气温基准平均值的差异及其统计检验

王效瑞 徐 敏
(安徽省气象局气象台, 合肥 230031)

摘要 文章利用安徽有记录以来的气温资料,讨论了单站及区域平均气温三个基准平均值的差异,并假定要素总体均方差未知,对三个基准年间的平均值差异进行了统计检验,现行气象业务以第二基准平均值(1961 ~ 1990 年)进行统计分析的适用性最好。
关键词 气温 基准平均值 统计检验

引言

在气象学的理论和实践中,气象因子的样本平均($\bar{x}$)、样本均方差(S_x)和样本距平($x_i - \bar{x}$)也许是最频繁使用的概念。样本的概念是相对于总体而言的,总体是统计分析对象的全体。总体确定后,其统计特征如总体平均(μ)和总体均方差(σ)应是不变的。由于总体中的个体很多,甚至是无限多,所以通常只能根据总体中有限多个样本资料去估计推断总体的 μ 和 σ。或者说,$\bar{x}$ 是 μ 的点估计值,S_x 是 σ 的点估计值[1]。显然,当样本容量(n)越大时,用样本反映总体的特征越全面,统计推断的精度越高。

气象业务工作中,通常取样本容量 $n \geqslant 30$,其资料统计参数视为总体的参数。目前,我国将 1961 ~ 1990 年的气象要素平均值作为基准值。

然而,样本 $x_1, x_2, ..., x_n$ 是从资料总体中随机抽取的,所以即使 n 不变,$x_1, x_2, ..., x_n$ 这些数值还是随机变化的,不同的样本就有不同的 $\bar{x}$,或者说 $\bar{x}$ 也是一个随机变量,也应服从某一概率分布。

在以往还未见到有人就样本容量不变,样本平均值的变化问题进行过研究。本文试图给出当样本容量 $n = 30$ 时,不同 $\bar{x}$ 及无偏估计方差 S_x^2 的变化,并在总体均方差未知时进行平均值差异的统计检验。在基准年平均值差异的统计检验基础上,亦即利用安徽温度资料对 20 世纪三个基准年平均值的变化进行分析,并对我国气象业务 1961 ~1990 年基准平均值的适用性作出评估。

1 资料与处理方法

1.1 资料选取

取安徽宿县、合肥、安庆和屯溪气象站建站至 2000 年的 1、4、7、10 月及年平均气温资料,计算各季代表月气温的三个基准平均值和无偏估计方差。安庆自 1951 年以来有完整的记录,其余 3 站均为 1953 年以来的记录。三个基准年的样本容量(n)安庆为 30,其余 3 站分别为 28、30 和 30,即 1953 ~1980 年、1961 ~1990 年和 1971 ~2000 年。

1.2 处理方法

利用安徽短期气候预测业务中全省及四个区域平均气温资料,计算淮北、江淮、沿江、江南及全省各基准年年平均气温、春季(3 ~ 5 月)、夏季(6 ~ 8 月)、秋季(9 ~11 月)、冬季(12 ~ 2 月)平均气温的平均值和无偏估计方差。由于各区域所选代表站的建站日期不同,所以各季节的气温资料截止于 1980 年的基准年 n 值在 27 ~ 30 之间,余者的基准年 n 值均为 30。

1.3 统计检验

基于将平均值差异的统计检验结果作为各基准平均值是否适用的判别标准,假定不同基准年间的气温资料序列是相互独立的,并且设定进行比较的两个资料总体遵从正态分布,其方差未知,但 $\sigma^2(1)$

安徽省气象局"安徽气候变化研究"课题资助
收稿日期:2002 年 8 月 19 日;定稿日期:2002 年 9 月 30 日
作者简介:王效端,男,1942 年生,高级工程师,主要从事气候诊断与基本气候学研究

$=\sigma^2(2)$。这样,我们可以根据总体中的两个随机样本 $x_1(1), x_2(1), ..., x_{n_1}(1)$ 和 $x_1(2), x_2(2), ..., x_{n_2}(2)$,检验假设 $H_0: \mu(1) = \mu(2)$。

统计量 $t =$

$$\frac{\bar{x}(1) - \bar{x}(2)}{\sqrt{\frac{(n_1 - 1)S_x^2(1) + (n_2 - 1)S_x^2(2)}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (1)$$

