第5卷 第2期　　大 气 科 学　　Vol. 5, No. 2
1981年6月　　SCIENTIA ATMOSPHERICA SINICA　　June, 1981

方差不相等的双样本回归分析

叶家东　范蓓芬
（南京大学气象系）

提　要

在人工降水试验效果的区域雨量回归分析中，当余方差不相等时，通常的多个事件检验法就不适用．本文在 Welch 检验法的基础上，提出了余方差不相等时适用的双样本回归分析方法，并且指出，我们根据双样本回归分析提出的统计量(19)，当余方差相等时要比多个事件检验法的统计量(6)有效．

利用双样本回归分析，对福建古田地区人工降水随机试验的效果进行评价．结果表明，人工降水引起的目标区平均相对增雨量随着对比区自然雨量的增大而减小．当对比区为小雨($x'<2$毫米/小时)时，催化效果显著($\alpha=0.05$)，相对增雨量达20—81%，但绝对增雨量不大，只有0.44—1.15(毫米/3小时)，而当对比区雨量$x'>2.8$(毫米/小时)时，催化效果就不显著．

人工影响天气试验效果的统计检验，通常都采用t-检验法或秩和检验法[1]．这类检验方法有一个共同的前提条件，就是假定待检验的二个样本所抽取的总体方差是相等的．但是，实际上这个条件并不总是满足的．在福建古田地区的人工降水试验中[2]，我们发现人工影响后的相对增雨量有随自然雨量强度增大而减小的趋势（表1）．这种趋势在美国佛罗里达[3]，苏联乌克兰[4]以及以色列[5]等地的试验中均有所发现．从统计上看，这意味着人工影响不仅改变了雨量的平均值，而且也改变了雨量的方差．如果这是真的，则通常的检验方法就不适用．同样，在区域雨量的回归分析中，利用多个事件检验法[6]检验平均增雨量的显著性时，也是假定了供分析的两组样本（称催化样本和对比样本）所抽取的总体余方差相等，如果这个条件不满足，多个事件检验法也不适用．

表 1

对比区雨量强度(毫米/小时)	$x'\leqslant 1$	$1<x'\leqslant 3$	$x'>3$
样本容量	18	22	12
目标区相对增雨量(%)	125.2	24.4	−1.0

方差不相等条件下的双样本检验问题，至今尚无一般性的解法，只有一些近似方法可供利用[7]，Behrens 曾提出过一个检验法，所以这种问题常称为 Behrens-Fisher 问题．对于人工影响天气试验，常采用 Welch 检验法[8]．至于余方差不相等条件下的双样本回归分

1979年9月28日收到修改稿．

析，尚未见到有人讨论过．本文在 Welch 检验法的基础上，提出一种余方差不相等条件下的双样本回归分析方法，并对余方差相等时适用的多个事件检验法作出修正．

一、Welch 检验法

设二独立样本 $x_{11}, x_{12}, \cdots, x_{1n_1}$ 和 $x_{21}, x_{22}, \cdots, x_{2n_2}$ 服从正态分布，总体平均值为 μ_1 和 μ_2，方差 $\sigma_1^2 \neq \sigma_2^2$，相应的样本统计量为 $\bar{x}_1$，$\bar{x}_2$ 和 s_1^2，s_2^2．现在要根据上述样本数据检验假设 $H_0: \mu_1 = \mu_2$．为此，需要决定统计量

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{1}$$

的分布．令

$$s^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

易知

$$z = \frac{\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\dfrac{s^2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

的分子是 $N(0, 1)$ 变量．如果分母是一个 χ-变量的倍数，则 z 是 t-变量．但现在 $\sigma_1^2 \neq \sigma_2^2$，所以分母不是一个 χ-变量的倍数． Welch 检验法的要点是以一个 χ^2-变量的倍数 s'^2 来近似代替 s^2，选择 s'^2 的自由度 ν'，使得 s'^2 与 s^2 的数学期望及方差都相等．这样，z 就近似地成为自由度是 ν' 的 t-变量．据此可以导出

$$\nu' = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{\sigma_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{\sigma_2^2}{n_2}\right)^2} \tag{2}$$

