

Seminar 3 in Applied Statistics

Carlos Buitrago

Table of contents

0.0.1	Подсказка	2
0.0.2	Подсказки	3
0.0.3	a)	3
0.0.4	b)	3
0.0.5	c)	4
0.0.6	d)	4
0.0.7	a)	4
0.0.8	b)	4
0.0.9	c)	5
0.0.10	d)	5
0.0.11	a)	5
0.0.12	b)	5
0.0.13	c)	6
0.0.14	d)	6
0.0.15	a)	7
0.0.16	b)	7
0.0.17	c)	7
0.0.18	d)	7
0.0.19	e)	7
0.0.20	a)	8
0.0.21	b)	8
0.0.22	c)	8
0.0.23	d)	9
0.0.24	e)	9
0.0.25	f)	9
1	Доп задача	9
1.0.1	a)	10
1.0.2	b)	10
1.0.3	c)	11
1.0.4	d)	11

Задача 1

Прямоугольный стол имеет неизвестные длины сторон $a > 0$ и $b > 0$. Его истинные площадь и периметр равны

$$S = ab, \quad P = 2(a + b).$$

Для измерения площади помощник независимо измеряет обе стороны стола. Каждый раз прибор допускает независимые гауссовские ошибки:

$$\varepsilon_i, \eta_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2),$$

где значение σ^2 известно.

Поэтому при i -м измерении прибор получает длины

$$X_i = a + \varepsilon_i, \quad Y_i = b + \eta_i,$$

а в качестве результата записывается произведение измеренных сторон:

$$S_i = X_i Y_i = (a + \varepsilon_i)(b + \eta_i).$$

После n независимых повторений наблюдается выборка

$$S_1, \dots, S_n.$$

1. Что можно сказать о периметре P , если $\sigma^2 = 0$?
2. Покажите, что при $\sigma^2 > 0$ периметр P идентифицируем по распределению S_i .
3. Постройте состоятельную оценку периметра на основе выборки

$$S_1, \dots, S_n.$$

4. Предложите способ построения приближённого 95% доверительного интервала для P с помощью бутстрэпа.

0.0.1 Подсказка

Исследуйте математическое ожидание и дисперсию случайной величины

$$S_i = (a + \varepsilon_i)(b + \eta_i).$$

В частности, полезно найти

$$\mathbb{E}[S_i] \quad \text{и} \quad \text{Var}(S_i).$$

💡 Задача 2

Пусть $X_1, \dots, X_n$ — случайная выборка из неизвестного абсолютно непрерывного распределения с математическим ожиданием m , дисперсией $\sigma^2 > 0$ и конечным четвёртым моментом.

Для математического ожидания m рассматривается стандартный асимптотический двусторонний доверительный интервал

$$I_n = \left[\bar{X} - q_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n}}, \bar{X} + q_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n}} \right],$$

где $\hat{\sigma}^2$ — состоятельная оценка дисперсии, а $q_{\alpha/2}$ — соответствующий квантиль стандартного нормального распределения.

1. Определите порядок ошибки покрытия этого доверительного интервала, то есть найдите порядок величины

$$\mathbb{P}\{m \in I_n\} - (1 - \alpha).$$

2. Объясните, почему для **двустороннего** доверительного интервала первый член поправки в разложении Эджворта может сократиться.
3. Что изменится, если исходное распределение не является абсолютно непрерывным и, в частности, может быть дискретным?

0.0.2 Подсказки

- Стандартизируйте выборочное среднее (с учётом оценки дисперсии) и обозначьте соответствующую статистику через S_n .
- Вероятность покрытия двустороннего интервала можно представить в виде

$$\mathbb{P}\{-q_{\alpha/2} \leq S_n \leq q_{\alpha/2}\}.$$

Таким образом, необходимо исследовать функцию распределения S_n одновременно в точках $q_{\alpha/2}$ и $-q_{\alpha/2}$.

- Используйте разложение Эджворта:

$$\mathbb{P}(S_n \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + O(n^{-3/2}),$$

где Φ и ϕ — функция распределения и плотность стандартного нормального распределения.

