

气象中使用统计检验的几个问题

黄嘉佑

(北京大学物理学院大气科学系, 北京 100871)

提 要

针对目前在气象和气候要素变化分析中,对相关系数和合成分析使用统计检验存在的若干问题进行评论。认为对相关系数等统计检验,不应该称为“信度检验”,其检验水平也不应该称为“95%置信度”,而应该称为“显著性检验”,“显著性水平为5%”。

关键词: 统计检验 显著性水平 相关系数

引 言

在气象统计分析中,经常要对资料进行分析,或制作气象水文要素的分析和预测。例如要分析测站资料的均一性,要研究气象与水文变量(降水量、气温等)不同时段是否有显著差异^[1],或者要检验两种不同条件下,例如 ENSO 年与非 ENSO 年的气候要素合成场差异性比较,要判别它们是否有显著差异,也需要进行检验^[2]。这些检验在气候诊断中有广泛的应用。另外,对气象水文要素进行预测,需要研究它们与其它因子之间的关系,建立它们关系的统计模型^[3]。度量它们关系的密切程度,经常要使用相关系数。计算值的大小判断要进行检验,其绝对值如果大于某个给定值,就称该关系在某种水平上是存在明显相关的。这种过程一般称为相关系数的检验。

在国内外一些天气气候诊断,应用相关系数分析的文章中,对这种检验过程的称呼有各种说法。例如把某两个相关场中高于或低于某一数值绘成阴影区域,对这些区域有的称为未达到“95%信度的区”;有的称“阴影区表示显著性信度检验超过 95% 的区域”;有的称“阴影区为 t 检验信度超过 95% 的区域”;有的称“阴影区为通过 95% 显著性检验的区域”;有的称“阴影区超过 99% 的可信度检验”;有的称“阴影区超过 5% 的可信度检

验”等等。有的把这种相关程度的检验过程称为“信度检验”(confidence test),把两个变量在某种水平上是存在的相关性判断,称为“信度水平”(confidence level),认为达到“95%信度”水平就表示两个变量相关是明显的。有人还认为:应该把“5%”称为“信度”,把“95%”称为“置信水平”,在文章中既用“5%信度”又用“95%置信水平”来说明两个变量的密切程度。

一般人认为,判断两个变量是否有关,如果计算的相关系数绝对值越大,表示两个变量相关的程度越高,其相关的可信的水平也应该越高,认为使用 95% 表示可信水平是合理的,可以称为 95% 信度,把这种检验过程称为信度检验,有些文章作者干脆就称此检验过程为“可信度检验”,把检验的判据程度称为“可信度水平”。这些提法存在很大的混乱,数值 95%、5% 也存在各种提法,到底哪种提法是正确呢? 正确的称呼是什么呢? 本文针对这些问题提出看法。

1 关于信度与置信度

上述关于相关系数等参数的检验,其理论基础均来自统计检验。在统计学中,提到“信度”这个名词是在参数区间估计中,其有关的论述为:“设总体 X 的分布含有未知参数 θ , 根据给定的置信度 $1 - \alpha$ ($0 < \alpha < 1$), 由样本 $X_1, X_2, \dots, X_n$ 确定两个统计量 θ_1 ,

θ_2 , 使得 $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$, 则称随机区间 (θ_1, θ_2) 为 θ 的置信度为 $1 - \alpha$ 的置信区间”(参见文献[4,5])。即只有参数估计才用“置信度”和“置信区间”的名称。有的统计学书称“ $1 - \alpha$ 是置信度(或信度), 置信度也称为置信概率”, 随机区间是“信度 $1 - \alpha$ ”的置信区间, 把“ α ”称为显著性水平(参见文献[6], 文献[7])。“置信度是指与一个区间估计相联系的概率。这个概率是该区间估计将包括总体参数的可信程度有多大。显然概率越大, 置信度就越大”(参见文献[8])。那么, 上面置信度的提法显然与此有关, 即与参数估计有关。

2 关于相关系数的检验

一般在气象中关于相关系数的检验问题, 是否属于参数的置信区间估计的问题, 即能否用置信区间来进行相关系数检验?

两个随机变量总体的相关系数, 在概率论中是一个描述两个总体相关程度的参数。如果一定要对相关系数作区间估计, 则需要对该参数的分布函数作计算才能够确定置信区间具体位置。而相关系数(r)的分布密度函数为^[3]

$$f(r) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} (1-r^2)^{\frac{n-4}{2}} \quad (1)$$

在一般的统计学书中, 没有给出相关系数的分布取值与对应的分布函数值的表。显然利用相关系数的计算值与相关系数(参数)的区间估计很难对应, 而我们根据相关系数的计算值进行置信区间的估计是十分困难的。有的统计学者提出用 Z 变换转化相关系数, 然后进行相关系数的置信区间估计, 但是这种过程与直接计算相关系数的置信区间完全不同。

我们常用的对相关系数的判断过程, 显然不是对这个参数进行区间估计, 而是根据相关系数的计算值, 对两个变量的相关密切程度作判断。

实际上, 对变量关系密切程度作出判断的过程, 使用的是统计学中的假设检验, 又称为显著性检验。根据传统的假设检验过程,

需要对两个变量总体相关系数作假设, 通常假设它们没有相关, 即原假设总体相关系数为零, 即 $H_0: \rho = 0$, 其对立假设 $H_0: \rho \neq 0$, 然后用与相关系数有关的统计量

