

Seminar 5 in Applied Statistics

Carlos Buitrago

Table of contents

0.0.1	a) Вероятностный предел	6
0.0.2	b) Состоятельность	7
0.0.3	a) Линейность Q_W	9
0.0.4	b) Частичное исключение W	9
0.0.5	c) Если присутствует квадратичная компонента от W	10

💡 Задача 1 (теорема Гаусса-Маркова)

Предположим, что регрессионная модель удовлетворяет следующим условиям:

ГМ-1 Модель является линейной относительно оцениваемых параметров;

ГМ-2 $\{(Y_i, X_i)\}_{i=1}^n$ — н.о.р. выборка, где X_i — k -мерный вектор регрессоров $(X_{i1}, \dots, X_{ik})$;

ГМ-3 Нет полной мультиколлинеарности, то есть выборочные реализации компонент X_i (которые являются векторами-столбцами матрицы регрессоров) образуют линейно независимый набор векторов в $\mathbb{R}^n$ с вероятностью 1;

ГМ-4 Нулевое условное среднее ошибки:

$$E(\varepsilon_i | X_{i1}, \dots, X_{ik}) = 0;$$

ГМ-5 Гомоскедастичность ошибки:

$$V(\varepsilon_i | X_{i1}, \dots, X_{ik}) = \text{const.}$$

Теорема: если вышеперечисленные предположения выполнены, то МНК-оценки являются лучшими линейными несмещёнными оценками коэффициентов линейной модели — то есть, имеют самую низкую дисперсию в соответствующем классе оценок.

📘 Решение

Рассмотрим

$$Y = X\beta + \varepsilon, \quad V(\varepsilon | X) = \sigma^2 I_n,$$

и

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y.$$

Для произвольного $c \in \mathbb{R}^k$ достаточно доказать

$$V(c^\top \hat{\beta}) \leq V(c^\top \tilde{\beta})$$

для любой другой линейной несмещённой оценки $\tilde{\beta}$.

Любая линейная оценка имеет вид

$$\tilde{\beta} = a^\top Y.$$

Из условной несмещённости:

$$E(\tilde{\beta} | X) = a^\top X \beta = \beta \implies a^\top X = I_k.$$

Для МНК:

$$V(c^\top \hat{\beta} | X) = \sigma^2 c^\top (X^\top X)^{-1} c.$$

Для $\tilde{\beta}$:

$$V(c^\top \tilde{\beta} | X) = \sigma^2 (ac)^\top (ac).$$

Пусть

$$P_X = X(X^\top X)^{-1} X^\top.$$

Поскольку P_X — ортогональный проектор,

$$\|ac\|^2 \geq \|P_X ac\|^2.$$

Следовательно,

$$\begin{aligned} V(c^\top \tilde{\beta} | X) &\geq \sigma^2 (ac)^\top P_X (ac) \\ &= \sigma^2 c^\top a^\top X (X^\top X)^{-1} X^\top ac \\ &= \sigma^2 c^\top (X^\top X)^{-1} c \\ &= V(c^\top \hat{\beta} | X), \end{aligned}$$

где использовано $a^\top X = X^\top a = I_k$.

Итак, МНК имеет минимальную дисперсию среди всех линейных несмещённых оценок. $\square$

🟡 Задача 2

Рассмотрим модель множественной регрессии

$$Y = X\beta + Z\gamma + \varepsilon, \quad \beta \in \mathbb{R}, \quad \gamma \in \mathbb{R}^k,$$

$$V(\varepsilon | X, Z) = \sigma^2 I_n, \quad E(\varepsilon | X, Z) = 0_n.$$

Докажите, что условная дисперсия МНК-оценки β имеет вид

$$V(\hat{\beta} | X, Z) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2 (1 - R_1^2)},$$

где R_1^2 — коэффициент детерминации во вспомогательной регрессии X на Z (предполагая, что константа содержится в Z).

