的综合影响。结果表明，利拉鲁肽在所有 BMI 水平下疗效较好，但其疗效对 BMI 的依赖性较小；二甲双胍次之，其疗效随 BMI 增加有更显著提升；阿卡波糖的疗效较低，且对 BMI 的依赖性不明显。

联合模型的调整后  $R^2=0.91$ ，表明模型拟合较好，能够有效解释血糖降低的主要影响因素。相比单一的方差分析或回归分析，联合模型不仅揭示了药物之间的疗效差异，还通过交互效应分析了 BMI 对疗效的调节作用。这一结果为糖尿病患者的个性化药物选择提供了科学依据。

## 5 方差分析与回归模型结合中的挑战与改进方向

尽管方差分析与回归模型地结合在多变量研究中展现了较强的解释力，但在实际应用中仍面临若干挑战。首先，在模型构建环节，如何兼顾分类变量与连续变量的兼容性是一个核心难题。虚拟变量的引入虽

然能实现两类变量的统一建模,但在分类变量数量较多或交互作用复杂时,模型规模容易膨胀,降低可解释性。尤其是在医学和社会科学研究中,多重交互项往往导致参数估计不稳定,增加结果解读的难度。临床或社会调查数据常常存在缺失值、异常值或分布不均衡等情况,这不仅影响方差分析对组间差异的检验效能,也削弱回归分析的稳定性。此外,当自变量之间存在高度相关时,模型会出现多重共线性,导致参数显著性检验失真。如何在建模前进行合理的数据预处理与变量筛选,成为提升模型稳健性的关键。方差分析与回归模型均依赖于正态性、独立性和方差齐性等统计假设,而这些条件在真实数据中往往难以完全满足。残差分布不符合正态、组间差异过大或样本量不足,都会削弱结果的解释力。此外,模型在外推应用时的可靠性有限,如果研究对象的范围过窄,将结论推广到更广泛人群可能导致偏差。

在推广与应用层面,该方法仍存在局限性。其一,该框架更适用于连续型因变量,对分类型或计数型结果变量需要扩展为广义线性模型。其二,模型复杂性较高,非专业人员往往难以理解和使用。为此,未来改进方向主要包括三个方面:一是开发模块化、易操作的软件工具,降低使用门槛;二是结合机器学习与大数据方法,提高对非线性关系与高维数据的处理能力;三是推动跨学科应用,通过统计学与医学、经济学、社会科学的融合,为复杂问题提供更具解释力和实践价值的解决方案。

## 6 结语

本研究结合方差分析与回归模型,探讨了在多变量研究中同时处理分类变量与连续变量的可行路径,并通过糖尿病患者临床数据进行验证。结果表明,不同药物在血糖降低效果上存在显著差异,利拉鲁肽整

体疗效较好,而二甲双胍在高 BMI 患者群体中表现出更明显的疗效提升。联合模型不仅揭示了药物类型的主效应,还阐明了 BMI 对疗效的调节作用,调整后的拟合优度较高,显示出该方法在解释复杂数据关系方面的优势。与单纯使用方差分析或回归分析相比,二者结合能够更全面地反映多因素交互影响,为医学研究和临床个性化治疗提供了更科学的统计依据。

未来研究应在三个方面深化。一是扩展数据规模与覆盖范围,涵盖不同人群和多中心样本,以提高模型的普适性和外推性。二是引入更多维度的变量和动态数据,如基因信息、长期疗效变化等,构建更高维度的联合模型,以捕捉更复杂的规律。三是推动与新兴方法的融合,包括机器学习与大数据分析,以提升对非线性关系和高维变量的处理能力。同时,应开发简便易用的软件工具,降低模型应用的技术门槛,促进其在医学、经济学和社会科学等多领域的推广与普及。通过这些改进,方差分析与回归模型的结合将展现出更强的解释力和应用价值。

## 参考文献:

- [1] 黄伯强,李启才.带交互作用的双因素方差分析的线性回归建模[J].统计与决策,2021,37(1):10-15.
- [2] 傅莺莺,田振坤,李裕梅.方差分析的回归解读与假设检验[J].统计与决策,2019,35(8):77-80.
- [3] 刘永辉,陈弘扬,王京.偏正态自回归模型的异常值得分检验[J].数理统计与管理,2024,43(5):847-855.
- [4] 陈卫,徐利娜,迷敏,等.线性回归模型在盐酸二甲双胍缓释片降糖效能非劣效性检验中的应用[J].现代预防医学,2010,37(3):407-409.
- [5] 汪桐含,张美钰,胡典顺.概率与统计的知识理解之单因素方差分析[J].数学通讯,2023(19):1-5.