实际上 σ_1^2 与 σ_2^2 是不知道的，以相应的样本估计量 s_1^2 与 s_2^2 近似代替，就得

$$\nu' \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} \tag{3}$$

ν' 称为有效自由度．于是，在原假设成立的前提下，统计量

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4}$$

可近似看作是自由度为 ν' 的 t-变量，从而可利用 t-检验法加以检验．

Welch 检验法与双样本 t-检验的比较，由(4)式和通常的双样本 t-检验的统计量

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \tag{5}$$

比较可知，它们的区别在于分母不同. 设 $\frac{s_1^2}{s_2^2}=F$，可得

$$\frac{z^2}{t^2}=\frac{\frac{\nu_1 s_1^2+\nu_2 s_2^2}{\nu_1+\nu_2}\left(\frac{1}{\nu_1+1}+\frac{1}{\nu_2+1}\right)}{\frac{s_1^2}{\nu_1+1}+\frac{s_2^2}{\nu_2+1}}=\frac{(F\nu_1+\nu_2)(\nu_1+\nu_2+2)}{(\nu_1+\nu_2)(1+F+\nu_1+F\nu_2)}$$

如果两个样本容量相等，则 $\nu_1=\nu_2=\nu$，这时

$$\frac{z^2}{t^2}=1$$

因此，对于容量相等的样本，统计量 z 和 t 数值相同，此时两种检验法的区别仅在于 t 的自由度是 2ν，而 z 的自由度为

$$\nu'=\nu\frac{(F+1)^2}{F^2+1}=\nu\left(1+\frac{2F}{F^2+1}\right)$$

当 $F=1$ 时，$\nu'=2\nu$，z 还原为 t. 这就是方差相等的情形；当 $F\neq 1$ 时，ν'比 2ν 小，但对于 $0.333<F<3$，自由度的偏差不超过20%. 所以当样本容量相等，而样本方差相差不大时，二种方法的差异是小的. 但是，当样本容量不相等时，例如当 $\nu_1\gg\nu_2$ 时，$z^2/t^2\to F$，此时用 t 统计量代替 z 统计量所产生的误差近似地随着 $\sqrt{F}$ 增大而增加，这种情况下就不宜用双样本 t-检验法.

二、双样本回归分析

通常，在人工降水效果检验的区域雨量回归分析中，利用统计量[6]

$$t=\frac{\bar{y}_k-\bar{\hat{y}}_k}{\sqrt{\frac{\sum_{j=1}^{n}(y_j-\hat{y}_j)^2}{n-2}\left[\frac{1}{k}+\frac{1}{n}+\frac{(\bar{x}_k-\bar{x}_n)^2}{\sum_{j=1}^{n}(x_j-\bar{x}_n)^2}\right]}} \tag{6}$$

检验平均差值

$$\bar{d}_k=\bar{y}_k-\bar{\hat{y}}_k$$

的显著性时，也是假定了供分析的两组样本(催化样本和对比样本)所代表的倚变量(目标区雨量)的余方差是相等的. 但是，从福建古田试验的回归分析发现，这个条件并不满足(表 2). 由表 2 可见，催化样本与对比样本的余方差有显著差异，显著度达到 0.05. 这时就不宜用上述多个事件检验公式(6)进行催化效果的统计检验. 为此，我们作如下双样本回归分析.

