

## 土壤中微生物含量影响因素的统计方法分析

李聪杰<sup>1</sup> 郝彦斌<sup>2</sup> 韩丛英<sup>1\*</sup>

(1. 中国科学院大学数学科学学院 北京 100049)

(2. 中国科学院大学生命科学学院 北京 100049)

**摘要:**【目的】采用多元方差分析、多元线性回归分析、通径分析等统计方法,并利用多元统计方法对土壤中微生物含量进行分析。【方法】通过研究这3种常用统计方法的原理、应用条件、适用于解决的问题等,用不同的方法分析影响土壤中微生物含量的各种因素,利用SPSS软件求解。【结果】得出不同降雨量下土壤中的微生物含量存在差异性;铵态氮含量和海拔高度与变形菌门含量呈显著的线性关系,土壤中变形菌门含量随着铵态氮含量增多而增多,且铵态氮含量对变形菌门含量所起的直接作用最大;土壤中铵态氮含量、总氮量和海拔高度与放线菌门含量呈显著的线性关系,且均为负相关;其中总氮量对土壤中放线菌门含量的变化所起的负面作用最大,但铵态氮含量对放线菌门含量所起的直接作用最大。对比描述性统计方法与多元方差分析方法可知,应用描述性统计方法只能得出降雨量是影响土壤中不同细菌群落相对含量的因素之一,而不知降雨量对微生物含量变化的影响程度,6个梯度降雨量引起地显著差异究竟是由哪种微生物引起的,方差分析方法可以解决以上的问题。对比多元线性回归分析方法与通径分析方法可知,多元线性回归分析得出的相关系数只能表示铵态氮含量、总氮量和海拔与变形菌门含量(放线菌门含量)之间关系的密切程度,但无法解释和分析这种关系的构成和来源,通径分析方法可以解决这个问题。【结论】通径分析方法在分析相关问题时要优于多元线性回归分析方法,通径分析比多元线性回归分析方法更能体现因素对变形菌门含量(放线菌门含量)的直接影响和间接影响,分析结论更直观、更能说明问题。

**关键词:** 方差分析,多元线性回归分析,通径分析,土壤中微生物含量

## Statistical analysis of influencing factors of soil microbial content

LI Cong-Jie<sup>1</sup> HAO Yan-Bin<sup>2</sup> HAN Cong-Ying<sup>1\*</sup>

(1. School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

(2. School of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** [Objective] We evaluated statistical methods of multivariate variance, multiple linear regression and path analysis and used multivariate statistical methods to analyze the microbial

**Foundation item:** National Natural Science Foundation of China (No. 71271204, 11331012, 11571014); Science and Education Integration Special Fund Project (No. KJRH2015)

\*Corresponding author: Tel: 86-10-88256908; E-mail: hancy@ucas.ac.cn

Received: January 09, 2016; Accepted: April 07, 2016; Published online (www.cnki.net): May 24, 2016

基金项目: 国家自然科学基金项目(No. 71271204, 11331012, 11571014); 科教融合专项基金项目(No. KJRH2015)

