文章编号: 1000-5404(1999)12-0890-03

重复观测数据的相关分析

王文昌[1],张 蔚[1],易 东[1],冷怀明[2],许汝福[1] (第三军医大学: [1]高原医学系医学统计学教研室; [2]学报编辑部;重庆 400038)

【提要】 目的: 本研究将建立重复测量数据的相关分析技术。方法: 利用半参数回归分析的原理与方法,结合重复观测数据的特点,建立重复观测数据的半参数回归模型,剔除重复因素的影响后,建立重复测量数据的相关系数公式,并讨论了相应的假设检验问题。结果: 给出了一个具体的实例,计算得其相关系数为 $r=0.3645$, $P>0.05$,与分别在各个重复点上计算得到的线性相关系数是吻合的。结论: 结果表明对重复观测数据进行相关分析时,所建立的方法是有效的、实用的,并且效果优于普通的最小二乘法和广义最小二乘法。

关键词 重复测量数据;相关分析;半参数回归
中图法分类号 R195.1 文献标识码 A

重复观测数据广泛存在于医学研究中,对重复测量数据进行相关分析无论在理论上还是实践上都有重要意义。假设在研究总体中随机地选出 n 个受试对象,分别对每个对象按一定的时间间隔重复测量 p 次,得观测值 $(X_{ij}, Y_{ij}, i=1,2,\cdots,n, j=1,2,\cdots,p)$ 这种观测数据称为纵向重复测量数据。由于对同一对象进行重复测量,因此其数据间存在自相关性,当对其进行相关回归分析时,通常的最小二乘法将失效[1]。前不久,我们运用统计学中才兴起的半参数回归方法[2,3,4]研究了此类数据的回归分析问题[5],建立了对其进行处理的半参数回归模型,讨论了模型的参数估计及假设检验等问题。本文将对重复测量数据的相关程度进行定量描述,并研究了相应的假设检验问题。

1 原理与方法

我们对纵向重复测量数据建立了如下半参数回归模型[5]:

$$Y_{ij}=b(X_{ij}-\bar{X})+g(j)+\varepsilon_{ij} \quad (1)$$

其中 b 为模型参数,主要反应了 X 对 Y 的影响。$\bar{x}=\frac{\sum_{ij}X_{ij}}{np}$ 为 X 的平均数。$g(j)$ 为未知函数,主要反应了重复测量的效应。$\varepsilon_{ij}\sim N(0,\sigma^2)$ 为随机误差且相互独立。对模型中的参数及未知函数可作如下估计:

$$\hat{g}(j)=\frac{\sum_i Y_{ij}}{n}-b\frac{\sum_i (X_{ij}-\bar{X})}{n} \qquad j=1,2,\cdots,p \quad (2)$$

$$\hat{b}=\frac{\sum_{ij}\tilde{X}_{ij}\tilde{Y}_{ij}}{\sum_{ij}\tilde{X}_{ij}^2} \quad (3)$$

其中, $\tilde{X}_{ij}=(X_{ij}-\bar{X})-\frac{\sum_i (X_{ij}-\bar{X})}{n}$, $\tilde{Y}_{ij}=Y_{ij}-\frac{\sum_i Y_{ij}}{n}$。

最后,可用表 1的方差分析表对模型进行假设检验。

表 1 模型(1)的方差分析表

Tab 1 Table of ANOV about Model (1)

Variant source	SS	v	MS	F
Total	$\sum_{ij}(Y_{ij}-\bar{Y})^2$	np-1		
Regression	$\sum_{ij}(\hat{Y}_{ij}-\bar{\hat{Y}})^2$	1	SS_{reg}/V_{reg}	MS_{reg}/MS_{err}
Repeated	$\sum_{ij}(\bar{Y}_j-\bar{Y})^2$	p	SS_{rep}/V_{rep}	MS_{rep}/MS_{err}
Residual	$\sum_{ij}(\tilde{Y}_{ij}-\tilde{\hat{Y}}_{ij})^2$	np-p-2	SS_{err}/V_{err}	

其中 $\bar{Y}_j=\frac{\sum_i Y_{ij}}{n}$。事实上,$(\tilde{X}_{ij},\tilde{Y}_{ij})$ 为原始数据消除了重复因素所导致的自相关性后所得的序列,因此 $\tilde{X}_{ij}$ 与 $\tilde{Y}_{ij}$ 间的相关性才能准确地反应原始数据消除重复因素的影响后的相关性。于是根据表 1,我们可定义原始数据 (X_{ij}, Y_{ij}) 的相关系数如下:

$$r=\frac{\sum_{ij}(\tilde{X}_{ij}-\bar{\tilde{X}})(\tilde{Y}_{ij}-\bar{\tilde{Y}})}{\sum_{ij}(X_{ij}-\bar{\tilde{X}})^2\sum_{ij}(Y_{ij}-\bar{\tilde{Y}})^2} \qquad v=np-p-2 \quad (4)$$