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (2)$$

作判断。因为该统计量遵从自由度为 $n-2$ 的 t 分布。判断时计算该统计量的观测值, 对照一般统计书所给的 t 分布表, 从而对变量关系密切程度作出判断(见文献[3]与[9])。在判断过程中, 需要事先确定检验水平。由于假设检验所依据的基本原理, 是小概率事件的实际推断原则。也就是说, 概率很小的事件在一次试验中实际上不大可能出现。如果在一次试验中, 它发生了, 我们就有理由怀疑原来对该事件提出的假设, 即原假设的正确性有问题, 从而拒绝原假设。通常的检验水平使用小概率, 即常用的 10%、5% 和 1%。统计书^[6,7]中说: “当 H_0 为真而作出拒绝 H_0 的判断, 称为犯第一类错误; 而 H_0 不真而作出接受 H_0 的判断, 称为犯第二类错误。犯第一类错误的概率不大于一个较小的正数 α , 而使犯第二类错误的概率 β 尽可能的小。这里 α 称为检验的显著水平, 或检验水平。”对于该检验水平有的书^[5]说: “给定的值 α ($0 < \alpha < 1$) 称作显著性水平, 通常取 $\alpha = 0.10, 0.05, 0.01$ 。”

显然, 上述关于相关系数的检验过程与相关系数置信区间估计是不同的。在文献[8]中更明确为: “这里要强调指出的是: 假设检验中的显著水平和区间估计中的置信度, 显然是不同的。与显著水平相反, 置信度是指落入指定的置信限以内的样本平均数所占的百分比。”而显著水平是否定假设可能犯错误的概率。同样, 即使在国外的经典统计学中, 也明确把犯第一类错误的概率称为“level of significant”^[9], 在文献[9]中说: “The word ‘significant’ is used in situations in which a null hypothesis is rejected. Type I error is called the level of significant, at $\alpha = 0.05$ or $\alpha = 0.01$ 。”

因此, 在相关场的阴影区的分界值的决

定,实际是判断两个变量相关密切程度的过程中的是否显著与零有差异的判断,实际就是相关系数检验的问题,应该按照显著性检验的规范进行检验,所提出的原假设和使用有关的统计量,应该严格根据统计量分布要求,查出在某个显著性水平 α 和某个自由度下的临界值。检验的可信程度取决于显著水平 α ,显著水平越小,犯第一类的错误越小,否定假设的可信度就越高。

至于认为使用 $1 - \alpha$ 表示可信水平,称为信度,在经典的假设检验中是没有的。所以,有些人认为既然检验的可信程度取决于检验水平 α ,就把 α 写为 95%,有人甚至称为“95%的显著性检验”或“99%的显著性检验”,也是不正确的。

而且信度不等于参数,信度是一个概率度,所谓“信度检验”的提法似乎也不妥,“信度”怎样进行“检验”呢?显然,统计学上的参数检验或非参数检验不能称为“信度检验”。因此,我们常用的相关系数检验应该使用假设检验或显著性检验,检验的水平用 α 表示。实际上,在很多有关的统计方法介绍的书中,一般的参数检验(包括相关系数的检验)和非参数检验,所使用的标准均使用以小概率程度作为显著性水平,并按规范的假设检验程序进行检验(参见文献[10~14])。

3 结 论

综合上面的论述,在相关系数的检验中使用置信度 95% (99%) 是不合适的。如果把某两个相关场中高于或低于某一数值绘成阴影区域,对这些区域称呼应该为“超过或低于显著性水平 5% 的区”,相关系数的检验过程应该称为“显著性检验”。“ α ”的选取,一般作为小概率的标准,通常在气象中,从应用

角度上多令它等于 5%。

实际上,对气象中有关参数的检验,例如平均值检验,两个变量平均值差异的检验也应该类似相关系数的参数检验,其相应的称呼也应该按照经典的显著性检验(或假设检验)过程的称呼为好。

参考文献

- Peterson, T., David, R. E., and T. R. Karl et al., Homogeneity adjustments of in site atmospheric climate data: a review, Int. J. Climatol. 1998, 18: 1493—1517.
- 黄嘉佑. 统计动力分析与预报. 北京: 气象出版社, 1993: 243.
- 黄嘉佑. 气象统计分析与预报方法(第三版). 北京: 气象出版社, 2004: 298.
- 《考研数学复习全书》编写组. 线性代数与概率统计复习指南. 上海: 东南大学出版社, 2000: 368.
- 《双博士考研数学》课题组. 2005 年硕士研究生入学考试应试教程. 北京: 机械工业出版社, 2004: 493.
- 袁卫, 赵彦云, 易丹辉等编. 统计学. 北京: 中国统计出版社, 1994: 331.
- 袁荫堂编. 概率论与数理统计. 北京: 中国人民大学出版社, 2003: 331.
- 何海燕, 张红元, 黄发贵等编. 商务统计. 北京: 广东经济出版社, 2001.
- Freund, J. E.: Modern Elementary Statistics (Tenth Edition), Prentice-Hall International (UK) Limited, London, 2001: 831.
- 陈家鼎, 刘婉如, 汪仁宫编. 概率统计讲义. 北京: 人民教育出版社, 1980: 311.
- G. W. 斯奈迪格等著. 数理统计方法. 北京: 科学出版社, 1986: 609.
- 郑德如主编. 统计学. 上海: 立信会计出版社, 1996: 327.
- 陈词成, 冯虹主编. 新编统计学原理. 北京: 北京经济学院出版社, 1996: 347.
- 何晓群, 刘文卿编. 应用回归分析. 北京: 中国人民大学出版社, 2001: 248.

Some Problems on Using Statistical Test in Meteorology

Huang Jiayou

(Department of Atmospheric Science, School of Physics, Peking University, Beijing 100871)

Abstract

The comment on statistical test of correlation coefficients and others are given. The statistical test is used to apply to some analysis of climate variation of meteorological elements. The conclusions are that it should be call as “significant test” and “significant level of 5%”, rather than as “confidence test” and “confidence level of 95%”.

Key Words: statistical test meteorology significant level