

回归方程模型的构建及变量的预测

针对数据 MRA_B，我们将探索哪些自变量（X1, X2, X3, X4, X5, X6）在预测因变量 Y 的变化中是有效的，并构建相应的回归方程模型，对回归方程的有效性、自变量对因变量的预测作用进行评估。有时在相关系数显著的情况下，一些变量仍然不能够预测 Y 变量的变化，对此我们将试着进行解释。

【分析】

我们将分别使用 forward 和 backward 方法对回归方程模型进行选择。

【forward 方法】

表 1. 皮尔逊相关系数分析表

	Y	X1	X2	X3	X4	X5	X6
Y	-						
X1	.825**	-					
X2	.426*	.558**	-				
X3	.624**	.597**	.493**	-			
X4	.590**	.669**	.445*	.640**	-		
X5	0.156	0.188	0.147	0.116	.377*	-	
X6	0.155	0.225	0.343	.532**	.574**	0.283	-

因为 X1 与 Y 的相关性最大，所以 X1 是最早被加工模型的变量。

在控制 X1 的情况下，剩下所有的变量对 Y 的 $p\text{-enter}>0.05$ ，又 X1 对 Y 的回归系数显著，所以回归模型最终选择的变量只有 X1。

表 2. 排除的变量^a

模型	输入	Beta	t	显著性	偏相关	共线性统计
						容差
1	X2	-.050 ^b	-0.386	0.702	-0.074	0.688
	X3	.204 ^b	1.571	0.128	0.289	0.644
	X4	.068 ^b	0.470	0.642	0.090	0.552
	X5	.002 ^b	0.014	0.989	0.003	0.965
	X6	-.032 ^b	-0.287	0.777	-0.055	0.950

a. 因变量: Y

b. 模型中的预测变量: (常量), X1

表 3. 输入/除去的变量^a

模型	输入的变 量	除去的变 量	方法
1	X1		向前（准则：要输入的 F 的概率 <= .050）

a. 因变量: Y

表 4. 系数^a

模型		未标准化系数		标准化		显著性
		B	标准错误	Beta	t	
1	(常量)	14.376	6.620		2.172	0.039
	X1	0.755	0.098	0.825	7.737	0.000

a. 因变量: Y

回归方程为: $Y = 14.376 + 0.755 \cdot X1$

ANOVA 分析知 $F(1,28) = 59.861$, $p < 0.001$, 即该回归方程模型有效。自变量对因变量的决定系数 $R^2 = 0.681$, 调整后的 $R^2 = 0.670$, 即 Y 的变异中有 68.1%可由 X1 解释。

【backward 方法】

表 1. 皮尔逊相关系数分析表

	Y	X1	X2	X3	X4	X5	X6
Y	-						
X1	.825**	-					
X2	.426*	.558**	-				
X3	.624**	.597**	.493**	-			
X4	.590**	.669**	.445*	.640**	-		
X5	0.156	0.188	0.147	0.116	.377*	-	
X6	0.155	0.225	0.343	.532**	.574**	0.283	-

模型 1 含所有的预测变量; 因为 X5 变量的预测系数 $B = 0.031$, $p\text{-remove} = 0.796$, 为其中的最大值, 所以模型 2 剔除 X5 变量; 同理, X4、X2、X6、X3 变量对应的 $p\text{-remove} > 0.1$, 所以依次剔除。最后回归模型 6 中的预测变量只剩 X1, X1 对 Y 的回归系数显著, $p\text{-remove} < 0.1$ 。

排除的变量^a

模型	输入	Beta	t	显著性	偏相关	容差
2	X5	.031 ^b	0.261	0.796	0.054	0.814
3	X5	.044 ^c	0.387	0.702	0.079	0.884
	X4	.084 ^c	0.470	0.643	0.095	0.353
4	X5	.043 ^d	0.391	0.699	0.078	0.884
	X4	.091 ^d	0.518	0.609	0.103	0.354
	X2	-.077 ^d	-0.588	0.562	-0.117	0.628
5	X5	.001 ^e	0.007	0.995	0.001	0.965
	X4	-.024 ^e	-0.157	0.876	-0.031	0.462
	X2	-.104 ^e	-0.799	0.432	-0.155	0.648
	X6	-.158 ^e	-1.291	0.208	-0.245	0.704
6	X5	.002 ^f	0.014	0.989	0.003	0.965
	X4	.068 ^f	0.470	0.642	0.090	0.552
	X2	-.050 ^f	-0.386	0.702	-0.074	0.688
	X6	-.032 ^f	-0.287	0.777	-0.055	0.950
	X3	.204 ^f	1.571	0.128	0.289	0.644