\*通讯作者: Tel: 86-10-88256908; E-mail: hancy@ucas.ac.cn

收稿日期: 2016-01-09; 接受日期: 2016-04-07; 优先数字出版日期(www.cnki.net): 2016-05-24

content in soil. **[Methods]** Through researching principle, application conditions and application to solve the problem of these three common statistical methods, we used different methods to analyze the impact of various factors on the microbial content in soil, and used SPSS software to solve. **[Results]** We found that there were differences in the amount of microbial content in the soil under different rainfall. Ammonium nitrogen content and altitude with Proteobacteria content showed linear relationship. In soil, Proteobacteria content increased with the increase of ammonium nitrogen content, and the direct effect of ammonium nitrogen content on Proteobacteria content was most important. Ammonium nitrogen content, total nitrogen, altitude with Actinobacteria content in soil showed adverse linear relationship, and the negative effect of total nitrogen content on Actinobacteria content in soil was the most, but the direct effect of ammonium nitrogen content on Actinobacteria content was the most. Compared with descriptive statistics and multivariate variance analysis showed that using descriptive statistical methods only draw rainfall was one of the factors of the relative content of different bacterial communities in soil and didn't know the impact of rainfall on the changes of content of microorganisms and what the cause of the significant difference between the 6 gradients of rainfall. Variance analysis methods could solve the above problems. Compared with multiple linear regression analysis and path analysis showed that correlation coefficient obtained by multiple linear regression analysis only said ammonium nitrogen content, total nitrogen content and altitude and deformation of bacterial phyla content (Actinobacteria content) the relationship between the close degree, but unable to explain and analyze the composition and source of the relationship, path analysis methods could solve this problem. **[Conclusion]** Path analysis method in the analysis of relevant problems was superior to multiple linear regression analysis method. Path analysis more than multiple linear regression analysis methods could reflect the direct and indirect impact of factors on the Proteobacteria content (Actinobacteria content). Its analysis conclusion was more intuitive, able to explain the problem.

**Keywords:** Variance analysis, Multivariate linear regression analysis, Path analysis, Soil microbial content

生物学的研究中会经常涉及到生物实验,从而产生海量的调查和实验数据。但是直接从大量的生物数据中分析得到生物学研究中所需的结果,并进行系统整理和深入分析是很困难的,也不现实。这就需要借助统计学原理和方法来分析实验数据背后隐藏的生物学意义,为研究问题的深入、精确化和定量化研究奠定了坚实的科学基础,使得解决生物学所研究的问题变得相对简单;同时可以对生物学研究中涉及到将来发生的事情进行预测,对一些事实用理论的方法加以证实,进行科学的解释<sup>[1-2]</sup>。另一方面,统计学方法<sup>[3]</sup>可以对所研究的生物学问题的实验设计的制定进行科学指导,以满足应用的统计方法所需数据结构的条件,从而更有利于对生物学问题的研究。因此,对生物学中的常用统计方法进行分析及应用就显得十分必要。

## 1 统计方法分析

### 1.1 方差分析

方差分析<sup>[1,3-4]</sup>就是对样本平均数进行假设检验,以鉴别各因素对研究对象影响程度的一种有效的统计学方法。以研究不同降水条件下的微生物含量的变化情况为例,每种微生物门类可以被看做一个整体,共涉及到7个总体,这类问题的研究就要运用多元方差分析方法。由单因素(一元)方差分析的原理及应用条件推广得到多元方差分析的原理及应用条件。

**1.1.1 方差分析的条件:** 方差分析的假设<sup>[1]</sup>: 有效应用方差分析的前提是样本要满足方差齐性、正态性和可加性。

**1.1.2 方差分析的数据转换:** 在生物学领域的研究中,往往会遇到一些不符合方差分析基本假设的样

本数据。在采用方差分析之前,要考虑对实验数据进行适当的数据转换。

(1) 平方根转换<sup>[1]</sup>:当生物学中的样本观测值的平均数与相应方差成一定的比例关系时采用,可以降低极端的变量对方差的影响,得到相同的方差;一般采用的方法是将原观测值转换为 $\sqrt{X_{ij}}$ ,当数据较小时转换为 $\sqrt{X_{ij}+1}$ 。

(2) 对数转换<sup>[1]</sup>:当生物学中的样本观测值的标准差与平均数成比例时采用,可以使处理效应与误差效应的关系由相乘性转换为可加性;一般采用的方法是将原始数据转换为对数 $\log_{10}^x$ 或 $\ln^x$ ,当生物数据中包括0时转换为 $\ln^{(x+1)}$ ;在通常情况下,对数转换方法对降低极端大值影响的效果比平方根转化好<sup>[1]</sup>。