- Обратите особое внимание на симметрию точек $\pm q_{\alpha/2}$ и чётность/нечётность полиномов p_1, p_2 . Именно здесь возникает сокращение члена порядка $n^{-1/2}$.
- После сокращения первого поправочного члена определите, какой член становится главным в ошибке покрытия.
- Для пункта 3 подумайте, какие условия необходимы для применения обычного разложения Эджворта. В дискретном случае у функции распределения появляются скачки, поэтому такое разложение в приведённой форме может быть неприменимо.
- Если разложение Эджворта использовать нельзя, вспомните универсальную оценку скорости сходимости в центральной предельной теореме — **неравенство Берри–Эссеена**. Сравните получаемый из него порядок ошибки с результатом для абсолютно непрерывного распределения.

💡 Задача 3

Рассмотрим стандартную парную линейную регрессию

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Введём вектор наблюдений и матрицу регрессоров

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}.$$

Предполагается, что столбцы матрицы X линейно независимы.

0.0.3 а)

Вычислите матрицы $X^T X$ и $X^T Y$, а затем найдите $(X^T X)^{-1}$.

0.0.4 б)

Используя результат пункта а), вычислите

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X^T X)^{-1} X^T Y$$

и покажите, что полученные компоненты совпадают с обычными скалярными МНК-оценками

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Подсказка. Удобно использовать тождества

$$\sum_{i=1}^n X_i = n\bar{X}, \quad \sum_{i=1}^n Y_i = n\bar{Y}.$$

0.0.5 c)

Рассмотрим матрицу

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

Докажите, что она является матрицей ортогонального проектора.

Подсказка. Достаточно проверить два свойства:

$$P_X^T = P_X, \quad P_X^2 = P_X.$$

На какое подпространство осуществляется эта проекция?

0.0.6 d)

Рассмотрим

$$Q_X = I - P_X.$$

Докажите, что Q_X также является матрицей ортогонального проектора, и определите подпространство, на которое она проектирует.

Подсказка. Используйте уже доказанное равенство $P_X^2 = P_X$ при вычислении Q_X^2 .

Дополнительно проверьте, что

$$P_X Q_X = 0.$$

Каков геометрический смысл этого равенства?

 Задача 4

Рассмотрим линейную модель

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 d_i + \beta_3 X_i d_i + \varepsilon_i,$$

где $d_i \in \{0, 1\}$ — дамми-переменная.

Цель задачи — понять, как включение d_i и взаимодействия $X_i d_i$ позволяет одновременно оценить две различные линии регрессии.

0.0.7 a)

Запишите модель отдельно для наблюдений с $d_i = 0$ и $d_i = 1$.

Какими будут свободный член и коэффициент наклона в каждой из двух групп?

Подсказка. Просто подставьте сначала $d_i = 0$, а затем $d_i = 1$.

0.0.8 b)

Поскольку порядок наблюдений не влияет на МНК-оценки, расположите данные так, чтобы

$$d_i = 0, \quad i = 1, \dots, n_0,$$

и

$$d_i = 1, \quad i = n_0 + 1, \dots, n.$$

Покажите, что сумма квадратов остатков представляется в виде суммы двух слагаемых, соответствующих двум группам.

Подсказка. Вы должны получить выражение вида

$$\sum_{i=1}^{n_0} (Y_i - \beta_0 - \beta_1 X_i)^2 + \sum_{i=n_0+1}^n (Y_i - (\beta_0 + \beta_2) - (\beta_1 + \beta_3) X_i)^2.$$

0.0.9 c)

Используя разложение из пункта **b)**, докажите, что МНК-оценки $\hat{\beta}_0$ и $\hat{\beta}_1$ в полной модели совпадают с обычными МНК-оценками, полученными только по подвыборке

$$d_i = 0.$$

Подсказка. Подумайте, можно ли минимизировать два слагаемых суммы квадратов независимо друг от друга.

0.0.10 d)

Дайте интерпретацию коэффициентам β_2 и β_3 .

В частности, выразите их оценки через оценки свободного члена и коэффициента наклона, полученные отдельно в группах $d_i = 0$ и $d_i = 1$.

Подсказка. Для группы $d_i = 1$ параметры линии регрессии равны

$$\beta_0 + \beta_2 \quad \text{и} \quad \beta_1 + \beta_3.$$

Задача 5

Пусть (Y_i, X_i) , $i = 1, \dots, n$, — независимые одинаково распределённые копии случайного вектора (Y, X) .

