

生物统计的各种检验方法和使用条件概述

房继明

刘来福

(北京师范大学生物学系, 100875) (北京师范大学数学系, 100875)

An Outline of Testing Methods and Conditions in Biological Statistics. Fang Jiming (*Department of Biology, Beijing Normal University, 100875*), Liu Laifu (*Department of Mathematics, Beijing Normal University, 100875*). *Chinese Journal of Ecology*, 1995, 14(2): 67-70.

This paper gives an outline of testing methods in biological statistics which can be used to different types of variables and experimental treatments, especially that of nonparametric analysis and test of related samples.

Key words: biological statistics, nonparametric analysis, test of related samples.

近年来国内生物统计学方面的教科书已有若干本,其内容结构基本相似,包括随机事件与概率、随机变量及其分布、抽样、参数估计、假设检验、方差分析、回归与相关、协方差分析、非参数检验(部分)等,所有论著可能都有这样一个缺欠,即对实验设计、取样、数据类型和其所对应的检验方法未予以全面的、整体的评述,尤其是对非参数检验的介绍很不全面,使得某些不符合参数检验的数据进行了参数检验,这都给从事生物科学的学生、学者进行数据分析时带来一定的困扰,如用错检验方法,或不知如何检验,即不能充分利用原始数据资料,挖掘其内在的信息。本文就生物科学研究中不同类型和实验条件所对应的检验方法,给予较全面的概述。本文主要目的是让生物科学工作者选择和使用正确的统计检验方法,而不必深入了解其统计学原理。此外,随着各类统计学软件包在国内的普及,比如 SPSS/PC+ 软件包,如何使用软件包进行数据分析对生物科学工作者而言也是一个难题,为此本文给出各种统计检验方法的中英文名称。

统计学所研究的数据类型不外乎四

种,额定值(Nominal),如生物的性别分别是雌和雄;序列值(Ordinal),如动物社群结构中的优势个体、亚优势个体、从属个体等;区间值(Interval),如动物的体温;比率值(Ratio),如体重、耗氧量等。不过数据还可以简单地分成离散型(Discrete)和连续型随机变量(Continuous)。

在生物学的研究设计中,首先要确定的是研究材料,它可以从生物大分子,到一个完整的有机体,到一个种群,乃至生物圈,然后是分析不同实验条件下研究材料特征的变化。实验条件的改变可以是生物体自身的,如研究不同的性别、年龄、体重、社会等级等之间的差异;也可以是非生物环境因素,如研究不同季节、温度、光周期、食物等对研究对象的影响。做为统计学上的变量,研究材料的特征和影响研究材料特征的实验因素,即可以是连续型随机变量,也可以是离散型随机变量。实验因素或处理(Factor or Grouping variable)一定含有两个以上的层次或称水平(Level),如水

作者简介:房继明,男,32岁,1992年毕业于北京师范大学生物学系动物生态学专业,获理学博士学位。现留校任副教授、博士生导师,从事动物行为学研究和动物生态学教学。

平数为2时的实验组和对照组。显然,当因素的水平等于1时,就不存在有实验条件了。下面根据研究特征变量的两种类型和个数,实验因素的个数和水平,逐一分析不同情况下所应选用的统计检验方法(表1)。

表1 根据变量类型和实验因素所决定的各种统计检验一览表

Tab. 1 Statistic tests for different types of variables and experimental factors

实验因素	因素的个数	0	1	2	n
	水平的个数	1	2→n	2×2→n×m	≥2
连续型变量的个数	1	A	B C	D	E
	2	F	G		
	n	H	I	J	
离散型变量的个数	1	K			
	2	L			

n, m: >2的整数

首先要明确生物统计中的检验方法可分成两大类:参数检验(Parametric test)和非参数检验(Nonparametric test)。前者依赖于样本的数据是正态分布或二项分布等,而后者不依赖于样本的数据为任何形式的分布。当然,参数检验要比非参数检验更有力。如果检验结果表明数据为非正态分布(方法见本文1.2),可以进行数据代换,使之成为正态分布。代换方式有:平方根代换、对数代换、正切代换等。如果经代换,仍然完全偏离正态分布,就只能用非参数检验了,否则用参数检验无效。