**1.1.3 方差分析的数学模型:**在上述假设下,方差分析问题归结为对原假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_r$ 做显著性检验。现在引入 $\alpha_i = \mu_i - \mu$ 即因素A在水平 $A_i$ 下对指标的效应,于是模型变为:

$$\begin{cases} X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \\ \sum_{i=1}^r n_i \alpha_i = 0 \\ \varepsilon_{ij} \text{相互独立且} \varepsilon_{ij} \sim N(0, \sigma^2) (i=1, 2, \dots, r; j=1, 2, \dots, n_i) \end{cases}$$

在此模型下,方差分析检验的问题等价于 $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ 。

其中 $\mu_i$ :水平 $A_i$ 的理论均值; $\varepsilon_{ij}$ :水平 $A_i$ 下的第 $j$ 次重复试验的试验误差,称为随机误差; $\alpha_i = \mu_i - \mu$ 为因素A在 $A_i$ 水平下对指标的效应。

## 1.2 多元线性回归分析

在生物学中,各变量之间在同一个过程中往往存在一定的相关关系。例如在相同的自然环境下,土壤理化指标、地理及环境等因素与土壤中微生物含量之间存在一定的相关关系,应该考虑这些因素对土壤中微生物含量变化的影响;而分析和处理这些变量之间的相关关系的一种有效方法就是多元线性回归分析方法。

**1.2.1 多元线性回归分析的原理:**多元线性回归分析<sup>[1]</sup>的基本方法是:利用多元线性回归模型,基于最小二乘法建立正规方程,求解得出多元线性回归方程,并对回归方程和偏回归系数进行检验,作出回归方程的区间估计。

**1.2.2 多元线性回归分析的数学模型:** $P$ 元线性回归模型,记为:

$$\begin{cases} y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon \\ E\varepsilon = 0, D\varepsilon = \sigma^2 \end{cases}, \text{其中} \varepsilon \text{是随机因素的总和。}$$

## 1.3 途径分析

多元线性回归分析中的偏回归系数是因变量 $y$ 随自变量变化而变化的反映量,但由于各个 $x_i$ 的单位和标准差不同,各个 $x_i$ 对 $y$ 的贡献大小就不能直接进行比较,而途径分析可以解决这一问题<sup>[1]</sup>。

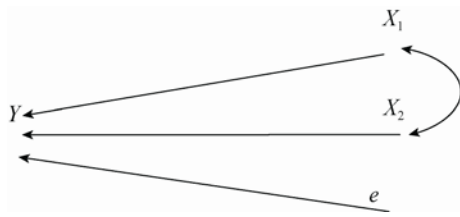
途径分析<sup>[1]</sup>可以将因变量与自变量的相互影响,即相关系数分解为直接影响(直接途径系数)和间接影响(间接途径系数),因此广泛应用于生物的遗传学等领域<sup>[5-6]</sup>。

**1.3.1 途径系数的理论:**图1中有两种路径,第一种是途径,即 $X_i$ 到 $Y$ 之间的单向路径(单箭头线),表示从因到果的路径;第二种是相关线,即 $X_i$ 之间平行关系的双向路径,表示互为因果的路径。所以, $X_i \rightarrow Y$ 的途径称为直接途径,对应着直接途径系数 $P_{iy}$ ;  $X_i \leftrightarrow X_j \rightarrow Y$ 的途径称为间接途径,对应着间接途径系数<sup>[1]</sup>。途径系数的求解:由于偏回归系数 $b_i$ 带单位,不能直接比较各个自变量 $X_i$ 对 $Y$ 影响程度的大小,所以必须对 $X_i$ 、 $Y$ 和剩余项 $e$ 进行标准化变换,使 $X_i$ 、 $Y$ 和 $e$ 转变为不带单位的相对数<sup>[1]</sup>。

转变后的回归方程为:  $\frac{(y - \bar{y})}{s_y} = b_1 \frac{s_1}{s_y} \times \frac{x_1 - \bar{x}_1}{s_1} + \dots + b_m \frac{s_m}{s_y} \times \frac{x_m - \bar{x}_m}{s_m} + \frac{s_e}{s_y} \times \frac{e}{s_e}$ , 则  $P_{iy} = b_i \frac{s_i}{s_y}$ ,  $P_{ey} = \frac{s_e}{s_y}$ ,