式(1)中, $\bar{x}(1) = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i(1)$ 和 $\bar{x}(2) = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i(2)$ 分别为两个总体的样本平均值; $S_x^2(1) = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i(1) - \bar{x}(1))^2$ 和 $S_x^2(2) = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_i(2) - \bar{x}(2))^2$ 分别为两个总体方差的无偏估计。

统计学证明,在总体与多年平均 μ_0 相等假设成立条件下,统计量 t 服从自由度为 $n-1$ 的 t 分布[1]。所以,利用样本资料计算出 $|t|$ 值,根据给定信度 α 与自由度 $(n_1 + n_2 - 2)$ 查 t 分布表可得到 t_α 值,若 $|t| \geqslant t_\alpha$ 则拒绝 H_0,若 $|t| < t_\alpha$,则接受 H_0。

2 气温基准平均值的变化与检验

2.1 单站气温

为叙述问题的方便,将建站至 1980 年气温平均值定义为第一基准平均值,1961~1990 年气温平均值为第二基准平均值,1971~2000 年气温平均值为第三基准平均值,建站至 2000 年的平均值为累年平均值,以下类同。表 1 给出宿县等 4 站 1、4、7、10 月和年平均气温基准平均值。

表 1 宿县等 4 站 1、4、7、10 月及年平均气温基准平均值 ℃

	宿县					合肥					安庆					屯溪				
	基准年			累年	最大差值	基准年			累年	最大差值	基准年			累年	最大差值	基准年			累年	最大差值
	1	2	3			1	2	3			1	2	3			1	2	3		
1 月	-0.2	0.4	0.8	0.3	1.0	2.1	2.4	2.6	2.4	0.5	3.5	3.8	4.0	3.8	0.5	3.9	4.0	4.2	4.0	0.3
4 月	14.4	14.6	15.0	14.7	0.6	15.5	15.5	15.9	15.7	0.4	16.0	16.1	16.6	16.3	0.6	16.2	16.2	16.4	16.3	0.2
7 月	27.4	27.4	27.3	27.4	0.2	28.3	28.1	28.1	28.2	0.2	28.8	28.7	28.7	28.8	0.1	28.1	27.9	27.7	27.9	0.4
10 月	15.7	16.0	16.2	16.0	0.5	17.0	17.1	17.2	17.1	0.2	18.0	18.1	18.3	18.2	0.3	17.3	17.6	17.7	17.5	0.4
年平均	14.4	14.5	14.7	14.6	0.3	15.7	15.7	15.8	15.8	0.1	16.5	16.5	16.7	16.6	0.2	16.2	16.2	16.3	16.3	0.1

由表 1 可知,宿县等 4 个单站四季代表月及年平均气温不同基准平均值的差值在 0.0~1.0℃之间。最大差值以宿县冬季(1 月)第一基准年与第三基准年相差 1.0℃为最大,以春季(4 月)第一基准年与第三基准年相差 1.0℃为最大,以春季(4 月)第一基准年与第三基准年相差 0.6℃次之。这种冬、春季第一和第三基准年间差值较大的现象是长江以北地区所共有的,而地处皖南的屯溪则相反,夏秋季的差值最大,冬春季反而较小。另外,宿县年平均气温第一和第三基准年间的差值最大,达 0.3℃。从各测站基准年平均值之间的差异及其变化趋势看,自 20 世纪 80 年代中期开始的暖冬和 90 年代春季气温偏高的现象在基准平均值中已有明显反映。另外,1994~1999 年秋季气温持续偏高,也导致第三基准年平均值偏高。表 2 给出宿县等 4 站 1、4、7、10 月和年平均气温各基准年无偏估计方差 S_x^2 与统计量 $|t|$。

从表 2 看出,各测站年平均气温各基准年 S_x^2 远小于四季代表月的值;就季节而言,各测站春季(4 月)和秋季(10 月)基准年 S_x^2 明显小于冬季(1 月)和夏季(7 月);各测站冬季(1 月)第一基准年的 S_x^2 值明显较大。沿江地区安庆和皖南地区屯溪各基准年 S_x^2 之间的稳定性优于宿县和合肥;另外,各季及年的 S_x^2 除淮北地区宿县外,均以第二基准年的 S_x^2 为最小。

根据表 2 中统计量 t,取信度 α 为 0.05,自由度为 $n_1 + n_2 - 2$,查表得 $t_\alpha = 2.002$。显然,除宿县冬季、春季和安庆春季的第一、第三基准年间 $t > t_\alpha$ 未通过检验外,其他测站各季及年平均气温各基准平均值均通过了检验。或者说合肥和屯溪各季及年平均气温的三个基准平均值之间未有明显差异,用任