1 描述性统计(Descriptive Statistics)

1.1 参数检验

在假设检验之前,对数据进行概括性描述统计是非常必要的,如了解样本的平均值(Mean)、方差(Variance)、标准差(Standard deviation)、标准误(Standard error of the mean)、区间估计等参数。参数

(Parameter)在统计学中是用来描述整体的数字特征。对于一个正态总体平均值、方差、百分数的假设检验方法请参考《生物统计》一书^[1]。

1.2 非参数检验

在非参数检验中,对数据进行概括性描述统计的参数有范围(Range)、众数(Mode)、中位数(Median)等。更重要的是对样本数据的分布型进行检验。检验方法:Kolmogorov-Smirnov one-sample test 总体分布的卡方检验。

以下所述的参数检验是建立在对样本特征一定的前提假设基础之上的,这些假设包括以下四项:(1)数据为正态分布(Normality);(2)样本之间的方差近似相等,即方差为齐性(Homogeneity of variance);(3)不同实验处理的效果是可加的(Additivity);(4)测量的数据为区间值或比率值,即为连续型随机变量。进行非参数检验的数据分布型可以随意,不过,除卡方检验以外的非参数检验所用的数据不能是额定值数据。

2 两个样本的比较(Comparison between two samples)

不同于标题1,由于对研究对象进行了两种实验处理,所以在进行假设检验之前,首先要明确样本之间的关系是独立的(Independent, unrelated),还是关联的(Matched, related)。这一点非常重要,但很容易被忽略,而用错检验方法。关联样本可以从研究同一组对象在不同实验处理条件下的反应而获得;或从关联的两组(或以上)研究对象在同一实验条件下获得^[2]。一般来说,关联样本比独立样本更容易检验出两组数据平均值之间的差异^[3]。

2.1 参数检验

(1)关联样本检验方法:成对数据平均数的比较,成对T-检验(Paired T-test)。

(2)独立样本检验方法:成组数据平均数的比较,当方差相等时,pooled variance T-test;当方差不等时,separate variance T-test。百分数的检验。

2.2 非参数检验

(1)关联样本检验方法:Wilcoxon rank test。

(2)独立样本检验方法:Mann-Whitney U test。

3 单因素分析

在此种条件下,一个实验因素有三个以上的水平,从而要进行多组数据间的平均值比较,以及进行研究变量与因素之间关系的回归分析。

3.1 参数检验

(1)关联样本检验方法:重复测量的方差分析^[2](Repeated measure MANOVA),一元线性回归($Y = a + bX$);非线性回归(Non-Linear Regression, $Y = b_0 + b_1x + b_2x^2 + \dots + b_nX^n$)。

(2)独立样本检验方法:单因素方差分析(One-way ANOVA)和多重比较(Multiple comparison);一元线性回归($Y = a + bX$);非线性回归(Non-Linear Regression, $Y = b_0 + b_1x + b_2x^2 + \dots + b_nX^n$)。

3.2 非参数检验

(1)关联样本检验方法:Friedman two-way ANOVA。

(2)独立样本检验方法:Kruskal-Wallis one-way ANOVA。

4 双因素分析

4.1 参数检验

(1)两个独立因子检验方法:双因素方差分析(Two-way ANOVA)。

(2)一个独立因子和一个关联因子检验方法:重复测量的方差分析。

(3)两个关联因子检验方法:重复测量的方差分析。

4.2 非参数检验

检验方法:尚不清楚。

5 多因素分析

5.1 参数检验

检验方法:多因素方差分析(More-way ANOVA)。

5.2 非参数检验

检验方法:尚不清楚。

6 相关和回归分析(Correlation and Regression)