其中 $s_i$ 是标准差。

**1.3.2 途径分析的原理:**途径分析在多元回归的基础上,将相关系数 $r_{iy}$ 分解为某一影响因素对因变量的直接作用和该影响因素通过其他因素对因变量

图1  $X_1$  与  $Y$  的通路分析Figure 1 Path analysis of  $X_1$  and  $Y$ 

注:  $X_1$  是自变量;  $Y$  是因变量。

Note:  $X_1$  is independent variable;  $Y$  is dependent variable.

的间接作用, 所以  $r_{iy}$  等于  $x_i$  到  $y$  的直接通路系数  $P_{iy}$  和通过与其相关的各个  $x_j (j=1, 2, \dots, m; j \neq i)$  对  $y$  的所有间接通路系数之和<sup>[7]</sup>。根据通路系数的理论可知,  $P_{iy}$  就是标准化了的偏回归系数, 间接通路系数

之和就是  $\sum_{j \neq i}^m r_{ij} P_{jy}$ <sup>[1]</sup>。

## 2 运用常用统计方法解决生物学实际问题

经过对以上 3 种生物学中常用统计方法的分析, 将运用方差分析、多元线性回归分析、通路分析等方法, 来研究影响土壤细菌群落相对含量变化的因素。样本数据是源于中国科学院大学生命科学学院全球变化生态学课题组在内蒙古天然水分梯度条件下开展的研究工作, 数据主要包括土壤细菌群落的相对多度以及土壤的理化性质等。

### 2.1 描述性统计在生物学中的应用

统计分析的目的是通过样本数据, 可以推断出总体的分布。试图推断出准确的整体分布, 应首先对所探究的数据集进行相应的描述性统计, 以便掌握现状、变量之间的关系等, 寻找出更适合解决问题的统计分析方法。下面以不同降雨量下土壤中各种微生物含量的数据为例进行描述性统计分析。

由图 2 中描述性统计分析可知, 在不同降雨量(如 P450、P400 等)情况下土壤中相同细菌群落的含量是不同的。可见, 降雨量是影响土壤中同种细菌群落相对含量的因素之一。下面将进一步探究降雨量这一因素是如何影响微生物含量的。

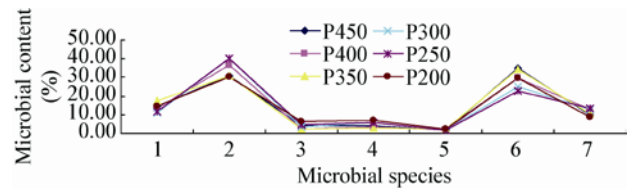


图2 不同降雨量下土壤中微生物含量

Figure 2 Microbial content in soil under different rainfall

注: 1: 酸杆菌门; 2: 放线菌门; 3: 拟杆菌门; 4: 芽单胞杆菌门; 5: 浮霉菌门; 6: 变形菌门; 7: 其他门类。

Note: 1: Acidobacteria; 2: Actinobacteria; 3: Bacteroidetes; 4: Gemmatimonadetes; 5: Planctomycetacia; 6: Proteobacteria; 7: Others.

根据此目的, 经过分析可以运用方差分析方法探究分析降雨量因素对研究对象即土壤中微生物含量的影响程度。

### 2.2 运用多元方差分析方法解决生物学中的实际问题

为了满足运用方差分析方法对数据的要求, 在做生物实验时应控制其他因素在同一个水平上, 需要让降雨量因素(A)改变各种不同的状态进行试验, 然后对所得数据进行分析。