令催化单元目标区雨量 y 倚对比区雨量 x 的区域回归方程为

表 2　回归分析中催化样本与对比样本的余方差相等性检验

	样本容量	$s^2_{余}$	$F=\frac{s^2_{2余}}{s^2_{1余}}$	α
催化样本(1)	$k=50$	0.0270	1.6556	<0.05
对比样本(2)	$n=51$	0.0447		

$$\hat{y}_1 = a_1 + b_1 x \tag{7}$$

而将对比单元的区域雨量回归方程写成

$$\hat{y}_2 = a_2 + b_2 x \tag{8}$$

相应的一元线性正态回归的结构模型是

$$y_1 = \alpha_1 + \beta_1 x + \varepsilon_1 \tag{7'}$$

$$y_2 = \alpha_2 + \beta_2 x + \varepsilon_2 \tag{8'}$$

其中 $\alpha_1 + \beta_1 x = y_{10}$, $\alpha_2 + \beta_2 x = y_{20}$ 是相应的总体回归值，而 $\varepsilon_1 \sim N(0,\sigma_1)$, $\varepsilon_2 \sim N(0,\sigma_2)$, $\sigma_1 \neq \sigma_2$.

由回归分析知[9]

$$\hat{y}_1 \sim N\left(\alpha_1 + \beta_1 x,\ \sigma_1\sqrt{\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum_{i=1}^{k}(x_i-\bar{x}_1)^2}}\right) \tag{9}$$

$$\hat{y}_2 \sim N\left(\alpha_2 + \beta_2 x,\sigma_2\sqrt{\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum_{j=1}^{n}(x_j-\bar{x}_2)^2}}\right) \tag{10}$$

其中 k, n 分别为催化样本和对比样本的容量，而

$$\bar{x}_1 = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \bar{x}_2 = \frac{1}{n}\sum_{j=1}^{n} x_j$$

现在要检验假设 $H_0: y_{10} = y_{20}$. 为此，作统计量

$$z_{\hat{y}} = \frac{\hat{y}_1 - \hat{y}_2 - (y_{10} - y_{20})}{\sqrt{\left(\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum_{i=1}^{k}(x_i-\bar{x}_1)^2}\right)s^2_{1余} + \left(\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum_{j=1}^{n}(x_j-\bar{x}_2)^2}\right)s^2_{2余}}} \tag{11}$$

其中

$$s^2_{1余} = \frac{1}{k-2}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2, \quad s^2_{2余} = \frac{1}{n-2}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2$$

分别是 σ_1^2, σ_2^2 的样本估计量.

仿照 Welch 的处理方法，在原假设成立的前提下，统计量

$$z_{\hat{y}} = \frac{\hat{y}_1 - \hat{y}_2}{\sqrt{\left(\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum_{i=1}^{k}(x_i-\bar{x}_1)^2}\right)s^2_{1余} + \left(\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum_{j=1}^{n}(x_j-\bar{x}_2)^2}\right)s^2_{2余}}} \tag{12}$$

表3 样本数据和基本统计量（x_i'，y_i' 与 x_j'，y_j' 均系三小时区域面积平均雨量）

催化样本(1)		对比样本 (2)		催化样本 (1)		对比样本 (2)		$k=50$
$x_i=\sqrt[4]{x_i'}$	$y_i=\sqrt[4]{y_i'}$	$x_j=\sqrt[4]{x_j'}$	$y_j=\sqrt[4]{y_j'}$	$x_i=\sqrt[4]{x_i'}$	$y_i=\sqrt[4]{y_i'}$	$x_j=\sqrt[4]{x_j'}$	$y_j=\sqrt[4]{y_j'}$	
0.9299	1.1815	1.4791	1.1659	1.1435	1.1469	1.9054	1.7248	$\bar{x}_1=1.5510$
1.6053	2.0774	0.7586	0.7640	1.2783	1.4333	1.6754	1.6807	$\bar{y}_1=1.6017$
1.3180	1.5464	1.7761	1.5983	1.2421	1.3437	1.6364	1.6022	$\sum_{i=1}^{k}(x_i-\bar{x}_1)^2=6.6241$
1.2510	1.5767	1.0193	0.5883	1.6604	1.6984	1.2009	0.9487	$\sum_{i=1}^{k}(y_i-\bar{y}_1)^2=5.6302$
0.9417	0.9693	0.7259	0.6785	1.6482	1.5035	1.8974	1.5799	$\sum_{i=1}^{k}(x_i-\bar{x}_1)(y_i-\bar{y}_1)=5.3559$
1.5930	1.7892	1.5694	1.3945	1.7556	1.7184	1.4557	1.6335	$a_1=0.3477$
1.2603	1.4526	1.6081	1.4180	2.2399	2.0148	1.7321	1.9871	$b_1=0.8085$
1.3244	1.5093	0.9304	1.0475	1.2866	1.3150	1.1402	1.3226	$r_1=0.8771$
1.9145	1.8012	1.7538	1.8647	1.4954	1.3301	1.5376	1.6505	$\hat{y}_1=0.3477+0.8085x$
2.3267	2.3500	1.6653	1.7979	0.8764	0.8144	1.0819	1.1158	$s_{1余}^2=\frac{1-r_1^2}{k-2}\sum_{i=1}^{k}(y_i-\bar{y}_1)^2=0.0270$
1.8310	1.7959	1.5265	1.3831	1.2052	1.2081	0.6999	0.6925	

