气象水文统计分析中显著性检验的应用

黄嘉佑

（北京大学物理学院大气科学系，北京，100871）

摘要

本文介绍在水文气象中经常遇到的显著性检验问题。例如两种不同条件下，差异检验。或者，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，对相关场的检验问题等。并提出关于合成场的检验方法，及其应用。

关键词：显著性检验；相关系数检验；合成场检验；

在水文气象中，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）不同时段是否要显著差异【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。或者，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的。这种过程一般称为相关系数的检验。

对这种关于两个变量关系的判断过程，或差异性检验，在这些检验中，通常使用统计学中的显著性检验。但是在国内外一些气候诊断或检验应用的文章中，对这种检验过程的称呼有各种说法。有的把这种过程称为‘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。也有些人认为，判断两个变量是否有关的，如果计算的相关系数越大，表示两个变量相关的程度越高，其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，有文章作者干脆就称为‘可信度检验’，‘可信度水平’。这些提法是否正确呢？在统计学中，提到‘信度’这个名词是在参数区间估计中，有关的论述：‘设总体 X 的分布含有未知参数θ，根据给定的置信度 1－α （0<α<1），由样本 $X_1, X_2, ..., X_n$ 确定两个统计量 q_1, q_2，使得 $P(q_1 < q < q_2) = 1 - a$，则称随机区间（q_1, q_2）为θ的置信度为 1－α 的置信区间。’（参见文献【4~5】）。有的统计学书称‘1－a 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－a’的置信区间，把‘a’称为显著性水平。（参见文献【6】）。‘a 是事先给定的一个小正数，它是指参数估计不准的概率（假设检验中称 a 为检验水平）。’（参见文献【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】）。显然，‘5％信度’和‘95％置信水平’在参数区间估计中是没有的。而‘95％信度’是指参数区间估计的水平，而不是判断相关系数密切程度的水平。即使在国外的经典统计学中，也把参数的置信区间称为‘confidence interval’，把置信区间两个端点称为‘confidence limits’，把 1－a 或（1－a）100％称为‘degree of confidence’（参见文献【9】）。显然，在统计学中，关于置信度，国内外的称呼是一致的。

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，上述对相关系数的判断过程，不是对这个参数的区间估计，而是对相关密切程度作判断。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行。而相关系数（r）的分布密度函数为（参见文献【3】）

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

在一般的统计学书中，没有给出相关系数的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，当然对变量关系密切程度也很难作出判断。

实际上，对变量关系密切程度作出判断的过程，使用的是显著性检验。统计学中把显著性检验称为假设检验。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 $H_0:r=0$，其对立假设 $H_0:r\neq 0$，然后用与相关系数有关的统计量

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

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，根据假设检验步骤，从而对变量关系密切程度作出判断。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 a，而使犯第二类错误的概率 b 尽可能的小。这里 a 称为检验的显著水平，或检验水平。’（参见文献【6～7】）。对于该检验水平有的书说：‘给定的值 $a(0<a<1)$ 称作显著性水平，通常取 $a=0.10,0.05,0.01$。’（参见文献【5】）。在文献【8】中更明确：‘这里要强调指出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。’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】。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）和非参数检验均使用显著性水平的称呼，并按规范的假设检验程序进行检验（参见文献【10～14】）。因此，在相关系数的检验中使用置信度是不合适的。因此，在判断两个变量相关密切程度的过程中，需要注意相关系数检验的问题，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 a 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 a，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 $1-a$ 表示可信水平是合理的，称为信度，既然检验的可信程度取决于检验水平 a，就把 a 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。信度不等于参数，信度是一个概率度，怎样进行‘信度检验’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’，而应该称为假设检验或显著性检验，检验的水平用α表示。