多元正态分布有关均值和方差检验的 SPSS 实现<sup>[8-9]</sup>: 由表 1 威尔克斯(Wilks)检验的  $P$  值为 0.000 可知: 6 个梯度的降雨量下土壤中 7 种微生物的含量有十分显著的差异。6 个梯度降雨量引起地显著差异究竟是由哪种微生物引起的? 对这 6 个梯度的降雨量分别用一元方差分析检验的结果见表 2。由表 2 得出的结果可知: 在不同降雨量下土壤中的酸杆菌门、放线菌门、拟杆菌门、单芽胞杆菌门、变形菌门的含量均有显著差异, 浮霉菌门( $P=0.063$ )这种微生物含量无显著差异。在一定的降雨量范围内, 降雨多少对浮霉菌门含量存在一定的影响, 但影响不是很大。

表 1 有关均值和方差的检验			
Table 1 Test of mean and variance			
Effect	Value	$F$	Sig.
Rainfall Pillai's tracing	2.865	2.460	0.002
Wilks's Lambda	0.000	6.322	0.000
Hotelling's tracing	61.991	11.158	0.000
Roy's maximum root	46.403	85.073	0.000

表 2 主体间效应检验  
Table 2 Test of inter subjective effect

Source	Dependent variables	Mean square	F	Sig.
Rainfall	Acidobacteria	16.209	6.937	0.003
	Actinobacteria	67.602	17.381	0.000
	Bacteroidetes	7.377	8.534	0.001
	Gemmatimonadetes	8.773	66.069	0.000
	Planctomycetacia	0.289	2.854	0.063
	Proteobacteria	70.937	10.648	0.000
	Others	9.394	4.042	0.022

### 2.3 研究影响因素与微生物含量之间的线性相关关系

据研究,土壤中微生物含量的结构受土壤自身理化指标因素、地理及气候因素以及生长在土壤之上的植被总量因素的影响。土壤自身因素即土壤的理化指标,包括土壤中含水量、酸碱度、铵态氮含量、硝态氮含量、总氮含量、总碳含量;地理及气候因素即海拔高度、年均温度;植被土壤根系分泌物影响土壤中微生物含量的变化。

根据 SPSS 相关性分析<sup>[8-9]</sup>结果知,酸杆菌门与土壤含水量、酸碱度、铵态氮、硝态氮、总氮、总碳、海拔、年均温度、植物总生物量都具有一定的线性关系。但与酸碱度(pH)的相关系数为-0.519\*,即存在的线性关系更密切;与海拔关系最不密切。

对样本数据进行逐步回归:由表 3 可知  $P=0.027$ ,说明酸杆菌门含量与各因素之间的逐步线性回归方程是显著的。回归方程的因变量系数和常数项系数的检验  $P$  值分别是 0.027 和 0.001,则得出最优的线性回归方程为:  $Y_1=31.443-2.587X_2$ 。由回归方程可知,酸杆菌门含量与土壤酸碱度因素呈负

表 3 回归模型检验<sup>b</sup>  
Table 3 Test of regression model<sup>b</sup>

Model	Mean square	$D_f$	F	Sig.
1 Regression	29.431	1	5.912	0.027 <sup>a</sup>
Residual	79.654	16		
Total	109.085	17		

注: <sup>a</sup>: 预测变量是(常量)酸碱度; <sup>b</sup>: 因变量是酸杆菌门。

Note: <sup>a</sup>: Predictive variables are (constant) pH; <sup>b</sup>: The dependent variable is the acid bacteria.

相关,随着土壤酸碱度的降低,土壤中酸杆菌门含量会逐渐增多。

与上述方法相同,将其余门类微生物含量分别与各个因素做线性回归,结合自变量与因变量的线性相关系数,经分析可知变形菌门含量分别与酸碱度(pH)、铵态氮、总氮和总碳含量等具有显著的线性关系;放线菌门含量分别与铵态氮、总氮和总碳含量等具有显著的线性关系。从而将变形菌门含量和放线菌门含量分别与影响因素做多元线性回归分析,基于得到的多元线性回归分析的部分结果,进行通径分析,并得出通径分析结果。