意基准平均值作统计分析都是适当的。但从表 2 中 t 值的变化来看,各测站用第二基准年平均值作统计分析最好,因为它与第一和第三基准年之间的统计量 t 值更小,仅在 0.0~1.8157 之间。

表 2　宿县等 4 站 1、4、7、10 月及年平均气温各基准无偏估计方差 S_x^2 与统计量$|t|$

	时间	S_x^2 第一基准年	S_x^2 第二基准年	S_x^2 第三基准年	$t_{1,2}$	$t_{2,3}$	$t_{1,3}$
宿县	1 月	2.4763	1.3612	1.2851	1.5747	1.6156	2.9314
	4 月	0.9632	0.9133	1.0313	0.7865	1.5707	2.2857
	7 月	1.7503	1.5079	1.5444	0.5974	0.3034	0.2969
	10 月	1.2417	1.0397	1.3617	1.0795	0.7067	1.6824
	年	0.3385	0.1822	0.3380	0.7502	1.5186	1.9634
合肥	1 月	2.4863	1.5425	1.5791	0.8108	0.6250	1.3405
	4 月	0.9432	1.7265	1.0688	0.0	1.3115	1.5152
	7 月	1.6739	1.6191	1.7272	1.5935	0.0	0.5838
	10 月	0.9732	0.9385	1.1212	0.3896	0.3817	0.7435
	年	0.2360	0.2074	0.3239	0.0	0.7513	0.7174
安庆	1 月	1.5579	1.5350	1.5471	0.9340	0.6238	1.5538
	4 月	1.0688	0.9465	1.0964	0.3858	1.9157	2.2330
	7 月	1.5038	1.4157	1.6124	0.3205	0.0	0.3102
	10 月	1.2017	0.9849	1.1444	0.3704	0.7505	1.0726
	年	0.1846	0.1619	0.2835	0.0	1.6407	1.6013
屯溪	1 月	2.2263	1.9833	1.5342	0.2627	0.5737	0.8357
	4 月	1.0488	0.7822	0.8858	0.0	0.8478	0.7752
	7 月	1.1099	1.0412	1.2196	0.7345	0.7283	1.4099
	10 月	0.9993	0.8157	1.1281	1.2010	0.3928	1.4749
	年	0.1432	0.1081	0.1454	0.0	1.0881	1.0020

注:统计量 t 的下标表示基准年

2.2　区域平均气温

包括若干代表测站的区域要素平均值滤掉了单个测站的局地变化,区域要素平均值也是现行短期气候预测及评分的主要依据,为此,本文讨论了安徽全省及四个区域四季、年气温各基准平均值的变化与平均值差异的统计检验。表 3 给出安徽全省及各区域四季及年平均气温各基准年平均值及其最大差值。

由表 3 可知,年平均气温各基准年间的平均值差异在 0.0~0.3 ℃之间,其中以江南地区各基准年平均值的差异最小,为 0.0 ℃;淮北和沿江地区为最大,为 0.3 ℃;值得注意的是,全省及各区域第一基准年与第三基准年的差值均大于第一与第二和第二与第三基准年的差值,或者说年平均气温第三基准年的平均值最大,亦即 20 世纪 80 和 90 年代气温记录在资料序列中有明显的贡献。

冬季气温各基准年间的平均值差异最大,为 0.0~0.7 ℃之间;淮北地区第一基准年与第三基准年之间的差值最大,为 0.7 ℃;江南地区第一基准年与第三基准年的差值最小,为 0.2 ℃;江淮和沿江地区居中,均为 0.4 ℃。与年平均气温一样,冬季气温第三基准年的平均值最大,但江南地区例外,却为最小。除江南地区外的其他地区,冬季气温第三基准年平均值为最高的事实表明,安徽长江以北地区自 20 世纪 80 年代中期开始的暖冬现象已在气温资料序列中有明显贡献。

春季气温各基准年间的平均值差异居中,为 0.0~0.4 ℃,江南地区差异最小,为 0.0~0.1 ℃,其他地区为 0.2~0.4 ℃;第三基准年平均值最大,第二基准年平均值居中,第一基准年平均值最小。这大概与 20 世纪 90 年代春季气温偏高有关。

