

Многоплановая и многоальтернативная линейная регрессионная модель эксперимента

Н. Бокучава

Проблемная лаборатория физической кибернетики,
Тбилисского Государственного Университета им. Ив. Джавахишвили

Абстракт

Представленную работу можно рассматривать как попытку унификации имеющихся достижений в научном направлении по планированию эксперимента и формирования обобщенной линейной информационно-статистической многоплановой и многоальтернативной регрессионной модели эксперимента.

ВВЕДЕНИЕ

В научных исследованиях основной задачей экспериментатора является отыскание (определение) наилучшего плана эксперимента, позволяющего с наименьшими усилиями и затратами (временными, техническими, финансовыми и др.) получить наибольшую информацию об исследуемом объекте. В процессе поиска такого оптимального плана экспериментатору приходится рассматривать несколько вариантов (моделей) планирования эксперимента с последующим определением некоторого количественного критерия предпочтительности выбора наилучшего из них.

1. ОБОБЩЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ ЭКСПЕРИМЕНТА

Обозначим через X_j ($j = \overline{1, K}$) число выбранных экспериментатором базисных матриц с x_{jr} независимыми элементами ($i = \overline{1, N_j}, r = \overline{1, m_j}, m_j < N_j$) и через Y_j соответствующие им матрицы столбцы наблюдений с y_{ji} элементами наблюдений. Тогда, следуя [1], линейные регрессии для каждой конкретной базисной матрицы X_j могут быть записаны следующим образом

$$Y_j = X_j \beta_j + \varepsilon_j \quad (1)$$

где

$$Y_j = [y_{j1}, y_{j2}, \dots, y_{jN_j}]^T, \quad X_j = [X_{j1}, X_{j2}, \dots, X_{jm_j}], \quad X_{jr} = [x_{j1r}, x_{j2r}, \dots, x_{jN_j r}]^T,$$

$\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm_j}]^T$ - вектор неизвестных параметров, $\varepsilon_j = [\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{jN_j}]^T$ - вектор ненаблюдаемых ошибок. *)

Следует отметить, что обычно вектор ненаблюдаемых ошибок представляет собой комбинацию "модельных" и "измерительных" ошибок. Вводя теперь обозначения,

$$X = \begin{bmatrix} X_1 & & & \\ & X_2 & & O \\ & & O & \\ & & & X_K \end{bmatrix}, \quad Y^{(1)} = \begin{bmatrix} Y_1 & & & \\ & Y_2 & & O \\ & & O & \\ & & & Y_K \end{bmatrix}$$

$$\beta^{(1)} = \begin{bmatrix} \beta_1 & & & \\ & \beta_2 & & O \\ & & O & \\ & & & \beta_K \end{bmatrix}, \quad \varepsilon^{(1)} = \begin{bmatrix} \varepsilon_1 & & & \\ & \varepsilon_2 & & O \\ & & O & \\ & & & \varepsilon_K \end{bmatrix}$$

мы можем записать $j = \overline{1, k}$ систем уравнений регрессии (1) в виде одного объединенного уравнения

*) Заглавными и строчными жирными буквами обозначены матрицы и векторы соответственно

$$Y = X\beta + \varepsilon. \quad (2)$$

Если при этом для каждой конкретной базисной матрицы X_j существует $H^{(l)}$ ($l=1, \overline{M}$) альтернативных гипотез подбора неизвестных параметров β_j , то, следуя (2), многоальтернативную регрессионную модель можем записать в следующей форме

$$Y^{(l)} = X\beta^{(l)} + \varepsilon^{(l)} \quad (3)$$

Исходя из многопланового и многоальтернативного регрессионного уравнения (3) следует, что вектор ошибок $\varepsilon^{(l)}$ зависит от вектора параметра $\beta^{(l)}$, и следовательно, от удачного выбора которого и будет зависеть качество линейной регрессионной модели. В качестве критерия наилучшей оценки параметра $\beta^{(l)}$ воспользуемся оценкой МНК. Тогда, обозначая эту оценку через $\hat{\beta}^{(l)}$, при условии $E(\varepsilon^{(l)}) = 0$, будем иметь

$$\hat{\beta}^{(l)} = \min_{\forall \beta_j^{(l)T} \in R^{m_j}} (\varepsilon^{(l)T} \varepsilon^{(l)}) = \min_{\forall \beta_j^{(l)T} \in R^{m_j}} [Y^{(l)} - X\beta^{(l)}]^T [Y^{(l)} - X\beta^{(l)}],$$

