松江地方の降雨特性について(Ⅱ)※※

線型回帰による雨量の推定

田中礼次郎※

Reijiro TANAKA

A Study on the Characteristic of Rainfall in Matsue District (II)
Estimation of Rainfall Data by Linear Regression

はじめに

一般に相関のある2変量の回帰直線式によって，1つの変量から推定される他方の変量は，相関係数が1の場合以外は回帰直線上にならぶ平均値であって，実際はこの直線のまわりにちらばりをもつはずである．そこでこのちらばりをあらわす項を回帰式にもたせると，より実際の分布によく適合する結果が得られることが予想される．この考え方を水文量を例にとり，松江市の雨量から宍道湖をへだてて約20kmはなれた平田市の雨量推定に適用し，予想どおりの好結果をうることができた．

1. 計算式（2変量正規分布）

x, y が2変量正規分布にしたがっているときの同時確率密度分布は(1)式で与えられる．

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\times\exp\left(-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right\}\right)\cdots\cdots\cdots\cdots(1)$$

ここに ρ は相関係数，μ_x, μ_y および σ_x^2, σ_y^2 はそれぞれ x, y の平均値，分散で，$|\rho|<1$, $\sigma_x>0$, $\sigma_y>0$ である．

また変量 y の $x=x$ なる条件つき確率密度関数 $f(y|x)$ は

$$f(y|x)=\frac{1}{\sqrt{2\pi}\ \sigma_y\sqrt{1-\rho^2}}\times\exp\left(-\frac{1}{2\sigma_y^2(1-\rho^2)}\left\{y-\left(\mu_y+\rho\sigma_y\frac{x-\mu_x}{\sigma_x}\right)\right\}^2\right)\cdots(2)$$

となるからその確率分布は正規分布 $N\left(\mu_y+\rho\frac{\sigma_x}{\sigma_y}(x-\mu_x),\ \sigma_y^2(1-\rho^2)\right)$ である．このときの変量はつぎの構造模型をもつ．

$$y_{x=x}=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)+\varepsilon \qquad \cdots\cdots\cdots(3)$$

ここで ε は $N(0, \sigma_y^2(1-\rho^2))$ にしたがう変量，したがって

$$\alpha=\mu_y-\rho\frac{\sigma_y}{\sigma_x}\mu_x,\ \beta=\rho\frac{\sigma_y}{\sigma_x} \qquad \cdots\cdots\cdots(4)$$

とおけば，(3)式は(3′)式の型となる．

$$y=\alpha+\beta\cdot x+\varepsilon \qquad \cdots\cdots\cdots(3')$$

ここに α, β は回帰係数，ε の分布は $N(0, \sigma^2)$ をなす確率変数，σ^2 は分散

(3′)式の α, β は標本 (x_1, y_1 ; x_2, y_2 ; ……… ; x_n, y_n) より(4′)式で計算できる．

$$\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x},\ \hat{\beta}=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}=r\frac{s_y}{s_x} \qquad \cdots\cdots\cdots(4')$$

ここに $s_x^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$, $s_y^2=\frac{1}{n}\sum_{j=1}^{n}y_j^2-\bar{y}^2$, r は標本相関係数

r は**表1**の相関表のように標本を整理すると(5)式で求められる．

$$r=\frac{\Sigma\Sigma u_iv_jf_{ij}-\frac{\Sigma u_if_i\cdot\Sigma v_jf\cdot j}{N}}{\sqrt{\left\{\Sigma u_i^2f_i\cdot-\frac{(\Sigma u_if_i\cdot)^2}{N}\right\}\left\{\Sigma v_j^2\cdot f\cdot j-\frac{(\Sigma v_jf\cdot j)^2}{N}\right\}}} \qquad \cdots\cdots\cdots(5)$$

※ 農業水利工学研究室
※※昭和45年度農業土木学会大講演会発表（昭和45年5月）

よって y の推定のための回帰直線の方程式は

$$y=\hat{\alpha}+\hat{\beta}x \qquad \cdots\cdots\cdots(6)$$

となる．また y を $\alpha+\beta x$ によって推定するときの分散 σ^2 は

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left\{y_i-(\hat{\alpha}+\hat{\beta}x_i)\right\}^2$$