分析两个变量间的关系,有两种方法,一个是相关系数的测定,一个是回归分析。

6.1 参数检验

(1)检验方法:相关系数(Correlation Pearson r)。

(2)检验方法:一元线性回归($Y = a + bX$);非线性回归(Non-Linear Regression, $Y = b_0 + b_1x + b_2x^2 + \dots + b_nX^n$)。

6.2 非参数检验

检验方法:秩相关系数(Correlation Spearman Rank r)。

7 协方差分析(Analysis of Covariance)

在所研究的两个变量中,一个为反应变量(Response variable),另一个与反应变量有线性回归关系,称为协变量(Covariate)。

7.1 参数检验

检验方法:协方差分析。

7.2 非参数检验

检验方法:尚不清楚。

以下标题 8、9、10 分析方法属于高级生物统计,是对多个研究变量在不同实验因素下的统计分析。

8 聚类分析、主成分分析、因子分析、多元回归分析(Cluster Analysis, Principle Analysis, Factor Analysis, Multiple Regression)

8.1 参数检验

检验方法:检验数据结构;主成分分析、因子分析;检验一个变量与其它变量之间的关系;多元回归分析。

8.2 非参数检验

检验方法:检验是否存在组;聚类分析。

9 判别分析(Discriminant Analysis)

9.1 参数检验

检验方法:判别分析;两组变量的典型相关(Canonical correlation)。

9.2 非参数检验

检验方法:尚不清楚。

10 多元分析

10.1 参数检验

检验方法:多元方差分析(MANOVA, Multivariate Analysis of Variance),包括:单因素的多元方差分析;双因素的多元

方差分析;多因素的多元方差分析;两组变量的典型相关。

10.2 非参数检验

检验方法:尚不清楚。

离散型数据的卡方检验为一种非参数检验,因为它需要假设数据的分布类型。

11 单样本卡方检验(Single-sample Chi-Square test)

检验方法:多项分布、总体分布的卡方检验。

12 多样本卡方检验(Multiple-sample Chi-Square test)

检验方法: $n \times m$ 列联表中独立性的卡方检验(Cross-tabulation test)。

上述这些检验除秩相关系数均可在SPSS/PC+软件包上运算。

参考文献

- [1] 刘来福,程书肖.生物统计.北京师范大学出版社,1988,190-196.
 [2] Stanton A. Glantz and Bryan K. Slinker. Primer of Applied Regression and Analysis of Variance. New York, McGraw-Hill, Inc. 1990.
 [3] 杜荣寿.生物统计学.北京:高等教育出版社,1985,131-132.
 (收稿:1993年11月30日,改回:1994年9月7日)

(上接45页)

- [34] Samson D. A. et al. Size-dependent effects in the analysis of reproductive effort in plants. *American Naturalist*, 1986, 127: 667-680.
 [35] Weiner J. The influence of competition on plant reproduction. *Plant Reproductive Ecology: Patterns and Strategies* (Ed. by J. Lovett Doust and L. Lovett Doust). Oxford University Press, Oxford. 1988. 228-245.
 [36] Klinkhamer P. G. L. et al. How to test for proportionality in the reproductive effort of plants. *American Naturalist*, 1990, 135: 291-300.
 [37] 周纪纶等.植物种群生态学.北京:高等教育出版社,1992,147-156.
 [38] Silvertown J. W. (祝宁等译).植物种群生态学导论.哈尔滨:东北林业大学出版社,1987. 134-140.

- [39] Porter J. R. A modular approach to analysis of plant growth. I. Theory and principles. *New Phytol.*, 1983, 94: 183-190.
 [40] Maillette L. Structural dynamics of silver birch 1. The fates of buds. *Journal of Applied Ecology*, 1982, 19: 203-218.
 [41] Levins R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 1969, 15: 237-240.
 [42] 中国科学院生物多样性委员会.生物多样性研究的原理与方法.北京:中国科学技术出版社,1994,55-68.
 [43] 祝宁主编.植物种群生态学研究现状与进展.哈尔滨:黑龙江科学技术出版社,1994,88-99.
 (收稿:1994年9月11日,改回:11月2日)