对相关系数 r 的假设检验方法与线性相关系数的假设检验相同。

2 实例

表 2为某幼儿园 9名幼儿连续 3年的体重与身高数据。文献[6]曾分析得受试对象内残差相关是显著的,并且断言使用普通最小二乘法得到的体重与身高数据的回归系数 $\hat{b}$ 的标准误将高估真实的标准误,于是在文献[6]中对 $\hat{b}$ 的标准误进行调整后,得到对 $\hat{b}$ 的假设检验的 t 值为 6.2667,同时采用广义

基金项目: 国家自然科学基金资助项目(39870691)
作者简介: 王文昌,男,1970.02.03生,贵州省大方县人,硕士,讲师,主要从事卫生统计方面的研究,发表论文 6篇。电话: (023)68752343
收稿日期: 1999-01-02 修回日期: 1999-09-23

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

最小二乘法也得到 $r=0.9808$,进一步说明了对 $\hat{b}$ 的假设检验是显著的,即幼儿身高与体重有显著关系的结论。事实上,我们分别对 1991年、1992年、1993年数据进行相关分析,则 3个 r 值分别为: 0.4891、0.2377、0.3878,其相应的 P 值均大于 0.05,也就是说,对此资料而言,如果剔除年龄因素的影响,幼儿的身高与体重应该是无关的。

表 2　某幼儿园幼儿的体重与身高

Tab 2　Height and weight of infants

Number		Observation date		
		6/7/1991	25/5/1992	24/5/1993
1	X(cm)	105.0	113.7	120.0
	Y(kg)	16.3	18.7	19.7
2	X(cm)	102.5	110.3	117.0
	Y(kg)	16.2	20.3	23.5
3	X(cm)	104.5	110.4	117.7
	Y(kg)	15.81	18.4	20.4
4	X(cm)	109.8	117.1	124.4
	Y(kg)	16.1	18.6	21.0
5	X(cm)	103.8	110.0	112.0
	Y(kg)	17.0	19.0	18.55
6	X(cm)	106.0	113.7	120.7
	Y(kg)	16.1	17.8	19.3
7	X(cm)	110.3	115.8	123.7
	Y(kg)	20.3	24.3	27.9
8	X(cm)	103.0	110.5	116.2
	Y(kg)	17.3	19.7	21.6
9	X(cm)	106.9	113.1	121.0
	Y(kg)	17.0	18.7	19.1

我们现用模型(1)来讨论表 2中的数据,则有: $\hat{b}=0.25236$, $\hat{g}(1)=18.6349$, $\hat{g}(2)=19.4477$, $\hat{g}(3)=19.5463$,并对其半参数回归模型进行假设检验得到: 对回归系数,$P>0.05$;对重复因素,$P<0.01$,即就此资料而言,如果排除年龄因素的影响后,由于幼儿的体重变异远远大于身高的变异而掩盖了身高对体重的影响,幼儿的身高与体重是无关的,而对体重影响的主要因素为年龄因素,这与分别在各时间点上对身高与体重的相关分析结果是吻合的。更进一步地,我们用本文方法计算在此 3年中,剔除年龄因素后,幼儿身高与体重的线性相关系数为 $r=0.3645$,与分别计算得到的相关系数较为接近,相应的 $P>0.05$,也说明了对此资料而言身高与体重是无关的。

3　讨论

对重复测量数据进行回归分析时,由于数据间的自相关性,一般的最小二乘法将失效。事实上,在文献[6]中无论是采用对标准误进行调整还是采用广义最小二乘法均得到了与实际资料所显示的信息不相符的结果。此实例的结果显示,半参数回归分析用于处理重复测量数据时的效果优于普通的最小二乘法和广义最小二乘法。

对重复测量数据作相关分析时,应考虑重复因素所导致的数据间的相关性。一般来说,重复因素所导致的相关性是比较显著的,如果不剔除重复因素的效应,将有可能夸大指标间的相关程度。

对重复测量数据作相关分析在实际中有着重要的意义。一般的模型仅能解决指标间有无相关性问题,而不可能对相关程度作出定量的描述。虽然模型的 P 值从某个角度来说能反映指标间的相关程度,但不利于实际分析使用。我们这里所导出的重复测量数据间的相关系数能真实地反应重复测量指标间的相关程度,无论是在理论上还是直观上它都是与线性相关系数一致的。

参考文献

[1] 陈希儒,王松桂.近代回归分析——原理方法及应用[M].合肥: 安徽教育出版社, 1987. 91-105.

[2] 高集体,洪对岩,梁　华,等.半参数回归模型研究的若干进展[J].应用概率统计, 1994, 10(1): 96-103.

[3] 洪圣岩.一类半参数模型的估计理论[J].中国科学(A辑), 1991, 12(12): 1258-1272.

[4] 王文昌,易　东,许汝福,等.季节性时间序列资料预测的半参数回归模型[J].中国卫生统计, 1997, 14(6): 4-7.

[5] 王文昌,易　东,张　蔚,等.重复观测数据的半参数回归分析[J].中国卫生统计, 1999, 16(4): 200-203.

[6] 赵景波,李　康.重复观测值线性回归分析及其在医学中的应用[J].中国卫生统计, 1995, 12(2): 18-19.