夏季气温各基准年间的平均值差异仅次于冬季,为 0.1~0.6 ℃,沿江地区第二基准年与第三基准年平均值相差 0.6 ℃,为最大,其他区域在 0.1~0.3 ℃之间。值得注意的是,与春季相反,全省及各

表 3 安徽全省及各区域四季及年平均气温各基准年平均值及其最大差值 ℃

		第一基准年	第二基准年	第三基准年	累年	最大差值
淮北	年	14.7	14.8	15.0	14.9	0.3
	冬	2.0	2.1	2.7	2.4	0.7
	春	14.5	14.7	14.9	14.7	0.4
	夏	26.8	26.7	26.6	26.7	0.2
	秋	15.7	15.8	16.0	15.9	0.3
江淮	年	15.6	15.7	15.8	15.7	0.2
	冬	3.6	3.7	4.0	3.8	0.4
	春	14.9	15.1	15.3	15.1	0.4
	夏	27.1	26.9	26.8	26.9	0.3
	秋	16.8	16.7	16.9	16.9	0.1
沿江	年	16.1	16.2	16.4	16.3	0.3
	冬	24.6	4.5	4.9	4.8	0.4
	春	15.4	15.36	15.8	15.6	0.4
	夏	27.3	27.1	26.7	27.0	0.6
	秋	17.5	17.4	17.5	17.5	0.1
江南	年	15.8	15.8	15.8	15.8	0.0
	冬	4.5	4.4	4.3	4.6	0.3
	春	15.3	15.3	15.4	15.4	0.1
	夏	26.6	26.4	26.3	26.5	0.3
	秋	16.8	16.9	16.9	16.9	0.1
全省	年	15.6	15.6	15.8	15.7	0.2
	冬	3.7	3.7	4.1	3.9	0.4
	春	15.0	15.2	15.4	15.2	0.4
	夏	26.9	26.8	26.7	26.8	0.2
	秋	16.7	16.7	16.8	16.8	0.1

区域均以第一基准年平均值为最大,第二基准年居中,第三基准年平均值最小。这反映出 20 世纪 80 和 90 年代的凉夏现象确实存在[2]。

秋季气温各基准年间的平均值差异与春季相近,为 0.0 ~ 0.3 ℃,除淮北地区第一基准与第三基准年平均值相差 0.3 ℃为最大外,其他区域各基准年间的平均值相差仅为 0.0 ~ 0.1 ℃,也是四季中各基准年平均值差异最小的季节。另外,第三基准年平均值稍大于第一和第三基准年的平均值,可能与 20 世纪 90 年代大多数年份秋季气温偏高有关。

表 4 给出安徽全省及各区域四季及平均气温各基准年的 S_x^2 与统计量 $|t|$。从表 4 可以看出,安徽全省及四个区域年平均气温各基准年无偏估计方差远小于各季平均气温的无偏估计方差;就各季而言,冬季气温各基准年的 S_x^2 最大,各区域在 0.8031 ~ 1.4040 之间,且各区域均以第一基准年的 S_x^2 最大,而第二与第三基准年的 S_x^2 十分接近;作为过渡季节春、秋两季气温各基准年的 S_x^2 最小,在 0.2937 ~ 0.5746 之间,各区域虽然其最大值在第一和第三基准年都有出现,但其 S_x^2 最小值均出现在第二基准年;夏季气温各基准年的 S_x^2 居中,在 0.4870 ~ 0.8269 之间,各区域均以第一基准年的 S_x^2 最大,第三基准年的 S_x^2 最小。

表 4 中 t 值未通过平均值差异统计检验则表明所在区域年或季节平均气温两基准年平均值之间有明显差异。本文取信度 $\alpha = 0.05$,自由度为 $n_1 + n_2 - 2$,并查 t 分布表得到 t_α 值。从表 4 中的 t 值来看,除全省和淮北地区的冬季平均气温外,全省及其他地区季节及年平均气温用第二基准年平均值作统计分析是最为适当的,它与第一和第三基准年之间的 t 值相对较小,或者说,第一与第二、第二与第三基准年平均值之间未有明显差异。