**2.3.1 变形菌门含量与各因素之间的多元线性回归分析:**逐步回归得出最优的多元线性回归方程。根据表 4 可知拟合优度  $R^2$  随着因素的逐步引入而有所增大,也就是说引入的因素对变形菌门在土壤中含量的影响在逐渐增加,进而得到有关变形菌门含量的最优的多元线性回归方程。

利用 SPSS<sup>[8-9]</sup>做多元回归方程的显著性检验  $P=0.000$  ( $<0.005$ ),所以此多元线性模型通过了检验,得到变形菌门在土壤中的含量与各因素之间的多元线性回归方程为:  $Y_6=9.460+1.459x_3+0.009x_7$ 。

多元回归模型的线性关系分析表明铵态氮含量和海拔非常显著地影响变形菌门含量,且均呈正相关,但土壤中铵态氮含量对土壤中变形菌门含量变化的影响较大,变形菌门含量随着铵态氮含量增多而增多。

**2.3.2 放线菌门含量与各因素之间的多元线性回归分析:**同理可以得到放线菌门含量的多元线性回

表 4 模型汇总<sup>c</sup>  
Table 4 Model summary<sup>c</sup>

Model	R	R square	Adjusting R square	F changes	Sig.
1	0.767 <sup>a</sup>	0.588	0.562	22.800	0.000
2	0.844 <sup>b</sup>	0.713	0.675	6.544	0.022

注: <sup>a</sup>: 预测变量是(常量)铵态氮; <sup>b</sup>: 预测变量是(常量),铵态氮海拔; <sup>c</sup>: 因变量是变形菌门。

Note: <sup>a</sup>: Predictive variables are (constant) ammonium nitrogen; <sup>b</sup>: Predictive variables are (constant) ammonium nitrogen, altitude; <sup>c</sup>: Dependent variable is Proteobacteria.

归结果。利用 SPSS<sup>[8-9]</sup> 可得最优的多元线性回归方程为： $Y_2 = 60.751 - 1.054x_3 - 14.896x_5 - 0.015x_7$ 。

由回归方程可知：土壤中铵态氮含量、总氮量和海拔高度与放线菌门含量呈显著的线性关系，且均为负相关；其中总氮量对土壤中放线菌门含量的变化所起的负面作用最大，即土壤中放线菌门含量随着总氮量增多而降低。

## 2.4 途径分析的应用

**2.4.1 影响变形菌门含量的途径分析：**根据理论知识可知，多元线性回归方程的标准系数就是途径分析的途径系数<sup>[6]</sup>，所以由途径系数可以看出因素铵态氮含量  $x_3$ 、海拔高度  $x_7$  对土壤中变形菌门含量  $Y_6$  的直接影响作用分别为  $P_{3y}=1.038$ ， $P_{7y}=0.446$ 。利用 SPSS 做各因素间的相关分析<sup>[8-9]</sup> 得到各因素间的相关系数，则  $r_{37}=r_{73}=-0.608$ 。下面分别计算土壤中铵态氮含量  $x_3$  通过海拔高度  $x_7$  对土壤中变形菌门含量  $Y_6$  的间接途径系数和海拔高度  $x_7$  通过铵态氮含量  $x_3$  对土壤中变形菌门含量  $Y_6$  的间接途径系数，即铵态氮(海拔)通过因素海拔(铵态氮)对变形菌门类含量的间接影响作用。计算结果如下：