1.3660	1.4114	1.4664	1.7724	1.4508	1.3619	1.0954	1.1229	
1.6795	1.4766	1.6742	1.6776	1.6089	1.4459	1.4653	1.6130	
1.2992	1.4440	1.2398	0.8721	1.0574	1.2524	1.1334	1.3888	
2.1723	2.0334	1.6232	2.0230	1.6213	1.2095	1.4169	1.4878	$n = 51$
1.2271	1.3975	1.5645	1.4839	1.4724	1.4748	0.9047	0.8098	$\bar{x}_2 = 1.4246$
2.4405	2.3001	1.4195	1.5849	1.9017	1.8915	0.6005	0.6687	$\bar{y}_2 = 1.4162$
1.3116	1.4678	1.5725	1.5210	2.2970	2.1719	1.3989	1.6647	$\sum_{j=1}^{n}(x_j - \bar{x}_2)^2 = 5.8354$
1.6274	2.0980	1.5789	1.5551	1.8554	2.1250	1.6598	1.6443	$\sum_{j=1}^{n}(y_j - \bar{y}_2)^2 = 6.8787$
1.3117	1.5625	1.3971	1.4281	1.6948	1.8241	1.8367	1.8701	$\sum_{j=1}^{n}(x_j - \bar{x}_2)(y_j - \bar{y}_2) = 5.2237$
1.4116	1.4778	1.4124	1.6663	2.0457	1.8322	1.7110	1.7024	$a_2 = 0.1409$
1.6271	1.8331	1.2912	0.8367	1.9607	1.8933	1.2854	1.6022	$b_2 = 0.8952$
1.4998	1.5799	1.0287	1.2151	1.7756	1.7349	1.5716	1.2549	$r_2 = 0.8245$
1.6771	1.9591	1.6329	1.4968	1.2165	1.2200	1.4142	1.7734	$\hat{y}_2 = 0.1409 + 0.8952x$
1.3519	1.4589	2.0236	1.7459			1.3888	1.3775	$s_{2*}^2 = \frac{1 - r_2^2}{n - 2}\sum_{j=1}^{n}(y_j - \bar{y}_2)^2 = 0.0447$
1.4589	1.5696	2.0708	1.6266					

可近似地看作是自由度为

$$\nu_{\hat{y}}=\frac{\left[\left(\frac{1}{k}+\frac{(x-\bar{x}_1)^2}{\sum(x_i-\bar{x}_1)^2}\right)s_{1余}^2+\left(\frac{1}{n}+\frac{(x-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)s_{2余}^2\right]^2}{\frac{1}{k-2}\left[\left(\frac{1}{k}+\frac{(x-\bar{x}_1)^2}{\sum(x_i-\bar{x}_1)^2}\right)s_{1余}^2\right]^2+\frac{1}{n-2}\left[\left(\frac{1}{n}+\frac{(x-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)s_{2余}^2\right]^2} \quad (15)$$

的 t-变量.