实际上，在气候诊断和气候数值模式的模拟与预测中，经常要对某些气候差异性问题进行检验。例如比较在某种条件下气候状态与一般情况下，或者与另一条件下的气候状态进行比较。在数值模拟中也经常要比较模拟场与观测场是否一致，或者在敏感性试验中比较正常状态下的试验与异常状态下的敏感性试验是否有差别等。如果它们不一致或有差别的话，在场中哪些地区有差别。而且要回答这些问题，还不能只根据一次试验就能作出定论，需要作多次试验。这就需要对多次试验结果作统计检验，然后作出客观的判断。又例如，在功率谱的显著性检验过程中，计算原假设功率谱的置信区间估计，当实际计算要检验的气候序列的功率谱值，超过假设功率谱置信区间上限时，表示落在否定域内，可以否定原功率谱的假设，认为所对应的序列周期振动是显著的。这种检验过程，实际上仍然属于传统的假设检验。应该按照统计学中的显著性检验过程进行检验和显著水平的设定。传统的错误称呼应该改变。

为了了解不同气候背景下，水文气象要素场是否有显著差异，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，通常取若干次（年）（样本）（设为 n）的距平场进行合成分析，得到样本 n 的合成场。但是合成场中仅看出距平符号均值（正负值）出现的区域，如果能够进一步对合成场中网格点（或不规则站点）的平均值，进行平均值的显著性检验，就能够明确合成场中的显著域。如果网格点要素遵从正态分布，并已经标准化，则可以用如下的统计量进行显著性检验：

$$t=\frac{\overline{X}-m}{s/\sqrt{n-1}} \qquad (2)$$

式中 $\overline{X}$、 s 分别为要素场中网格点 n 容量样本的平均值和标准差, m 为所有样本的平均值，由于要素值是标准化值，在所有样本的平均值应该为 0[10]。由此可以计算出要素场的合成场 t 值场，该场可以是表示合成场中某个格点的合成要素值与一般无条件情况的比较，是否有显著差异的统计显著性检验。为应用简便起见，可以把场中某格点的 t 值的绝对值大于 2.0 作为判定标准，这一标准近似对应为 5%的显著水平。

例如，使用在长江中游地区暴雨发生的气候异常环境研究中，附图 1 该地区给出 14 场暴雨过程第一日 500hPa 信号场的合成 t 值图，与 255 个无雨日 500hPa 信号场的合成 t 值图（附图 2）比较，可以发现它们的显著域（由绝对值 2.0 所包围的区域）分布形势有明显的差异。

参考文献

【1】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.

【2】黄嘉佑,统计动力分析与预报,北京，气象出版社，1993:pp.243。

【3】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:pp.298。

【4】《考研数学复习全书》编写组，线性代数与概率统计复习指南 ，南京，东南大学出版社，2000：pp,384

【5】《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：pp,384。

【6】袁卫等人编，《统计学》，北京，中国统计出版社，1994：pp,493。

【7】袁荫棠编，《概率论与数理统计》，2003：北京，中国人民大学出版 pp,284。

【8】何海燕，张红元，黄发贵等编，《商务统计》，2001：广东经济出版 pp,228。

【9】Freund,J.E.：Modern Elementary Statistics (Tenth Ediation), Prentice-Hall International (UK) Limited, London, 2001：pp588。