によって推定されるが，これに(4′)を代入すれば

$$\hat{\sigma}^2=s_y^2(1-r^2) \qquad \cdots\cdots\cdots(7)$$

となる．(3′)式において

$$\varepsilon=\sigma\varepsilon'$$

とおくと，ε' は $N(0,1^2)$ の正規分布にしたがう変量となるからこれに正規乱数を用いて(3′)式を(8)式で表わせば，任意の x に対して回帰直線式(6)式のまわりのちらばりを考慮した y を推定することができる．

$$y=\alpha+\beta x+\sigma\varepsilon' \qquad \cdots\cdots\cdots(8)$$

ここに ε' は正規乱数

また(3′)，(4′)，(7)式より

$$s_y^2=r^2s_y^2+(1-r^2)s_y^2 \qquad \cdots\cdots\cdots(9)$$

が得られる．(9)式は(3′)式から推定した y の分散 s_y^2 は，(6)式から得られる y の分散 $r^2s_y^2$ と，ε 項による $(1-r^2)s_y^2$ の和で与えられることを示している．

表1　r 計算のための相関表（松江日雨量30mm以上）

平田雨量 Y (mm)	$y_i=\log Y$	松江雨量 X (mm)	30.2–35.5	35.5–41.7	41.7–49.0	49.0–57.5	57.5–67.6	67.6–79.4	79.4–93.3	93.3–109.6	109.6–128.8	128.8–177.8	151.4–177.8	$f_{\cdot j}=\sum_i f_{ij}$	$v_jf_{\cdot j}$	$v_j^2f_{\cdot j}$	$\sum_i u_if_{ij}$	$v_j\sum_i u_if_{ij}$
		$x_i=\log X$	1.48–1.55	1.55–1.62	1.62–1.69	1.69–1.76	1.76–1.83	1.83–1.90	1.90–1.97	1.97–2.04	2.04–2.11	2.11–2.25	2.18–2.25					
		v_j \ u_i	-5	-4	-3	-2	-1	0	1	2	3	4	5					
20.0-24.0	1.30-1.38	-5	1	2	1									4	-20	100	-16	80
24.0-28.8	1.38-1.46	-4	2			1		1						4	-16	64	-12	48
28.8-35.0	1.46-1.54	-3	6	4		1			1					12	-36	108	-47	141
35.0-41.7	1.54-1.62	-2	7	3	5			1		1				16	-32	64	-60	120
41.7-50.1	1.62-1.70	-1	3	3	3	3	2	3	1					18	-18	18	-43	43
50.1-60.3	1.70-1.78	0	1	3	2	1	2	1	1			1		12	0	0	-22	0
60.3-72.4	1.78-1.86	1		1	2	5	4	3				1		16	16	16	-20	-20
72.4-87.1	1.86-1.94	2						3	1	1	1			7	14	28	5	10
87.1-104.7	1.94-2.02	3										1		1	3	9	4	12
104.7-125.9	2.02-2.10	4							1		2		1	4	16	64	12	48
125.9-151.4	2.10-2.18	5									1			1	5	25	3	15
$f_{i\cdot}=\sum_j f_{ij}$			20	16	13	11	9	11	5	2	4	3	1	95	-68	496	-196	497
$u_if_{i\cdot}$			-100	-64	-39	-22	-9	0	5	4	12	12	5	-196				
$u_i^2f_{i\cdot}$			500	256	117	44	9	0	5	8	36	48	25	1048				
$\sum_j v_jf_{ij}$			-48	-30	-16	-5	4	2	2	0	15	4	4	-68				
$u_i\sum_j v_jf_{ij}$			240	120	48	10	-4	0	2	0	45	16	20	497				

したがって回帰直線式(6)式から推定した y の標準偏差は rs_y となり，$|r|$ が小さくなるほど小さな値となり，ε 項をもつ(3′)式から推定した y に比べて母集団分布への適合が悪くなる．