表 3 是福建古田地区三年人工降水试验中目标区和对比区雨量的基本资料. 计算表明,变量取雨量的四次方根值的正态性很好,拟合度达 0.94, 所以分析时统计变量均取雨量的四次方根值。表 3 还列入有关的统计量. 我们根据上述方法对古田地区试验的效果进行双样本回归分析, 计算结果列于表 4. 催化样本(1)和对比样本(2)的区域雨量回归线及实测雨量散布图见图 1;催化效果随对比区自然雨量而变化的关系见图 2.

表 4 试验效果的双样本回归分析

对比区自然雨量		目标区平均增雨量			显著性检验		
x	x'(毫米)	$\hat{y}_1-\hat{y}_2$	绝对增雨量(毫米)	相对增雨量(%)	$z_{\hat{y}}$	$\nu_{\hat{y}}$	α(单边)
0.8	0.41	0.1376	0.4388	81.38	1.6801	94.76	<0.05
1.0	1.00	0.1201	0.6345	55.05	1.8766	95.81	<0.05
1.3	2.86	0.0941	0.9308	32.12	2.2193	96.03	<0.025
1.5	5.06	0.0768	1.0840	22.37	1.9845	93.63	0.025
1.5510	5.79	0.0723	1.1101	20.09	1.8682	87.84	<0.05
1.7	8.35	0.0594	1.1524	15.08	1.3026	84.33	<0.10
1.8	10.50	0.0507	1.1393	12.07	0.9584	82.56	<0.20
2.0	16.00	0.0334	0.9877	7.10	0.4834	82.50	>0.25

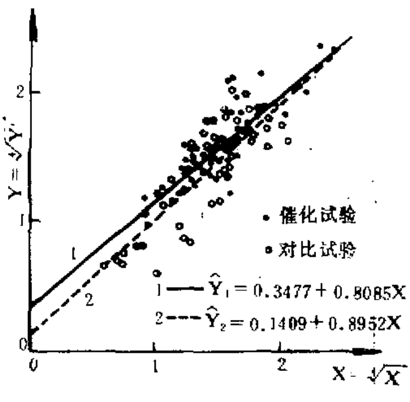

图 1 区域雨量回归线和实测雨量散布图

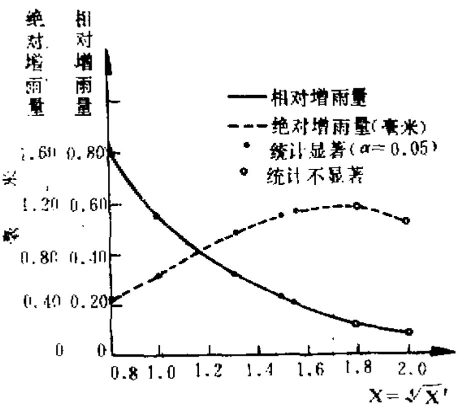

图 2 增雨效果随对比区自然雨量而变化的关系曲线

由所列图表可见: (1)平均相对增雨量是随对比区雨量的增加而减小的, 当 $x>1.7$ (相当于雨量 8.35 毫米/3 小时)时,催化效果就不显著了($\alpha=0.05$);(2)平均绝对增雨量

介于 0.44—1.15（毫米/3 小时）之间，对比区雨量 $x \leqslant 1.7$ 时是随着 x 增加而增大的，$x=1.7$ 时达到最大，约 1.15(毫米/3 小时)，但此时显著度不高（$\alpha<0.10$）．看来这种小火箭催化冷云降雨的方法在小雨时效果较好，可增加 20—81%，但绝对增雨量并不大，只有 1（毫米/3 小时）左右．