【10】陈家鼎，刘婉如，汪仁宫编，《概率统计讲义》，1980：北京，人民教育出版 pp,265。

【11】G.W.斯奈迪格等著，《数理统计方法》，1986：北京，科学出版社 pp,567。

【12】郑德如主编，《统计学》，上海，立信会计出版社，1996：pp,267。

【13】陈词成，冯虹主编，《新编统计学原理》，1996：北京，北京经济学院出版社 pp,308。

【14】何晓群，刘文卿编，《应用回归分析》，2001：北京，中国人民大学出版社 pp,273。

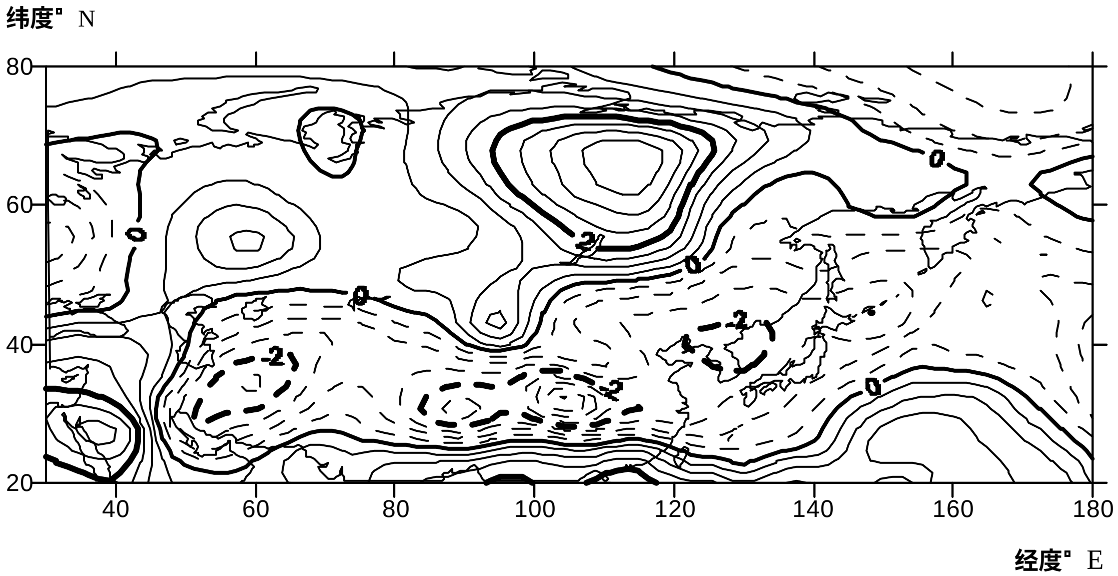

图 1　14 场暴雨过程第一日 500hPa 信号场的合成 *t* 值图

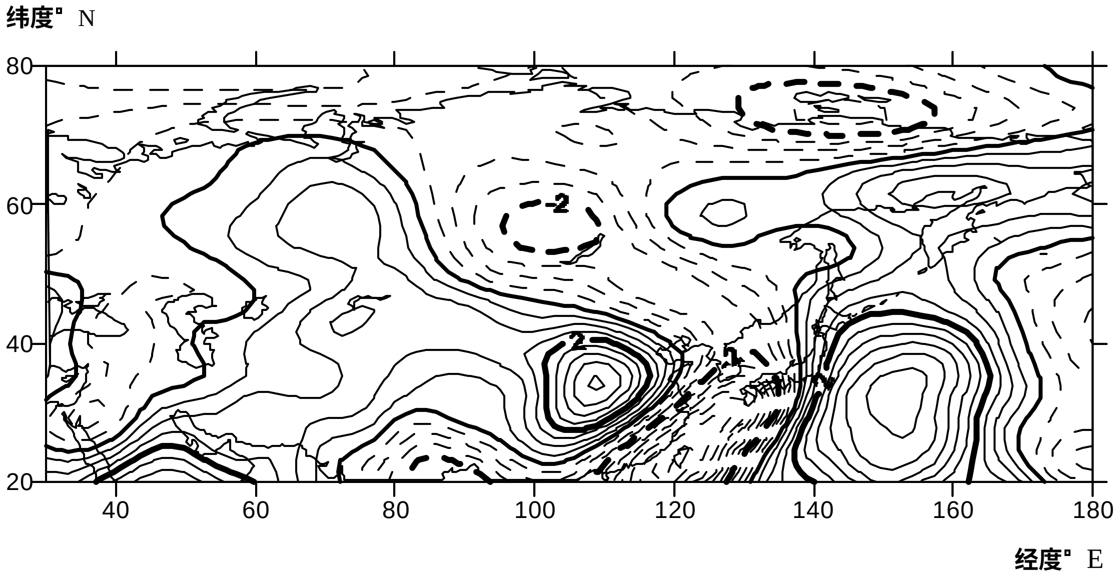

图 2　255 个无雨日 500hPa 信号场的合成 *t* 值图

气象水文统计分析中显著性检验的应用

作者： 黄嘉佑
作者单位： 北京大学物理学院大气科学系,北京,100871