2. 適　用　例

2.1　日雨量

1の計算式を2点間の雨量の関係に適用するために，

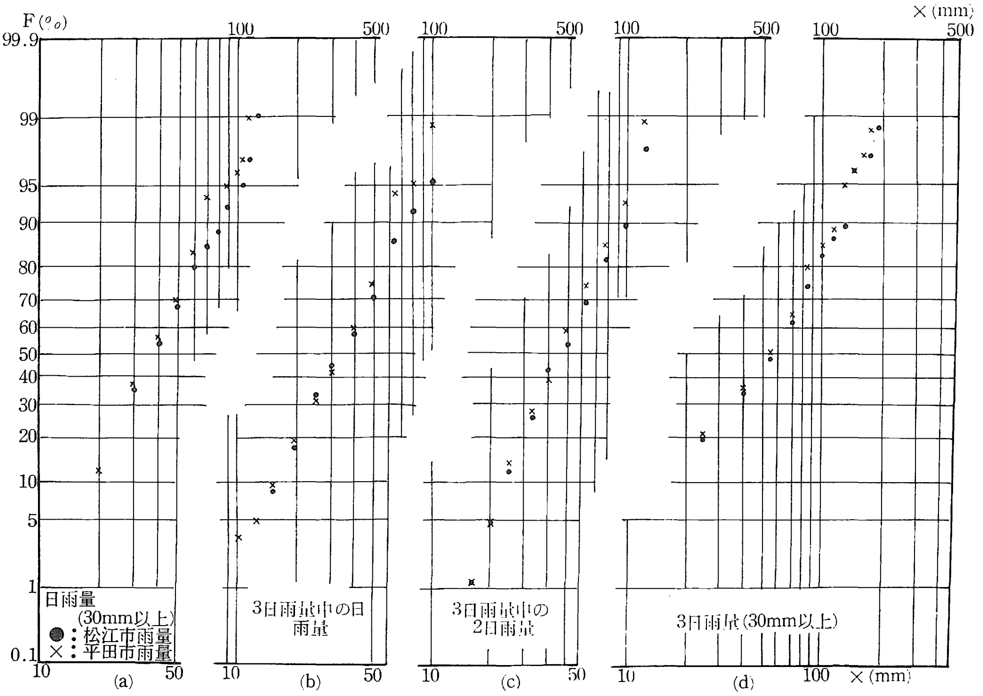

図1　松 江，平 田 雨 量 分 布 図

まず雨量分布を正規化し，正規化された雨量間の標本相関係数 r を求めて母相関係数 ρ の推定値とし，(6′)，(8′)式を求める．すなわち松江市の11年間の日雨量観測記録（昭和28〜38年）から，機械排水の対象を30㎜以上として松江の30㎜以上の日雨量と，これに対応する同日の平田市日雨量95個を取り出して対数確率紙にプロットすれば，**図1**（a）のように両方ともほぼ直線状に分布するので対数正規分布をなすと見なすことができる．そこでおのおのの雨量 X，Y の対数値 x，y をとって分布を正規化し，**表1**のように整理して相関係数 r を(5)式より求めると，$r=0.665$ となる．

つぎに(4′)式より回帰係数 α および β を計算すると

$$\hat{\beta}=r\frac{s_y}{s_x}=0.6336,\quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=0.5925$$

したがって回帰直線は

$$y=0.6336x+0.5925 \qquad \cdots\cdots\cdots(6')$$

となる．y を(6′)式によって推定するときの分散 $\hat{\sigma}^2$ は(7)式から，$\hat{\sigma}^2=0.0168$，$\hat{\sigma}=0.1297$ が得られる．したがって(8)式より

$$y=0.6336x+0.5925+0.1297\varepsilon' \qquad \cdots\cdots\cdots(8')$$

をうる．(8′)式の x に松江日雨量 X の対数値，ε' に正規乱数を順次代入して平田日雨量の対数値 y を求め，X に対応する Y を推定することができる．これを観測雨量と対比した相関図に示すと**図2**のとおりである．