$$r_{37} \times P_{7y} = -0.608 \times 0.446 = -0.271\ 168；$$

$$r_{73} \times P_{3y} = -0.608 \times 1.038 = -0.631\ 104。$$

直接影响变形菌门在土壤中含量的两个因素中，土壤中铵态氮含量对变形菌门含量的直接影响作用比地理因素海拔相对大，且起到正向促进作用。在间接影响变形菌门在土壤中含量的两个因素中，海拔高度通过铵态氮含量对变形菌门在土壤中含量的负面间接作用较大，为  $-0.631\ 104$ ，而且海拔高度与变形菌门含量的简单相关系数  $r_{7y}=-0.186$  为负值，说明海拔高度对变形菌门含量的影响总体来说是起负作用，即随海拔的升高土壤中的变形菌门含量会逐渐减少；而铵态氮含量通过海拔因素对变形菌门的间接影响虽然也为负值 ( $-0.271\ 168$ )，但由于铵态氮含量对放线菌门含量的直接影响  $P_{3y}=1.038$  较大，从而使铵态氮含量对变形菌门含量的总体影响相对较大。因此，理化指标铵态氮含量和地理因素海拔对变形菌门含量的增加具有重要

的影响作用，土壤中变形菌门含量随着铵态氮含量、海拔高度的降低而增加。

**2.4.2 影响放线菌门含量的途径分析：**土壤中铵态氮含量、总氮含量和海拔因素对土壤中放线菌门含量的直接影响通过通径系数来表示，即直接途径系数为： $P_{3y}=-0.797$ ， $P_{5y}=-0.341$ ， $P_{7y}=-0.783$ ；间接途径系数分别为：

$$r_{35} \times P_{5y} = 0.652 \times (-0.341) = -0.222\ 332；$$

$$r_{37} \times P_{7y} = -0.608 \times (-0.783) = 0.476\ 064；$$

$$r_{53} \times P_{3y} = 0.652 \times (-0.797) = -0.519\ 644；$$

$$r_{57} \times P_{7y} = -0.331 \times (-0.783) = 0.259\ 173；$$

$$r_{73} \times P_{3y} = -0.608 \times (-0.797) = 0.484\ 576；$$

$$r_{75} \times P_{5y} = -0.331 \times (-0.341) = 0.112\ 871；$$

$$r_{35} \times P_{5y} + r_{37} \times P_{7y} = 0.652 \times (-0.341) + (-0.608) \times (-0.783) = 0.253\ 732；$$

$$r_{53} \times P_{3y} + r_{57} \times P_{7y} = 0.652 \times (-0.797) + (-0.331) \times (-0.783) = -0.260\ 471；$$

$$r_{73} \times P_{3y} + r_{75} \times P_{5y} = -0.608 \times (-0.797) + (-0.331) \times (-0.341) = 0.597\ 447。$$

土壤中铵态氮含量、总氮含量和海拔因素对土壤中放线菌门含量变化的直接作用比通过其他因素的间接作用大，且均为负值，说明随着土壤中铵态氮含量、总氮量的增加和海拔高度的增高，土壤中放线菌门的含量会逐渐降低。

## 3 几种统计方法的对比

分别采用描述性统计方法和方差分析方法分析天然水分梯度条件下土壤细菌群落相对多度的数据，从研究结果出发对比描述性统计方法与多元方差分析方法。在研究各因素对研究对象影响的时候，应用描述性统计方法只能得出降雨量是影响土壤中不同细菌群落相对含量的因素之一，而不知道降雨量这一因素对微生物含量变化的影响程度，6个梯度降雨量引起的显著差异究竟是由哪种微生物引起的，方差分析方法可以解决以上的问题。可见，方差分析方法在解决各因素对研究对象影响的问题上明显优于描述性统计方法，更能说明问题，但描述性统计方法是统计分析的基础，是不可忽视的，描述性统计方法和方差分析方法之间存在递进