(编辑　陈聪连)

Correlative analysis of repeatedly measured data

WANG Wen-chang, ZHANG Wei, YI Dong, LENG Huai-ming, XU Rufu (Department of Medical Statistics, Third Military Medical University, Chongqing 400038)

Abstract　Objective: To seek the method for correlative analysis of repeatedly measured data. Methods: Using the theories and methods of semiparametric regression, the semiparametric regression model of repeated measurements was established with consideration of characteristics of repeated

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

measurements. By the means of this model, the factors affecting repeated measurements were eliminated and the formula of the correlative coefficient was obtained. **Results**: When a real example was given, the correlative coefficient was 0. 3645. The coefficient was similar as that of each repeated measurement. **Conclusion**: This method is effective and practical for the correlative analysis of repeatedly measured data. Its effect is better than that of the ordinary least square and the generalized least square analyses.

Key words repeated measurement; correlative analysis; semiparametric regression

Foundation item: National Natural Science Foundation of China(39870691)

文章编号: 1000-5404(1999)12-0892-01

一氧化氮对大鼠胃窦肌间神经丛胆碱能神经的影响

Effects of nitric oxide on cholinergic nerves of myenteric plexus in gastric antrum of rats

龙庆林,王振华 (第三军医大学附属西南医院消化内科,重庆 400038)

一氧化氮是胃肠道主要的非胆碱能非肾上腺素能抑制性神经递质,它在胃肠道分布广泛且含量较高。形态学研究表明一氧化氮可能影响其它神经递质的释放[1]。本试验拟通过观察正常大鼠注射一氧化氮供体硝普钠后胃窦肌间神经丛内胆碱能神经变化,间接了解一氧化氮对乙酰胆碱合成的影响。

1 材料与方法

健康 Wistar大鼠 24只,雌雄各半,体重 250~ 350 g,随机分为 3组,每组 8只:① 正常对照组;② 一氧化氮组,每天腹腔注射硝普钠 0. 5 mg/kg, 3/d;③ 左型精氨酸甲脂 (L-NAME)组,每天腹腔注射 L-NAME 15 mg/kg, 2/d。

4周后将各组大鼠用 3%戊巴比妥钠麻醉后取出全胃,向胃内灌注适量 4%的多聚甲醛液后放入固定液中固定 12 h,再放入 30%蔗糖液脱水。将胃窦剪成 0. 4 cm× 0. 8 cm大小的组织块,胃窦粘膜表面固定后呈绒毛状,粘膜及粘膜下层连接较紧密,可用钟表镊一块撕掉,根据平滑肌纤维走向判断出斜肌层和环肌层并依次撕去,即得到附于纵肌之上的肌间神经丛。采用 Kamovsky-Root's直接法显示胆碱能神经。

采用 MD20型真彩色图像分析仪进行图像分析,于 25倍物镜下每个标本随机取 3~ 5个视野,计算阳性神经平均面密度(目标面积占统计面计的百分比),反映阳性神经在铺片中的含量。

统计学处理采用单因素方差分析。

2 结果

光镜下,正常大鼠胃窦肌间神经丛中的胆碱能神经纤维呈"树根状"或"网状"分布,其神经节较粗,阳性神经元胞体较多;节间纤维相对细小,有少许阳性神经元散在分布。L-NAME组的胆碱能神经分布与正常组类似,而一氧化氮组胃窦肌间神经丛内胆碱能神经节更粗大,阳性神经元胞体密集,体积较大,其节间纤维也增粗。

图像分析显示,一氧化氮组胆碱能神经平均面密度为(26. 75± 4. 72)%,正常对照组为(22. 04± 5. 80)%,两组比较,相差显著($P < 0.05$); L-NAME组为(21. 17± 4. 00)%,与正常组比较,无显著差异。

3 讨论

一氧化氮能调控神经释放乙酰胆碱,但不同的动物和不同组织在不同实验条件下其结果不一样[2]。在胃肠道,学者们对一氧化氮是促进还是抑制乙酰胆碱释放意见不一, Katalina[3]认为一氧化氮的作用是双向的,基础状态下,一氧化氮供体明显促进豚鼠肌间神经元释放乙酰胆碱,且与浓度成正相关;局域性电刺激时,一氧化氮却抑制乙酰胆碱释放。

本试验以大鼠为研究对象,通过观察胃窦肌间神经丛胆碱能神经含量变化间接了解乙酰胆碱的变化。注射硝普钠后,大鼠胃窦肌间神经丛内胆碱能神经节增大,胞体密集,节间束增粗;图像分析也表明一氧化氮组胆碱能神经平均面密度较正常组高,故认为一氧化氮能促进大鼠胃窦肌间神经元合成乙酰胆碱。

近来有作者认为 L-NAME能抑制毒蕈受体,但如果对胆

作者简介: 龙庆林,男, 1970. 10. 01生,湖南省茶林县人,硕士,住院医师,主要从事消化道动力学方面的研究,发表论文 3篇。电话: (023) 68754114-73049

收稿日期: 1999-04-10;修回日期: 1999-07-15

(下转 895页)

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net