Рассмотрим линейную модель

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

Не предполагается заранее, что регрессор X_i независим или некоррелирован с ошибкой ε_i .

0.0.11 a)

Запишите МНК-оценку коэффициента наклона $\hat{\beta}$ через центрированные наблюдения $X_i - \bar{X}$ и $Y_i - \bar{Y}$. Затем подставьте

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

и представьте $\hat{\beta}$ в виде

$$\hat{\beta} = \beta + \text{дополнительный член.}$$

Подсказка. После центрирования константа α исчезает.

0.0.12 b)

Найдите вероятностный предел $\hat{\beta}$ при $n \rightarrow \infty$.

Подсказка. Разделите числитель и знаменатель дополнительного члена на n и примените закон больших чисел.

В результате должны появиться величины

$$\text{Cov}(X, \varepsilon) \quad \text{и} \quad \text{Var}(X).$$

0.0.13 c)

Сформулируйте необходимое и достаточное условие состоятельности оценки $\hat{\beta}$, то есть условие, при котором

$$\hat{\beta} \xrightarrow{P} \beta.$$

Подсказка. Используйте найденный в пункте b) вероятностный предел.

0.0.14 d)

Объясните статистический смысл полученного условия.

Обязательно ли для состоятельности требовать независимость X и ε ? Достаточно ли более слабого условия?

Подсказка. Сравните условия

$$X \perp\!\!\!\perp \varepsilon$$

и

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

💡 Задача 6 — теорема Фриша–Воу–Ловелла в генеральной совокупности

Пусть

$$Y = X^T \beta + \varepsilon, \quad \text{Cov}(X, \varepsilon) = 0,$$

где вектор регрессоров разбит на две группы:

$$X = \begin{pmatrix} D \\ W \end{pmatrix}.$$

Здесь $D \in \mathbb{R}^{p_1}$ — интересующие нас регрессоры, а $W \in \mathbb{R}^{p_2}$ — контрольные переменные. Таким образом,

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

Для случайной величины V определим оператор частичного исключения влияния W :

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

где

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

Предполагается существование необходимых вторых моментов и невырожденность соответствующих матриц вторых моментов.

0.0.15 а)

Найдите явный вид коэффициента γ_{VW} .

Подсказка. Выпишите условие первого порядка для задачи

$$\min_b \mathbb{E}(V - W^T b)^2.$$

0.0.16 б)

Докажите, что Q_W является линейным оператором, то есть

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

для любых случайных величин V, Z и $a, b \in \mathbb{R}$.

Подсказка. Используйте явную формулу для γ_{VW} из пункта а).

0.0.17 в)

Примените Q_W к обеим частям исходной модели

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

Покажите, что влияние W исчезает и получается регрессия между «очищенными» величинами

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

Подсказка. Сначала определите $Q_W(W^T \beta_2)$ и $Q_W(\varepsilon)$.

0.0.18 д)

Покажите, что ошибка в преобразованной модели остаётся некоррелированной с регрессорами $\tilde{D}$.

После этого выразите β_1 через $\tilde{Y}$ и $\tilde{D}$.

Подсказка. Рассмотрите

$$\text{Cov}(\tilde{D}, \varepsilon)$$

и воспользуйтесь исходным условием $\text{Cov}(X, \varepsilon) = 0$.

0.0.19 е)

Предположим теперь, что контрольные переменные входят в модель не только линейно, но и квадратично:

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

Как следует изменить оператор Q_W , чтобы полностью исключить влияние контрольных переменных?

Подсказка. Расширьте вектор контрольных регрессоров, включив в него необходимые квадратичные произведения компонент W . После этого примените ту же идею линейного проектирования уже к расширенному вектору.

 Задача 7 — оценивание модели с фиксированными эффектами

Рассмотрим панельные данные

$$\{(Y_{it}, X_{it}) : i = 1, \dots, n, t = 1, \dots, T\},$$

где i обозначает индивида, t — момент времени, а $X_{it} \in \mathbb{R}^k$ — вектор регрессоров. Пусть

$$Y_{it} = \alpha + X_{it}^T \beta + \varepsilon_{it},$$

причём ошибка имеет структуру

$$\varepsilon_{it} = u_i + \tilde{\varepsilon}_{it}.$$

Компонента u_i является индивидуальным эффектом, который может быть связан с регрессорами, тогда как $\tilde{\varepsilon}_{it}$ имеет нулевое математическое ожидание.