的关系。

分别采用多元线性回归分析方法和通径分析方法分析天然水分梯度条件下土壤理化指标等的的数据,从研究结果出发对比多元线性回归分析方法与通径分析方法。这两种方法都是研究多个相关变量之间线性关系的统计方法<sup>[1]</sup>,应用多元线性回归方法只能得出:土壤中铵态氮含量和海拔高度与变形菌门含量呈显著线性关系,土壤中铵态氮含量、总氮量和海拔高度与放线菌门含量呈显著线性关系,却不知道这些相关因素对土壤细菌群落相对多度的直接影响和间接影响;通过多元线性回归分析得出了相关系数,而相关系数只能表示铵态氮含量、总氮量和海拔与变形菌门含量(放线菌门含量)之间关系的密切程度,但无法解释和分析这种关系的构成和来源<sup>[1]</sup>,而通径分析方法则可以解决这个问题。可见,通径分析方法在分析相关问题时要优于多元线性回归分析方法,通径分析比多元线性回归分析方法更能体现因素对变形菌门含量(放线菌门含量)的直接和间接影响,使得出的分析结论更直观、更客观、更能说明问题;然而多元线性回归分析方法和通径分析方法又有着相辅相成和递进的关系,两种方法结合使用会更好。

#### 4 结论

本文通过对生物学中常用的3种统计方法进行分析可知:方差分析<sup>[1,4]</sup>方法可以用于分析影响研究对象变化的因素有哪些,以鉴别各因素对研究对象影响的程度;但有效应用方差分析的前提是样本要满足方差齐性、正态性和可加性。多元线性回归<sup>[1,3]</sup>可以用于分析和处理各种因变量与自变量之间的

相关关系,建立多元线性回归模型,通过偏相关回归系数来说明因素对因变量的影响程度。通径分析<sup>[1,5]</sup>(路径分析)是在多元线性回归分析基础上进行的一种实用的统计方法,这种方法比多元线性回归分析更能体现因素对 $Y$ 的直接和间接影响,使我们得出的分析结论更直观、更能说明问题。

#### 参考文献

- [1] Li CX, Shao Y, Jiang LN, et al. Biostatistics[M]. 5th Edition. Beijing: Science Press, 2013: 94-253 (in Chinese)  
李春喜, 邵云, 姜丽娜, 等. 生物统计学[M]. 第5版. 北京: 科学出版社, 2013: 94-253
- [2] Ramette A. Multivariate Analyses in Microbial Ecology[M]. FEMS Microbiology Ecology, 2007,62(2): 142-160
- [3] Jia JP, He XQ, Jin YJ. Statistics[M]. 3rd Edition. Beijing: China Renmin University Press, 2007 (in Chinese)  
贾俊平, 何晓群, 金勇进. 统计学[M]. 第3版. 北京: 中国人民大学出版社, 2007
- [4] Gao ZJ, Shi SL, Li Y. Application of SPSS in ANOVA of biological statistics[J]. Progress in Modern Biomedicine, 2008, 8(11): 2116-2120 (in Chinese)  
高忠江, 施树良, 李钰. SPSS 方差分析在生物统计的应用[J]. 现代生物医学进展, 2008, 8(11): 2116-2120
- [5] Jing YH, Xing LW. Path analysis and its application[J]. Statistical Education, 2006(2): 24-26 (in Chinese)  
敬艳辉, 邢留伟. 通径分析及其应用[J]. 统计教育, 2006(2): 24-26
- [6] Du JJ, Chen ZW. Method of path analysis with SPSS linear regression[J]. Bulletin of Biology, 2010, 45(2): 4-6 (in Chinese)  
杜家菊, 陈志伟. 使用 SPSS 线性回归实现通径分析的方法[J]. 生物学通报, 2010, 45(2): 4-6
- [7] Zhang Q, Cong P, Peng L. The realization of path analysis in Excel and SPSS[J]. Agriculture Network Information, 2007(3): 109-110,91 (in Chinese)  
张琪, 丛鹏, 彭励. 通径分析在 Excel 和 SPSS 中的实现[J]. 农业网络信息, 2007(3): 109-110,91
- [8] Zhang WT, Yan J. Fundamental Tutorial for SPSS Statistics[M]. Beijing: High Education Press, 2004 (in Chinese)  
张文彤, 闫洁. SPSS 统计分析基础教程[M]. 北京: 高等教育出版社, 2004
- [9] George AM. SPSS for Introductory Statistics[M]. Lawrence Erlbaum Associates, 2004