откуда с учетом условия

$$\frac{\partial}{\partial \beta^{(l)}} \left([Y^{(l)} - X\beta^{(l)}]^T [Y^{(l)} - X\beta^{(l)}] \right) = 0 \quad (4)$$

и правила дифференцирования произведения матриц [2]

$$\frac{\partial}{\partial \beta^{(l)}} (A(\beta^{(l)})B(\beta^{(l)})) = \frac{\partial A(\beta^{(l)})}{\partial \beta^{(l)}} B(\beta^{(l)}) + A(\beta^{(l)}) \frac{\partial B(\beta^{(l)})}{\partial \beta^{(l)}}. \quad (5)$$

Для $\hat{\beta}^{(l)}$ получим

$$\hat{\beta}^{(l)} = (X^T X)^{-1} X^T Y^{(l)} \quad (6)$$

3. Определение наилучшей функции регрессии.

После того, как определены оценки $\hat{\beta}^{(l)}$ (следовательно и $\hat{\beta}^{(l)}$), встает вопрос определения наилучшей функции регрессии. Обозначим через

$$\eta^{(l)} \equiv \eta(X, \beta^{(l)}) = X\beta^{(l)} \quad (7)$$

функции регрессии уравнения (3), где

$$\eta^{(l)} = \begin{bmatrix} \eta_1^{(l)} & & & \\ & \eta_2^{(l)} & & 0 \\ & & & 0 \\ 0 & & 0 & \\ & & & \eta_K^{(l)} \end{bmatrix}, \quad \eta^{(l)}_j = [\eta_{j1}^{(l)}, \eta_{j2}^{(l)}, \dots, \eta_{jn_j}^{(l)}] J^T$$

и разложим ее в ряд Тейлора по параметрам $\beta_j^{(l)}$ в окрестности $\hat{\beta}_j^{(l)}$, ограничиваясь первыми производными, тогда

$$\begin{aligned} \eta(X, \beta^{(l)}) &= \eta(X, \hat{\beta}^{(l)}) + \frac{\partial \eta^{(l)}}{\partial \beta_1^{(l)}} \Big|_{\beta_1^{(l)} = \hat{\beta}_1^{(l)}} (\beta_1^{(l)} - \hat{\beta}_1^{(l)}) + \frac{\partial \eta^{(l)}}{\partial \beta_2^{(l)}} \Big|_{\beta_2^{(l)} = \hat{\beta}_2^{(l)}} (\beta_2^{(l)} - \hat{\beta}_2^{(l)}) + \Lambda \\ \Lambda + \frac{\partial \eta^{(l)}}{\partial \beta_K^{(l)}} \Big|_{\beta_K^{(l)} = \hat{\beta}_K^{(l)}} (\beta_K^{(l)} - \hat{\beta}_K^{(l)}) &= \eta(X, \hat{\beta}^{(l)}) + \sum_{j=1}^k X_j \Delta \beta_j^{(l)} \end{aligned} \quad (8)$$

где

$$\Delta \beta_j^{(l)} = \beta_j^{(l)} - \hat{\beta}_j^{(l)}, \quad \frac{\partial \eta^{(l)}}{\partial \beta_j^{(l)}} = X_j.$$

откуда

$$\Delta\eta^{(l)} = \eta(X, \beta^{(l)}) - \eta(X, \hat{\beta}^{(l)}) = \sum_{j=1}^k X_j \Delta\beta_j^{(l)}. \quad (9)$$

Условие $\Delta\eta^{(l)} < \delta$, где δ наперед заданное сколь угодно малое положительное число, будем именовать условием отсева тех функции регрессии, которые не удовлетворяют этому неравенству; постепенным добавлением в разложение (8) производных высших порядков отсева будем продолжать до тех пор, пока не будет выявлена наилучшая (единственная) функция регрессии.