Рассмотрим два способа оценивания β :

1. **LSDV (least squares dummy variables)**: добавить отдельную дамми-переменную для каждого индивида;
2. **FE (fixed effects)**: вычесть из каждой переменной её среднее по времени для соответствующего индивида.

Докажите, что оба метода дают одну и ту же МНК-оценку $\hat{\beta}$.

0.0.20 а)

Запишите LSDV-модель в матричном виде

$$Y = Z\mu + X\beta + \varepsilon,$$

где Z — матрица индивидуальных дамми-переменных.

Опишите структуру и размерность матрицы Z .

Подсказка. Каждый столбец Z соответствует одному индивиду. Для наблюдений этого индивида он содержит T единиц, а для остальных — нули.

0.0.21 б)

Вычислите

$$Z^T Z$$

и найдите $(Z^T Z)^{-1}$.

Подсказка. Столбцы Z , соответствующие различным индивидам, ортогональны, а каждый столбец содержит ровно T единиц.

0.0.22 в)

Рассмотрите матрицу проекции

$$P_Z = Z(Z^T Z)^{-1} Z^T.$$

Опишите её блочную структуру и выясните, что происходит с произвольным вектором $V \in \mathbb{R}^{nT}$ после умножения на P_Z .

Подсказка. Покажите, что на диагонали P_Z находятся блоки

$$\frac{1}{T} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

размера $T \times T$.

Какое статистическое преобразование выполняет такой блок?

0.0.23 d)

Пусть

$$Q_Z = I - P_Z.$$

Покажите, что

$$Q_Z Y$$

получается вычитанием из каждого Y_{it} среднего значения данного индивида

$$\bar{Y}_{i\cdot} = \frac{1}{T} \sum_{t=1}^T Y_{it}.$$

Аналогично опишите $Q_Z X$.

Подсказка. Для каждого индивида рассмотрите соответствующий блок из T наблюдений отдельно.

0.0.24 e)

Примените теорему Фриша–Вой–Ловелла к LSDV-модели и покажите, что оценка $\hat{\beta}$ совпадает с МНК-оценкой в регрессии

$$Q_Z Y = Q_Z X \beta + v.$$

Используя пункт **d)**, докажите, что эта регрессия в точности совпадает с FE-регрессией

$$Y_{it} - \bar{Y}_{i\cdot} = (X_{it} - \bar{X}_{i\cdot})^T \beta + (\varepsilon_{it} - \bar{\varepsilon}_{i\cdot}).$$

Таким образом, объясните, почему **LSDV** и **within/FE-оценивание** дают одну и ту же оценку $\hat{\beta}$.

0.0.25 f)

Объясните, почему преобразование из пункта **e)** устраняет ненаблюдаемый индивидуальный эффект u_i .

Подсказка. Вычислите

$$u_i - \frac{1}{T} \sum_{t=1}^T u_i.$$

Почему это особенно полезно, если u_i коррелирует с регрессорами X_{it} ?

1 Доп задача

Задача

Пусть $X_1, \dots, X_n$ — независимые одинаково распределённые случайные векторы из параметрического семейства с плотностью

$$p(x; a, b),$$

где $a \in \mathbb{R}$ — **параметр интереса**, а $b \in \mathbb{R}^s$ — **мешающий параметр**.

Истинное значение параметров обозначим через (a_0, b_0) . Предполагается, что при проверке нулевой гипотезы значение a_0 известно, тогда как b_0 неизвестно.

Пусть

$$\ell_i(a, b) = \log p(X_i; a, b)$$

— вклад i -го наблюдения в логарифмическую функцию правдоподобия. Рассмотрим скор-функцию по параметру интереса

$$S_a(b) = \sum_{i=1}^n \frac{\partial \ell_i(a_0, b)}{\partial a}.$$

Предполагается выполнение стандартных условий регулярности для ММП и информационной матрицы Фишера.