📘 Решение

Пусть

$$Q_Z = I - P_Z$$

— проектор на $L(Z)^\perp$.

По теореме FWL:

$$\hat{\beta} = (X^\top Q_Z X)^{-1} X^\top Q_Z Y.$$

Так как $Q_Z^\top = Q_Z$ и $Q_Z^2 = Q_Z$,

$$\begin{aligned} V(\hat{\beta} | X, Z) &= (X^\top Q_Z X)^{-1} X^\top Q_Z (\sigma^2 I_n) Q_Z X (X^\top Q_Z X)^{-1} \\ &= \frac{\sigma^2}{X^\top Q_Z X}. \end{aligned}$$

Но

$$X^\top Q_Z X = \|Q_Z X\|^2 = SSR_X,$$

где SSR_X — сумма квадратов остатков регрессии X на Z .

Поскольку Z содержит константу,

$$SSR_X = SST_X(1 - R_1^2), \quad SST_X = \sum_{i=1}^n (X_i - \bar{X})^2.$$

Следовательно,

$$V(\hat{\beta} | X, Z) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2 (1 - R_1^2)}.$$

Интерпретация: чем сильнее X объясняется остальными регрессорами Z ($R_1^2 \rightarrow 1$), тем больше дисперсия $\hat{\beta}$. Это эффект мультиколлинеарности.

💡 Задача 3

Докажите, что если модель множественной регрессии удовлетворяет условиям Гаусса-Маркова, то

$$E\hat{\varepsilon}_i = 0.$$

📖 Решение

Остатки:

$$\hat{\varepsilon} = Y - X\hat{\beta} = (I - P_X)Y.$$

Поэтому

$$\begin{aligned} E(\hat{\varepsilon} | X) &= (I - P_X)E(Y | X) \\ &= (I - P_X)X\beta \\ &= 0, \end{aligned}$$

поскольку $P_X X = X$.

По закону повторного математического ожидания

$$E(\hat{\varepsilon}) = E[E(\hat{\varepsilon} | X)] = 0.$$

Следовательно, для каждого i :

$$E\hat{\varepsilon}_i = 0.$$

💡 Задача 4

Пусть истинная зависимость между X_i и Y_i имеет вид

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

и условия Гаусса-Маркова выполнены для выборки размером n . Проанализируйте МНК-оценки параметров и R^2 у каждой из следующих моделей:

a)

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad Y_i^* = aY_i, \quad a \in \mathbb{R};$$

b)

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad Y_i^* = Y_i + b, \quad b \in \mathbb{R};$$

c)

$$Y_i = \beta_0 + \beta_1 X_i^* + \varepsilon_i, \quad X_i^* = cX_i, \quad c \in \mathbb{R};$$

d)

$$Y_i = \beta_0 + \beta_1 X_i^* + \varepsilon_i, \quad X_i^* = X_i + d, \quad d \in \mathbb{R}.$$

i Решение

Пусть $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^\top$.

a) $Y^* = aY$.

$$\hat{\beta}^* = (X^\top X)^{-1} X^\top (aY) = a\hat{\beta}.$$

То есть

$$\hat{\beta}_0^* = a\hat{\beta}_0, \quad \hat{\beta}_1^* = a\hat{\beta}_1.$$

При ненулевом a корреляция меняется максимум знаком, поэтому её квадрат не меняется:

$$R_*^2 = R^2.$$

b) $Y^* = Y + b$.

$$\hat{\beta}_0^* = \hat{\beta}_0 + b, \quad \hat{\beta}_1^* = \hat{\beta}_1.$$

Добавление константы не меняет корреляцию:

$$R_*^2 = R^2.$$

c) $X^* = cX$. Для $c \neq 0$ пространство регрессоров не меняется:

$$L(1, X^*) = L(1, X).$$

Следовательно, $\hat{Y}$ и R^2 те же, а

$$\hat{\beta}_0^* = \hat{\beta}_0, \quad \hat{\beta}_1^* = \frac{\hat{\beta}_1}{c}, \quad R_*^2 = R^2.$$

d) $X^* = X + d$. Поскольку $X = X^* - d$,

$$\hat{\beta}_0 + \hat{\beta}_1 X = (\hat{\beta}_0 - \hat{\beta}_1 d) + \hat{\beta}_1 X^*.$$