Если после добавления в (8) членов высших производных вопрос однозначного определения наилучшей функции регрессии, с учетом критерия отсева, остается не решенным, то экспериментатору следует:

а) продолжить накопление информации относительно измеряемых (наблюдаемых) переменных, не меняя уже избранного плана эксперимента, путем проведения повторных наблюдений;

б) изменить план эксперимента с помощью увеличения или уменьшения числа элементов базисной матрицы, тем самым изменяя соответственно и размерность параметрического пространства $\beta^{(l)}$, провести дополнительные измерения.

Рассмотрим подробно каждый из этих вариантов.

Вариант а. Вследствии того, что на каждое повторное измерение будут влиять т.н. "скрытые параметры", связанные с "модельными" и "измерительными" ошибками, недоступные контролю со стороны экспериментатора и меняющиеся со временем (например: износ установки, психологический настрой исследователя, изменение внешней среды и др.), для исключения большой разрозненности результатов повторных измерений следует, насколько это возможно:

1) внести определенные коррективы в пространство параметров $\beta^{(l)}$, 2) ограничиться разумным числом повторных измерений.

Обозначим через $\mu_{\eta_j^{(l)}}^{(p)} = [\mu_{\eta_{j_1}^{(l)}}^{(p)}, \mu_{\eta_{j_2}^{(l)}}^{(p)}, \Lambda, \mu_{\eta_{j_{N_j}}^{(l)}}^{(p)}]$ и $f^{(p)}(\eta_{j_1}^{(l)}, \eta_{j_2}^{(l)}, \Lambda, \eta_{j_{N_j}}^{(l)})$ соответственно средние значения и плотности распределения векторов $(\eta_j^{(l)})^T$ при повторных измерениях ($p = \overline{1, c}$ - разумное число повторных измерений). Тогда, исходя из известного факта теории регрессии [2] о нормальном распределении параметров $\beta^{(l)}$ и вектора $\varepsilon^{(l)}$ и принципа максимума информационной энтропии [3] плотности $f^{(p)}(\eta_{j_1}^{(l)}, \eta_{j_2}^{(l)}, \Lambda, \eta_{j_{N_j}}^{(l)})$ могут быть представлены в следующем виде *):

$$f^{(p)}\left((\eta_j^{(l)})^T \right) = \frac{1}{(2\pi)^{N_j/2} |\Lambda_{\eta_j^{(l)}}^{(p)}|^{1/2}} \exp\left(-\frac{1}{2} \left[\eta_j^{(l)} - \mu_{\eta_j^{(l)}}^{(p)} \right]^T \Lambda_{\eta_j^{(l)}}^{(p)-1} \left[\eta_j^{(l)} - \mu_{\eta_j^{(l)}}^{(p)} \right] \right) \quad (10)$$

где

$$\left[\eta_j^{(l)} - \mu_{\eta_j^{(l)}}^{(p)} \right]^T = \left[\left(\eta_{j_1}^{(l)} - \mu_{\eta_{j_1}^{(l)}}^{(p)} \right), \left(\eta_{j_2}^{(l)} - \mu_{\eta_{j_2}^{(l)}}^{(p)} \right), \Lambda, \left(\eta_{j_{N_j}}^{(l)} - \mu_{\eta_{j_{N_j}}^{(l)}}^{(p)} \right) \right],$$

$$\Lambda_{\eta_j^{(l)}}^{(p)} = cov \left\{ \left[\eta_j^{(l)} - \mu_{\eta_j^{(l)}}^{(p)} \right] \left[\eta_j^{(l)} - \mu_{\eta_j^{(l)}}^{(p)} \right]^T \right\}.$$

Исходя из (10), выбирая для варианта а в качестве критерия выбора наилучшей функции регрессии меру различающей информации [3]