与多个事件检验法比较，多个事件检验法是检验 k 次试验的平均效果的．为了比较起见，在(12)式中取 $x=\bar{x}_1=\bar{x}_k$，于是 $\hat{y}_1=\hat{y}_k$，$\bar{y}_k=\hat{y}_2$．令

$$\frac{s_{1余}^2}{s_{2余}^2}=F,$$

且因 $\bar{x}_2=\bar{x}_n$，则由(12)和(6)式可得

$$\frac{t^2}{z_{\hat{y}}^2}=\frac{\left[\frac{1}{k}\cdot F+\frac{1}{n}+\frac{(\bar{x}_1-\bar{x}_n)^2}{\sum(x_j-\bar{x}_n)^2}\right]}{\left[\frac{1}{k}+\frac{1}{n}+\frac{(\bar{x}_k-\bar{x}_n)^2}{\sum(x_j-\bar{x}_n)^2}\right]} \tag{14}$$

如果 $F=1$，则 $\frac{t^2}{z_{\hat{y}}^2}=1$，$z_{\hat{y}}$ 还原为 t；如果 $F<1$，则 $t<z_{\hat{y}}$；反之，当 $F>1$ 时 $t>z_{\hat{y}}$，即人工影响后目标区雨量指标的余方差变小时，统计量 $z_{\hat{y}}$ 比 t 有效；反之，则差．

至于自由度，如果 $F=1$，且 $n=k$，则由(13)式可得

$$\nu_{\hat{y}}=(n-2)\left\{1+\left[1+\frac{\left(\frac{(\bar{x}_1-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)^2}{\frac{2}{n}\left(\frac{1}{n}+\frac{(\bar{x}_1-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)}\right]^{-1}\right\}\approx 2(n-2)=2\nu$$

因为在一般情况下，

$$\left[\frac{(\bar{x}_1-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right]^2 与 \frac{2}{n}\left[\frac{1}{n}+\frac{(\bar{x}_1-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right]$$

相比，是二阶小项．在这种情况下，$z_{\hat{y}}$ 的自由度在 $n=k$，$F=1$ 时几乎是多个事件检验法中 t 的自由度的两倍．

我们认为，问题在于多个事件检验法没有充分利用样本资料所提供的全部信息，所以它不是一个最有效的统计量．对此，我们作如下方差相等时的双样本回归分析．

在 $\sigma_1=\sigma_2=\sigma$ 的情况下，当假设 $H_0: y_{10}=y_{20}=y_0$ 成立时，对任一固定的 x 值而言，由(9)式和(10)式可知统计量

$$\begin{aligned} u&=\frac{\hat{y}_1-\hat{y}_2}{\sqrt{\sigma^2\left(\frac{1}{k}+\frac{(x-\bar{x}_1)^2}{\sum(x_i-\bar{x}_1)^2}\right)+\sigma^2\left(\frac{1}{n}+\frac{(x-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)}}\\ &=\frac{\hat{y}_1-\hat{y}_2}{\sigma\cdot\sqrt{\left(\frac{1}{k}+\frac{(x-\bar{x}_1)^2}{\sum(x_i-\bar{x}_1)^2}\right)+\left(\frac{1}{n}+\frac{(x-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)}} \end{aligned} \tag{15}$$

服从 $N(0,1)$ 分布，于是

$$t=\frac{\hat{y}_1-\hat{y}_2}{s_{余}\cdot\sqrt{\left(\frac{1}{k}+\frac{(x-\bar{x}_1)^2}{\sum(x_i-\bar{x}_1)^2}\right)+\left(\frac{1}{n}+\frac{(x-\bar{x}_2)^2}{\sum(x_j-\bar{x}_2)^2}\right)}} \tag{16}$$

是 t-变量，其中 s_*^2 是余方差 σ^2 的样本估计量．我们来考虑 s_*^2．既然样本 1 和 2 都是从正态回归模型

$$y = \alpha + \beta x + \varepsilon = y_0 + \varepsilon$$

中抽取的独立样本，式中 $\varepsilon \sim N(0, \sigma)$，则样本余方差

$$s_{1*}^2 = \frac{1}{k-2}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2;\quad s_{2*}^2 = \frac{1}{n-2}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2$$

以及

$$s_*^2 = \frac{1}{n+k-4}\left[\sum_{i=1}^{k}(y_i - \hat{y}_i)^2 + \sum_{j=1}^{n}(y_j - \hat{y}_j)^2\right]$$