1.0.1 а)

Найдите предельное распределение статистики

$$\frac{1}{\sqrt{n}} S_a(b_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ell_i(a_0, b_0)}{\partial a}.$$

Выразите дисперсию предельного распределения через соответствующий блок информационной матрицы Фишера.

Подсказка. При выполнении стандартных условий регулярности скор-функция в истинной точке имеет нулевое математическое ожидание:

$$\mathbb{E} \left[\frac{\partial \ell_i(a_0, b_0)}{\partial a} \right] = 0.$$

После этого примените многомерную/одномерную ЦПТ и вспомните связь между дисперсией score и информацией Фишера.

1.0.2 б)

Пусть $\hat{b}$ — ММП для мешающего параметра b при фиксированном $a = a_0$, причём

$$\sqrt{n}(\hat{b} - b_0)$$

имеет асимптотически нормальное распределение.

Исследуйте предельное распределение

$$\frac{1}{\sqrt{n}} S_a(\hat{b}).$$

Покажите, что оно в общем случае **не обязано совпадать** с предельным распределением

$$\frac{1}{\sqrt{n}} S_a(b_0).$$

Подсказка. Разложите скор-функцию по b в окрестности b_0 :

$$S_a(\hat{b}) = S_a(b_0) + \frac{\partial S_a(b_0)}{\partial b^T} (\hat{b} - b_0) + \text{остаточный член}.$$

Затем используйте асимптотическое представление ММП $\hat{b}$ через score по b .

Обратите внимание: поскольку

$$\hat{b} - b_0 = O_p(n^{-1/2}),$$

$a \partial S_a / \partial b$ представляет собой сумму по n наблюдениям, дополнительный член после нормировки на $\sqrt{n}$ может иметь порядок $O_p(1)$ и потому не исчезает.

Подумайте, какую роль здесь играет блок I_{ab} информационной матрицы Фишера.

1.0.3 c)

Чтобы устранить влияние оценивания мешающего параметра, рассмотрим **ортогонализированный score**

$$B_n(b) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\partial \ell_i}{\partial a} \Big|_{(a_0, b)} - I_{ab} I_{bb}^{-1} \frac{\partial \ell_i}{\partial b} \Big|_{(a_0, b)} \right],$$

где I_{ab} и I_{bb} — соответствующие блоки информационной матрицы Фишера.
Найдите предельное распределение

$$B_n(b_0)$$

и вычислите его асимптотическую дисперсию.

Подсказка. Обозначьте индивидуальные score через

$$s_{a,i} = \frac{\partial \ell_i}{\partial a}, \quad s_{b,i} = \frac{\partial \ell_i}{\partial b}.$$

Тогда нужно найти дисперсию величины

$$s_{a,i} - I_{ab} I_{bb}^{-1} s_{b,i}.$$

Используйте соотношения

$$\text{Var}(s_{a,i}) = I_{aa}, \quad \text{Var}(s_{b,i}) = I_{bb}, \quad \text{Cov}(s_{a,i}, s_{b,i}) = I_{ab}.$$

Полученная величина связана с **эффективной информацией** для параметра a при наличии мешающего параметра b .

1.0.4 d)

Пусть в ортогонализированной статистике вместо неизвестного b_0 используется состоятельная ММП $\hat{b}$.
Исследуйте

$$B_n(\hat{b})$$

и сравните его предельное распределение с распределением $B_n(b_0)$.

Подсказка. Снова выполните разложение по Тейлору в окрестности b_0 :

$$B_n(\hat{b}) = B_n(b_0) + \frac{\partial B_n(b)}{\partial b^T} \Big|_{b=b_0} (\hat{b} - b_0) + \dots$$

Ключевой вопрос: имеет ли теперь линейный поправочный член порядок $O_p(1)$, как в пункте **b)**, или он исчезает асимптотически?

Используйте специальный выбор коэффициента

$$I_{ab} I_{bb}^{-1}$$

и проверьте, что он делает ортогонализированный score локально нечувствительным к малым изменениям мешающего параметра b .

Объясните, почему именно это свойство позволяет заменить неизвестное b_0 его оценкой $\hat{b}$, не изменяя предельное распределение статистики.