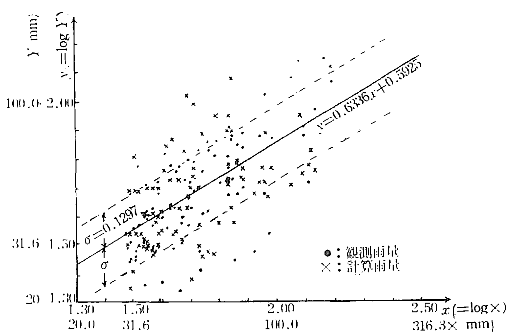

図2　松江，平田日雨量相関図

2.2　3日雨量および3日雨量中の日，2日雨量

農地の豪雨を対象とする排水計画には，2日〜3日雨量を基準雨量とすべき場合が多い．そこで松江の30㎜以上3日連続雨量を対象として，その第1日目，2日目，3日目の雨量をそれぞれ X_1，X_2，X_3 で表し，X_1+X_2 X_2+X_3 の大きい方を3日雨量に含まれる2日雨量としこの2日雨量中の日雨量のうち，大きい方を2日雨量に含まれる日雨量として取り出すと，それぞれ85個のデータが得られる．これらに対応する平田の雨量についても日，2日，3日雨量がそれぞれ85個できるので，これらの雨量を対数確率紙にプロットすると**図1**(*b*)，(*c*)，(*d*)のようにいずれもほぼ直線上に配列するから，対数正規

分布と見なして各雨量間の相関係数を(5)式から求めると，　3日雨量の相関係数　$r_3 = 0.845$

2日雨量　〃　$r_2 = 0.838$

日 雨 量　〃　$r_1 = 0.768$

となり**2.1**で扱った日雨量の相関係数 $r = 0.665$ よりいずれも大きく，より相関の程度が高いことがわかる．これは連続して3日も降雨があるときは両市の距離程度にある2地点においては，単独の日雨量の場合に比較して降雨の地域性が少なくなると解することができる．

2.1 と同じ手順で(6)，(8)式に対応する計算式を求めればつぎのとおりである．

日 雨 量　$y = 0.7349x+0.4091$　………(6_1)

$y = 0.7349x+0.4091+0.1057\varepsilon_1'$ …(8_1)

表2　平田市観測雨量に対する計算雨量の分布比較

階級雨量 (mm)	日雨量				連続3日雨量											
					日雨量				2日雨量				3日雨量			
	(6′)式		(8′)式		(6_1)式		(8_1)式		(6_2)式		(8_2)式		(6_3)式		(8_3)式	
	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}	f_i	f_{mi}
10 — 20					2	7	5	7	1 }6	2 }10	1 }10	2 }10				
20 — 30	—	9	8	9	17	18	13	18	5	8	9	8	—	8	3	8
30 — 40	31	23	26	23	21	13	16	13	15	14	8	14	11	9	7	9
40 — 50	28	19	17	19	17	11	15	11	16	9	11	9	16	9	13	9
50 — 60	18	14	20	14	14	8	13	8	8	14	15	14	9	9	10	9
60 — 70	9	14	10	14	4	11	11	11	14	9	12	9	8	8	9	8
70 — 80	2	10	6	10	3	5	3	5	5	8	7	8	12	10	11	10
80 — 90	6	—	4	—	4 }7	2 }5	3 }4	2 }5	10 }11	4 }9	7 }10	4 }9	9	7	8	7
90 —100	1	1	1	1	3	3	1	3	1	5	3	5	6	9	5	9
100—110	—	1	1	1	—	—	2	—	1 }8	3 }6	3 }5	3 }6	2 }4	2 }5	5 }8	2 }3
110—120	— }7	1 }6	1 }8	1 }6	—	3	1	3	7	3	2	3	2	3	3	3
120—130	—	2	—	2	— }0	2 }7	1	2	1	2	—	2	5 }7	1 }5	1 }2	1 }5
130—140	—	1	1	1	—	2	—	2	1	1	3	1	2	4	1	4
140—150							— }5	— }7	—	2	—	2	1	2	2	2
150—160							—	—	— }2	— }6	2	—	1	1	3	1
160—170							—	—	—	—	— }7	— }6	—	1	1	1
170—180							—	—	—	—	—	—	— }3	1 }6	1	1
180—190							1	—	—	1	1	1	1	—	— }9	— }6
190—200											1	—	—	—	—	—
200—210													—	1	1	1
210—220															1	—
計	95	95	95	95	85	85	85	85	85	85	85	85	85	85	85	85
$\chi^2 = \sum\frac{(f_i-f_{mi})^2}{f_{mi}}$	25.541		6.694		29.377		8.804		17.368		4.657		18.360		12.704	
$\chi^2_n\,(0.05)$	$\chi^2_6\,(0.05)=12.592$				$\chi^2_8\,(0.05)=15.507$				$\chi^2_8\,(0.05)=15.507$				$\chi^2_{10}(0.05)=18.307$			
適合度	×		○		×		○		×		○		×		○	