都是 σ^2 的无偏估计量，于是

$$t_1 = \frac{\hat{y}_1 - \hat{y}_2}{\sqrt{\frac{1}{k-2}\sum(y_i - \hat{y}_i)^2\left[\left(\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum(x_i - \bar{x}_1)^2}\right) + \left(\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum(x_j - \bar{x}_2)^2}\right)\right]}} \tag{17}$$

$$t_2 = \frac{\hat{y}_1 - \hat{y}_2}{\sqrt{\frac{1}{n-2}\sum(y_j - \hat{y}_j)^2\left[\left(\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum(x_i - \bar{x}_1)^2}\right) + \left(\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum(x_j - \bar{x}_2)^2}\right)\right]}} \tag{18}$$

以及

$$t = \frac{\hat{y}_1 - \hat{y}_2}{\sqrt{\frac{\sum(y_i - \hat{y}_i) + \sum(y_j - \hat{y}_j)^2}{n+k-4}\left[\left(\frac{1}{k} + \frac{(x-\bar{x}_1)^2}{\sum(x_i - \bar{x}_1)^2}\right) + \left(\frac{1}{n} + \frac{(x-\bar{x}_2)^2}{\sum(x_j - \bar{x}_2)^2}\right)\right]}} \tag{19}$$

分别是自由度为 $\nu_1 = k - 2$，$\nu_2 = n - 2$ 以及 $\nu = n + k - 4$ 的 t-变量，都可以用来检验假设

$$H_0: y_{10} = y_{20}$$

但是，以 t 值最有效，因为它的余方差估计量最充分地利用了全部样本数据．

这样，若令 $x = \bar{x}_1 = \bar{x}_k$，则(18)式就还原为(6)式，而(19)式所表达的 t 就是(12)式所表达的 z_j 当 $F = 1$，且 $n = k$ 时的特例．

所以我们认为，要检验假设 $H_0: y_{10} = y_{20}$，在 $\sigma_1^2 = \sigma_2^2$ 的情况下，宜用(19)式，它比多个事件检验法中的(6)式更充分地利用了样本资料，且不限于只检验 $(\bar{y}_k - \hat{y}_k)$ 的显著性．在 $\sigma_1^2 \neq \sigma_2^2$ 的情况下，宜用(12)式．当 $F = 1$ 且 $k = n$ 时，(12)式还原为(19)式．

三、讨论和结论

1. 双样本检验中，当方差不相等，且二个样本容量相差较大时，宜采用 Welch 检验法；如果二个样本容量接近相等，且样本方差相差不大时 $(0.333 < F < 3)$，采用 t-检验法误差不大，只在自由度上有不超过20%的误差．

2. 回归分析中，当催化单元倚变量的余方差与对比单元的不相等时，可利用双样本回归分析方法，根据统计量(12)检验二条回归线的差异．这种检验方法比多个事件检验法进了一步．首先，它不需要假定二个回归方程的余方差相等；其次，它可以检验自变量 x

取不同值时的倚变量 $\hat{y}$ 的差异显著性．就人工降水试验来说，即可以检验对比区雨量不同时的催化效果．即使在总体余方差相等的条件下，多个事件检验法也不是最有效的．采用(19)式作为待检验的统计量，由于它充分利用了样本资料，所以比多个事件检验法的统计量(6)式更有效．它是(12)式当 $F=1$，且 k 和 n 相差不大时的特例．

3. 福建古田地区人工降水试验的催化效果．区域回归分析表明，平均相对增雨量是随对比区自然雨量的增加而减小的，小雨时（$x'<2$ 毫米/小时）效果比较显著，相对增雨量达 20—81%，但绝对增雨量并不大，一次试验(三小时时段）约增加 0.44—1.15（毫米）左右．当 $x>1.7$（相当于 $x'>8.35$ 毫米/3 小时)时，效果就不显著了．