Поэтому

$$\hat{\beta}_1^* = \hat{\beta}_1, \quad \hat{\beta}_0^* = \hat{\beta}_0 - \hat{\beta}_1 d, \quad R_*^2 = R^2.$$

Во всех четырёх случаях соответствующее пространство подгонки не меняется (с точностью до масштабирования Y в а)), поэтому R^2 инвариантен.

💡 Задача 5 (МНК в нелинейной модели)

Рассмотрим уравнение

$$Y = (\alpha + \beta X)u,$$

где Y и X — скалярные наблюдаемые, а u — ненаблюдаемая величина. Пусть

$$E(u | X) = 1 \quad \text{и} \quad V(u | X) = \sigma^2.$$

Как бы вы оценили параметры (α, β) методом наименьших квадратов? Как бы вы построили стандартные ошибки?

📌 Решение

Несмотря на мультипликативную запись исходной модели,

$$E(Y | X) = (\alpha + \beta X)E(u | X) = \alpha + \beta X.$$

Поэтому её можно записать как линейную регрессию

$$Y = \alpha + \beta X + \varepsilon,$$

где

$$\varepsilon = (\alpha + \beta X)(u - 1), \quad E(\varepsilon | X) = 0.$$

Значит, α и β можно оценивать обычным МНК.

Но

$$\begin{aligned} V(\varepsilon | X) &= (\alpha + \beta X)^2 V(u - 1 | X) \\ &= \sigma^2 (\alpha + \beta X)^2, \end{aligned}$$

то есть ошибки **гетероскедастичны**.

Следовательно:

- коэффициенты оцениваем обычным МНК;
- обычная гомоскедастическая формула стандартных ошибок неприменима;
- используем **робастные к гетероскедастичности стандартные ошибки**.

OLS coefficients + heteroskedasticity-robust standard errors.

💡 Задача 6

Докажите, что в модели множественной регрессии с константой выполняется соотношение

$$R^2 = \widehat{\text{con}}^2(\hat{Y}, Y).$$

📌 Решение

Пусть

$$Y = \hat{Y} + \hat{\varepsilon}, \quad \hat{Y} = X\hat{\beta}.$$

В модели с константой

$$\bar{\hat{Y}} = \bar{Y}, \quad \sum_{i=1}^n \hat{\varepsilon}_i = 0.$$

Центрируем:

$$\tilde{Y} = Y - \bar{Y}\mathbf{1}, \quad \tilde{\hat{Y}} = \hat{Y} - \bar{Y}\mathbf{1}.$$

Тогда

$$\tilde{Y} = \tilde{\hat{Y}} + \tilde{\varepsilon},$$

причём

$$\tilde{\hat{Y}} \perp \tilde{\varepsilon}.$$

Следовательно,

$$\langle \tilde{\hat{Y}}, \tilde{Y} \rangle = \|\tilde{\hat{Y}}\|^2.$$

Теперь

$$R^2 = \frac{SSE}{SST} = \frac{\|\tilde{\hat{Y}}\|^2}{\|\tilde{Y}\|^2}.$$

С другой стороны,

$$\widehat{\text{сог}}^2(\hat{Y}, Y) = \frac{\langle \tilde{\hat{Y}}, \tilde{Y} \rangle^2}{\|\tilde{\hat{Y}}\|^2 \|\tilde{Y}\|^2}.$$