a. 因变量: Y

b. 模型中的预测变量: (常量), X6, X1, X2, X3, X4

c. 模型中的预测变量: (常量), X6, X1, X2, X3

d. 模型中的预测变量: (常量), X6, X1, X3

e. 模型中的预测变量: (常量), X1, X3

f. 模型中的预测变量: (常量), X1

系数^a

模型		未标准化系数		标准化系数		显著性
		B	标准错误	Beta	t	
1	(常量)	10.787	11.589		0.931	0.362
	X1	0.613	0.161	0.671	3.809	0.001
	X2	-0.073	0.136	-0.073	-0.538	0.596
	X3	0.320	0.169	0.309	1.901	0.070
	X4	0.082	0.221	0.070	0.369	0.715
	X5	0.038	0.147	0.031	0.261	0.796
	X6	-0.217	0.178	-0.183	-1.218	0.236
2	(常量)	12.798	8.491		1.507	0.145
	X1	0.613	0.158	0.671	3.885	0.001
	X2	-0.072	0.133	-0.073	-0.543	0.592
	X3	0.312	0.162	0.301	1.924	0.066
	X4	0.098	0.208	0.084	0.470	0.643
	X6	-0.211	0.173	-0.178	-1.218	0.235
3	(常量)	14.303	7.740		1.848	0.076
	X1	0.653	0.131	0.715	5.006	0.000

	X2	-0.077	0.131	-0.077	-0.588	0.562
	X3	0.324	0.157	0.312	2.058	0.050
	X6	-0.172	0.149	-0.145	-1.151	0.261
4	(常量)	13.578	7.544		1.800	0.084
	X1	0.623	0.118	0.681	5.271	0.000
	X3	0.312	0.154	0.301	2.026	0.053
	X6	-0.187	0.145	-0.158	-1.291	0.208
5	(常量)	9.871	7.061		1.398	0.174
	X1	0.644	0.118	0.704	5.432	0.000
	X3	0.211	0.134	0.204	1.571	0.128
6	(常量)	14.376	6.620		2.172	0.039
	X1	0.755	0.098	0.825	7.737	0.000

a. 因变量: Y

回归方程为: $Y = 14.376 + 0.755 \cdot X1$

ANOVA 分析知 $F(1,28) = 59.861$, $p < 0.001$, 即该回归方程模型有效。自变量对因变量的决定系数 $R^2 = 0.681$, 调整后的 $R^2 = 0.670$, 即 Y 的变异中有 68.1% 可由 X1 解释。

由表 1 可知 X1, X2, X3, X4 与 Y 的相关系数均显著, 但只有 X1 能预测 Y 的变化, 可能是因为 X2, X3, X4 与 X1 存在较高的相关。由表 1 知, X2, X3, X4 均各自与 X1 为显著的高相关。再做一个 X2, X3, X4 关于 X1 的回归方程可得:

$$X1 = 3.376 + 0.3 \cdot X2 + 0.213 \cdot X3 + 0.546 \cdot X4$$

ANOVA 分析知 $F(3,26) = 10.644$, $p < 0.001$, 即该回归方程模型有效。自变量对因变量的决定系数 $R^2 = 0.551$, 调整后的 $R^2 = 0.499$, 即 X1 的变异中有 55.1% 可由 X2, X3, X4 解释(X2: $t = 1.791$, $p = 0.085$; X3: $t = 1.047$, $p = 0.305$; X4: $t = 2.443$, $p = 0.022$)。即可知可能是因为 X2, X3, X4 与 X1 存在较高的相关, X4 对 X1 有较高的预测性, 所以它们不能预测 Y 变量。