由古田试验看来，利用小火箭发射冷云催化剂（碘化银或介乙醛）对古田地区 4—6 月份的降水性云系进行人工催化的方法，一次试验平均在三小时时段内可增加目标区雨量 1 毫米左右．而且，人工降水的效果，看来既不是可加的，也不一定象通常认为的那样是可乘的[10]．因为不少地区的试验都曾发现相对增雨量随自然雨量增加而减小，绝对增雨量则在某一自然雨量强度下达到极大．

古田试验平均每次试验的催化剂用量为碘化银（AgI)220 克或介乙醛（MA)2450 克．室内成冰核率试验表明[11]，在 $-10°$—-15℃范围内，MA 的成冰核率约 10^{10}—5×10^{10}(个/克)；在 $-12°$—-13.5℃时，AgI 的成冰核率为 10^{11}—10^{12}（个/克)．于是，每次试验平均输入云中的有效冰核数约为 2.5×10^{13}—1.2×10^{14} 个(MA）或 2.2×10^{13}—2.2×10^{14} 个（AgI)．如果每个有效核在降水性云系中都能长成 $r=1$ 毫米的雨滴，且都落在目标区，则在目标区 1500 平方公里面积上平均每次可增加雨量约 0.07—0.34 毫米(MA)或 0.07—0.70 毫米（AgI)．如果冰晶在云中有繁生作用，或者大多数人工冰核都先后在不同温度下活化；则人工冰核产生的冰晶再增加几倍甚至 1 个量级是有可能的．这样，只要自然云中水分是充足的，人工催化增加 1 毫米或更多一些雨量是可能的．大雨时效果不显著可能是自然起伏大，检查不出来；或者自然降水过程本身已充分有效，增加冰核无济于事．

参　考　文　献

[1] 叶家东，人工降水的统计研究方法，江西省气象科学研究所印，1977．

[2] 福建省气象局气科所，南京大学气象系，大气科学，3 卷 2 期，131—140，1979．

[3] Simpson, J., Woodley, W. L., Miller, A. R. and Cotton, G. F. *J. Atmos. Sci.*, **10**, 526—544, 1971.

[4] Лесков, Б. Н., Proc. of the WMO/IAMAP Sci. Conf. on weather Modification, Tashkant, 17, 10, 1973, WMO-No. 399, 143—146, 1974.

[5] Gabriel, K. R., The Israeli Rainmaking Experiment 1961—1967 Final Statistical Tables and Evaluation Hebrew University, 1970.

[6] Thom, H. C. S., Final Report of the Advisory Committee on Weather Control, Vol. II., Tech. Rep. No. 1, 1—25, 1957.

[7] Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, vol. 2, p. 146, 1961.

[8] Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering. p. 299, 1965.

[9] 上海师范大学数学系概率统计教研组，回归分析及其试验设计，上海教育出版社，1978．

[10] Scott, E. L., Third Conf. on Probabi. and Statis. in Atmos. Sci., 65—72, 1973.

[11] 福建省气象局、南京大学、西安化工研究所，福建气象科技，2 期，51—54，1977．

TWO SAMPLE REGRESSION ANALYSIS WITH UNEQUAL VARIANCES

Yeh Jia-dong Fan Pei-fen

(Department of Meteorology, Nanjing University)

Abstract

In the regression analysis with control area of rainfall, the usual method of multiple event test is unsuitable in the case of unequal variances. In this paper, based on Welch test, a method of the two-sample regression analysis was proposed, which is suitable in the case of unequal variances. Futher analysis shows that the statistic (19) with equal variances is more powerful than the statistic of multiple event test (6).

The effect of cloud seeding experiment in the area of Gutian, Fujian province, has been evaluated with the two-sample regression analysis. It is shown that on average in the target area the relative enhancement of rainfall by seeding decreases with increase of the rainfall in control area. When the rainfall in control area is small ($x'<2$mm/hr), the seeding effect is significance ($\alpha=0.05$), the relative enhancement of rainfall reaches 20—81%, but absolute enhancement is only 0.44—1.15 (mm/3hrs), and the effect is nonsignificant when $x'>2.8$ (mm/hr).