*) $\Lambda_{\eta_j^{(l)}}^{(p)-1}$ и $\Lambda_{\eta_j^{(l)}}^{(s)-1}$ соответственно обратные матрицы матриц $\Lambda_{\eta_j^{(l)}}^{(p)}$ и $\Lambda_{\eta_j^{(l)}}^{(s)}$.

$$J(p : s, \eta_j^{(l)r}) = \int f^{(p)}(\eta_j^{(l)r}) \ln \frac{f^{(p)}(\eta_j^{(l)r})}{f^{(s)}(\eta_j^{(l)r})} d\eta_j^{(l)r} \quad (p, s = \overline{1, c}; p \neq s), \quad (11)$$

которая, с учетом известного соотношения из матричной алгебры [4] $b^T A b = \text{Sp}(A b b^T)$, где $A \in M_{n \times n}$, $b \in R^n$, принимает следующий вид:

$$J(p : s, \eta_j^{(l)r}) = \frac{1}{2} \ln \frac{|\Lambda_{\eta_j^{(l)}}^{(s)}|}{|\Lambda_{\eta_j^{(l)}}^{(p)}|} + \frac{1}{2} \text{Sp} \left\{ \Lambda_{\eta_j^{(l)}}^{(p)} \left(\Lambda_{\eta_j^{(l)}}^{(s)-1} - \Lambda_{\eta_j^{(l)}}^{(p)-1} \right) \right\} + \frac{1}{2} \text{Sp} \left\{ \Lambda_{\eta_j^{(l)}}^{(s)-1} \left(\mu_{\eta_j^{(l)}}^{(p)} - \mu_{\eta_j^{(l)}}^{(s)} \right) \left(\mu_{\eta_j^{(l)}}^{(p)} - \mu_{\eta_j^{(l)}}^{(s)} \right)^T \right\} \quad (12)$$

будем отдавать предпочтение той функции регрессии, для которой (12) принимает максимальное значение. Если таковыми окажутся несколько $\eta_j^{(l)}$, то в этом случае следует последовательно увеличивать количество повторных измерений (в разумном объеме) до получения однозначного ответа, в противном случае надо переходить на вариант б.

Вариант б. В этом варианте возможно:

1) не изменяя числа столбцов в исходных базисных матрицах X_j , постепенно увеличивать число строк от N_j до $N_j + n_{j_1}^*$ ($n_{j_1}^* = \overline{1, n_{j_1}}, n_{j_1} \leq N_j$); в результате этого размерности векторов β_j останутся прежними, а размерности векторов $\eta_j^{(l)}$ увеличатся на то же самое число $n_{j_1}^*$, предоставляя тем самым дополнительную информацию о функциях регрессии;

2) не изменяя числа строк в исходных базисных матрицах X_j , постепенно увеличивать число столбцов от m_j до $m_j + n_{j_2}^*$ ($n_{j_2}^* = \overline{1, n_{j_2}}, n_{j_2} \leq m_j$); хотя в этом случае размерности векторов $\eta_j^{(l)}$ остаются прежними (т.е. неизменными), исследователь получает дополнительную информацию о функциях регрессии за счет появления дополнительных слагаемых в каждом из $\eta_j^{(l)}$.

Проводя повторные наблюдения, поиск наилучшей функции регрессии прекращаем на том случае варианта б, для которого информационный критерий $\max J(p : s)$, соотношения (12), дает единственное решение для $\eta_j^{(l)}$.

ЛИТЕРАТУРА

1. Н. Бокучава, Н. Николадзе, Г. Мамасахлисов. *Информационно статистическая модель планирования эксперимента*. Тр. ТГУ. Том 330(19), 1998.
2. С.М. Ермаков, А.А. Жигловский. *Математическая теория оптимального эксперимента*. Москва, "Наука", 1987.
3. Н. Бокучава, Н. Николадзе. *О принципах информационного моделирования*. Тр. ТГУ. Том 330(19), 1998.
4. Ф.Р. Гантмахер. *Теория матриц*. Москва, Наука, 1967.