表 4　安徽全省及各区域四季与年平均气温各基准年的 S_x^2 与统计量 $|t|$

		S_x^2			$t_{1,2}$	$t_{2,3}$	$t_{1,3}$
		第一基准年	第二基准年	第三基准年			
淮北	年	0.2600	0.1766	0.2870	0.8104	1.6090	2.1598 *
	冬	1.4040	0.9177	0.8031	0.3571	2.5052 *	2.5650 *
	春	0.4947	0.4510	0.5290	1.1179	1.1068	2.1470 *
	夏	0.7830	0.6160	0.4870	0.4600	0.5200	0.9660
	秋	0.3219	0.3430	0.4404	0.6658	1.2376	1.8634
江淮	年	0.2377	0.1866	0.2986	0.8212	0.7863	1.4515
	冬	1.1458	0.8694	0.8143	0.3830	1.2660	1.5540
	春	0.5081	0.2937	0.4681	1.2158	1.2547	2.2002 *
	夏	0.8269	0.8012	0.6714	0.8514	0.4513	1.3324
	秋	0.5746	0.3524	0.5355	0.0	0.5814	0.5157
沿江	年	0.1326	0.1317	0.2212	1.0373	1.8440	2.6708 *
	冬	0.9031	0.8401	0.8644	0.4115	1.6779	1.2260
	春	0.4200	0.3205	0.5393	1.2731	1.1813	2.2371 *
	夏	0.7728	0.7148	0.6653	0.8909	1.8648	2.7192 *
	秋	0.5612	0.4793	0.4949	0.5328	0.5549	0.0
江南	年	0.1154	0.1081	0.1487	0.0	0.0	0.0
	冬	0.9976	0.9440	0.9143	0.3829	0.4018	0.7719
	春	0.4978	0.4454	0.3813	0.0	0.6024	0.5757
	夏	0.6056	0.5845	0.5141	0.9960	0.5225	1.5409
	秋	0.5358	0.5035	0.5657	0.5283	0.0	0.5128
全省	年	0.1612	0.1367	0.2241	0.0	1.8238	1.7101
	冬	1.0170	0.7830	0.7766	0.0	2.4390 *	1.5980
	春	0.3799	0.2651	0.5071	1.3550	1.2469	2.3041 *
	夏	0.7061	0.6431	0.5455	0.4678	0.5024	0.9723
	秋	0.3374	0.3409	0.5106	0.0	0.5935	0.5824

注:带“ * ”的统计量 $|t|$ 为未通过检验

3　讨论与结论

(1) 单站与区域平均气温资料均表明,冬季气温各基准年间的平均值差异最大,居各季节之首,其第三基准年平均值为最大的事实表明,安徽长江以北地区自 20 世纪 80 年代中期开始的暖冬现象已有明显贡献;夏季气温各基准年间的平均值差异居次,其第一基准年的平均值最大,第二基准年次之,第三基准年最小的分布特征与春季的分布完全相反,这可能分别与 20 世纪 80、90 年代的凉夏和 90 年代暖春现象有关;秋季气温第三基准年平均值稍大于第一和第二基准年平均值,可能与 20 世纪 90 年代多数年份秋季气温明显偏高相联系。

(2) 基准年间平均值差异的统计检验表明,单站与区域平均气温的检验结果不尽一致:合肥各季节及年平均气温三个基准年平均值间未有明显差异,用任意一个基准年平均值作统计分析都是适当的,但从统计量 t 值的变化来看,除年平均气温外,各季节气温均以用第二基准平均值作统计分析最好;淮北地区冬季气温第二和第三基准年间,第一和第三基准年间,春季气温第一和第三基准年间,年平均温度第一和第三基准年间,江淮地区春季气温第一和第三基准年间,沿江地区年平均气温、春季气温、夏季气温的第一、第三基准年间,全省冬季气温第二、第三基准年间,春季气温第一、第三基准年间未通过统计检验,从 t 值变化来看,除全省和淮北地区冬季气温外,全省及其他地区季节与年平均气温用第二基准平均值进行统计分析最好。

参考文献

1　施能.气象统计预报中的多元分析方法.北京:气象出版社,1992
2　王效瑞,方茸.近 50 年来安徽气候及其变化研究.减灾与发展,1999,(3):41 - 44

(下转 57 页)

(上接 37 页)

THE DIFFERENT TEMPERATURE VALUES OF STANDARD YEAR AVERAGE AND ITS STATISTICAL TEST

Wang Xiaorui Xu Min

(Anhui Meteorological Observatory, Hefei 230031, China)

Abstract: The different values of standard temperature year average using the meteorological data in Anhui Province are discussed. The statistical test of these differences of element average is made on the supposition that standard square deviation for whole data is unknown. The current year average based on 1960 - 1990 data for China meteorological operation is suitable to the statistical analysis.

Key words: temperature, standard average, statistical test