Используя

$$\langle \tilde{\hat{Y}}, \tilde{Y} \rangle = \|\tilde{\hat{Y}}\|^2,$$

получаем

$$\widehat{\text{сог}}^2(\hat{Y}, Y) = \frac{\|\tilde{\hat{Y}}\|^2}{\|\tilde{Y}\|^2} = R^2.$$

Итак,

$$R^2 = \widehat{\text{сог}}^2(\hat{Y}, Y).$$

💡 Задача 7

Предположим, у вас есть набор реализаций независимых и одинаково распределённых (н.о.р.) копий случайных величин Y и X , и вы хотите оценить регрессионную модель

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

а) Каков вероятностный предел МНК-оценок этой модели (как обычно, ε — ненаблюдаемая ошибка регрессии)?

б) Каковы условия состоятельности МНК-оценки $\hat{\beta}$?

📘 Решение

0.0.1 а) Вероятностный предел

Для парной регрессии с константой

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

Подставим

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Поскольку

$$Y_i - \bar{Y} = \beta(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon}),$$

получаем

$$\hat{\beta} = \beta + \frac{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})(X_i - \bar{X})}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

По ЗБЧ

$$\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})(X_i - \bar{X}) \xrightarrow{p} \text{cov}(\varepsilon, X),$$

а

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow{p} V(X).$$

Следовательно,

$$\text{plim } \hat{\beta} = \beta + \frac{\text{cov}(\varepsilon, X)}{V(X)}.$$

Для константы

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}.$$

Поэтому

$$\text{plim } \hat{\alpha} = E(Y) - (\text{plim } \hat{\beta})E(X).$$

Так как

$$E(Y) = \alpha + \beta E(X) + E(\varepsilon),$$

то

$$\text{plim } \hat{\alpha} = \alpha + E(\varepsilon) - E(X) \frac{\text{cov}(\varepsilon, X)}{V(X)}.$$

0.0.2 б) Состоятельность

Для состоятельности $\hat{\beta}$ необходимо и достаточно, чтобы

$$\text{plim } \hat{\beta} = \beta.$$

Следовательно,

$$\text{cov}(\varepsilon, X) = 0.$$

То есть регрессор X должен быть некоррелирован с ошибкой регрессии.

Если мы также хотим состоятельность $\hat{\alpha}$ для α , дополнительно необходимо

$$E(\varepsilon) = 0.$$

Таким образом,

$$\text{cov}(\varepsilon, X) = 0 \implies \hat{\beta} \xrightarrow{p} \beta.$$

А при

$$E(\varepsilon) = 0, \quad \text{cov}(\varepsilon, X) = 0,$$

имеем

$$(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} (\alpha, \beta).$$

💡 Задача 8 (FWL в регрессии по генеральной совокупности)

Пусть $Y, X_1, \dots, X_p, \varepsilon$ — набор случайных величин, где

$$X = (X_1, \dots, X_p)^T$$

и

$$\text{cov}(X, \varepsilon) = 0_p.$$

Коэффициент β в линейной модели неизвестен:

$$Y = X^T \beta + \varepsilon.$$

Предположим, что вектор X можно разделить на две группы:

$$X = (D^T, W^T)^T,$$

где подвектор D размерности p_1 представляет «целевые» регрессоры, а подвектор W размерности p_2 представляет другие регрессоры, иногда называемые контрольными переменными. Уравнение принимает вид

$$Y = D^T \beta_1 + W^T \beta_2 + \varepsilon.$$

Коэффициент β_1 показывает, как меняется прогноз Y при изменении переменной D и фиксированных значениях W . Это предполагает, что D предсказывает Y даже после исключения (линейного) влияния W .

В генеральной совокупности определим оператор частичного исключения Q_W , который берёт случайную величину V и создаёт $\tilde{V}$ по правилу:

$$Q_W(V) = \tilde{V} = V - W^T \gamma_{VW},$$

где

$$\gamma_{VW} = \arg \min_{b \in \mathbb{R}^{p_2}} E(V - W^T b)^2.$$

Если V — случайный вектор, то применение оператора интерпретируется покомпонентно. Для корректного определения этого оператора необходимо, чтобы V и W имели конечные вторые моменты.