註　f_i：計算雨量度数，$\chi^2_n(0.05)$：自由度 n の χ^2 分布5%点の値
　　f_{mi}：観測　〃

2日雨量　$y = 0.7986x+0.3346$　………(6_2)

$y = 0.7986x+0.3346+0.1256\varepsilon_2'$ …(8_2)

3日雨量　$y = 0.8295x+0.2929$　………(6_3)

$y = 0.8295x+0.2929+0.1184\varepsilon_3'$ …(8_3)

(8_1)～(8_3)式の x に松江雨量の対数値，ε_i' に正規乱数を代入して対応する平田雨量を求めることができ

る．また同様にして30㎜以上の連続２日雨量とこれに含まれる最大日雨量について関係式を求めるとつぎのようになる．

２日雨量　$y = 0.9378x + 0.0538 + 0.1925\varepsilon_4'$ ……(8_4)

日 雨 量　$y = 0.8052x + 0.2870 + 0.1935\varepsilon_5'$ ……(8_5)

3．適合度の検定

回帰式(６)式よりもちらばりを考慮した(3′)式の方が観測値母集団によく適合する結果が得られることは，理論的には(９)式で説明したとおりであるが，(10)式のχ^2－検定からもたしかめることができる．

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_{mi})^2}{f_{mi}} \qquad \cdots\cdots\cdots(10)$$

ここに，k は階級で $5 \leq k \leq 20$，f_i は計算より求めた雨量度数，f_{mi} は観測雨量度数，f_i または$f_{mi} \geq 5$ になるようにまとめる．

すなわち上述の各式から求めた平田雨量．および観測雨量の各階級ごとの度数 f_i，f_{mi} をしらべ，(10)式の計算値と自由度 $n = k-1$ の χ^2 分布の５％点，χ_n^2（0.05）の値を比較すると，**表２**のとおりでいずれの雨量分布についても回帰式(６)式よりも(８)式の方が非常に適合度が高いことがわかる．

4．む　す　び

従来から一般に使用されている線型回帰式に，その回帰直線式のまわりのちらばりを考慮した ε 項をもたせこれに正規乱数を使用して相関関係にある一方から他方を推定する方法を提示し，これを松江と平田の雨量関係に適用して非常によく母集団分布に適合する結果が得られた．この場合正規乱数の性質からわかるように標本数が多いほど適合度がよくなることはいうまでもない．この方法は，実用的には水工計画において確率水文量などを計算する場合に，水文データーの母集団をこれと相関のある既存資料から推定するのに有効であろう．

Summary

If variables x and y come under the bivariate normal distribution, the relation between x and y is as follows :

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

in which α and β are coefficients of regression, and ε a stochastic variable. The data obtained by applying this equation to the rainfall of Matsue and Hirata are believed to be more suitable for the population than the data estimated by the conventional equation of regression line $y = \alpha + \beta x$.