а) Покажите, что Q_W является линейным оператором.

б) Примените Q_W к обеим сторонам уравнения. Будет ли результирующая ошибка регрессии некоррелирована с регрессорами? Выразите коэффициент регрессии β_1 через $\tilde{Y}$ и $\tilde{D}$.

с) Как следует модифицировать Q_W в случае, если в уравнении модели присутствует квадратичная компонента от W ?

i Решение

0.0.3 а) Линейность Q_W

Из условия первого порядка для линейной проекции V на W :

$$E[W(V - W^T \gamma_{VW})] = 0.$$

Следовательно,

$$\gamma_{VW} = [E(WW^T)]^{-1} E(WV).$$

Поэтому для $a \in \mathbb{R}$

$$\gamma_{aV, W} = [E(WW^T)]^{-1} E(WaV) = a\gamma_{VW},$$

и

$$Q_W(aV) = aV - W^T(a\gamma_{VW}) = aQ_W(V).$$

Аналогично,

$$\gamma_{V+Z, W} = [E(WW^T)]^{-1} \{E(WV) + E(WZ)\} = \gamma_{VW} + \gamma_{ZW},$$

откуда

$$Q_W(V + Z) = Q_W(V) + Q_W(Z).$$

Следовательно,

$$Q_W(aV + bZ) = aQ_W(V) + bQ_W(Z)$$

и Q_W — линейный оператор.

0.0.4 б) Частичное исключение W

Исходная модель:

$$Y = D^T \beta_1 + W^T \beta_2 + \varepsilon.$$

Применим Q_W к обеим частям. По линейности:

$$Q_W(Y) = Q_W(D^T \beta_1) + Q_W(W^T \beta_2) + Q_W(\varepsilon).$$

Но

$$Q_W(W^T \beta_2) = 0.$$

Кроме того, из

$$\text{cov}(W, \varepsilon) = 0$$

следует, что линейная проекция ε на W равна нулю, поэтому

$$Q_W(\varepsilon) = \varepsilon.$$

Обозначая

$$\tilde{Y} = Q_W(Y), \quad \tilde{D} = Q_W(D),$$

получаем

$$\tilde{Y} = \tilde{D}^T \beta_1 + \varepsilon.$$

Проверим некоррелированность:

$$\text{cov}(\tilde{D}, \varepsilon) = \text{cov}(D - W^T \gamma_{DW}, \varepsilon).$$

Так как D и W являются компонентами X и

$$\text{cov}(X, \varepsilon) = 0,$$

имеем

$$\text{cov}(\tilde{D}, \varepsilon) = 0.$$

Поэтому коэффициент регрессии в генеральной совокупности равен

$$\beta_1 = [E(\tilde{D}\tilde{D}^T)]^{-1} E(\tilde{D}\tilde{Y}).$$

Это и есть версия **теоремы Фриша–Воу–Ловелла (FWL)** для генеральной совокупности.

0.0.5 с) Если присутствует квадратичная компонента от W

Пусть теперь

$$Y = D^T \beta_1 + W^T \beta_2 + W^T B W + \varepsilon,$$

где B — неслучайная квадратная матрица.

Обычного исключения только линейного влияния W теперь недостаточно: необходимо исключить как линейные, так и квадратичные функции W .

Поэтому определяем

$$Q_{\tilde{W}}(V) = V - W^T \gamma_{VW} - W^T B_V W,$$

где коэффициенты выбираются из задачи

$$(\gamma_{VW}, B_V) = \arg \min_{b, B} E (V - W^T b - W^T B W)^2.$$

Идея остаётся той же: прежде чем оценивать влияние D на Y , из обеих величин необходимо удалить **всё влияние контрольных переменных W , которое присутствует в исходной модели.**