

ЧАСТЬ II Решение

Задача 1

Оценивание производственной функции Кобба-Дугласа $Q = AK^\alpha L^\beta$ по $n = 30$ наблюдениям дало следующие результаты:

$$\ln Q = 1.37 + \underset{(0.231)}{0.656} \cdot \ln K + \underset{(0.182)}{0.473} \cdot \ln L, \quad R^2 = 0.98, \quad \widehat{Cov}(\hat{\alpha}, \hat{\beta}) = 0.043$$

(в скобках даны стандартные ошибки). Проверьте гипотезу о постоянстве отдачи от масштаба (*constant return to scale*).

Решение. Задача равносильна тестированию гипотезы $H_0: \alpha + \beta = 1$ против альтернативы

$H_a: \alpha + \beta \neq 1$ для модели $\ln Q = c + \alpha \ln K + \beta \ln L + \varepsilon$. Обозначим $a = \hat{\alpha}$, $b = \hat{\beta}$

МНК-оценки коэффициентов α , β , соответственно. Применим t -тест. Тестовая статистика:

$$t = \frac{a+b-1}{s_{a+b}}, \quad \text{где } s_{a+b}^2 = s_a^2 + s_b^2 + 2 \cdot \widehat{Cov}(a, b).$$

Имеем:

$$s_{a+b} = \sqrt{0.231^2 + 0.182^2 + 2 \cdot 0.043} = 0.415, \quad t = \frac{0.656 + 0.473 - 1}{0.415} = 0.311.$$

При нулевой гипотезе статистика t имеет распределение Стьюдента с $30 - 3 = 27$ степенями свободы. Поскольку $t_{0.25}(27) = 0.684$, то P -значение теста больше, чем 0.5.

Таким образом, на любом разумном уровне значимости нулевая гипотеза не отвергается.

Задача 2

Для оценки потребления кофе в некотором регионе рассматривается простая линейная модель

$$Q_t = \alpha + \beta P_t + \varepsilon_t,$$

где Q_t – количество кофе, проданное в регионе в год t , P_t – мировая цена кофе в год t .

По данным за последние 12 лет известны следующие выборочные статистики:

- коэффициент корреляции между объемом продаж и ценами, равный -0.856 ;
- средняя цена кофе, равная 0.610;
- средний объем продаж, равный 617.2;
- стандартное отклонение цены, равное 0.377;
- стандартное отклонение объема продаж, равное 40.76.

(а) Вычислите МНК-оценки коэффициентов α , β .

(б) Вычислите коэффициент детерминации R^2 .

(в) В текущем году из-за неблагоприятных погодных условий ожидается, что мировая цена кофе будет 1.2. Оцените средний объем продаж кофе в регионе и постройте для него двусторонний 95%-ный доверительный интервал.

Напоминание. Для двух наборов $\{x_i, i = 1, \dots, n\}$, $\{y_i, i = 1, \dots, n\}$ выборочный коэффициент

корреляции r_{xy} вычисляется по формуле
$$r_{xy} = \frac{1/(n-1) \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y}.$$

Решение

(а) Обозначим

$$S_{PP} = \sum_{t=1}^{12} (P_t - \bar{P})^2, \quad S_{QQ} = \sum_{t=1}^{12} (Q_t - \bar{Q})^2, \quad S_{PQ} = \sum_{t=1}^{12} (P_t - \bar{P})(Q_t - \bar{Q}).$$

Заметим, что $S_{PP} = (n-1)s_P^2 = 1.56$, $S_{QQ} = (n-1)s_Q^2 = 18275.15$

Тогда

$$b = \frac{S_{PQ}}{S_{PP}} = \frac{S_{PQ}}{\sqrt{S_{PP}} \cdot \sqrt{S_{QQ}}} \cdot \frac{\sqrt{S_{QQ}}}{\sqrt{S_{PP}}} = \frac{S_{PQ}/(n-1)}{\sqrt{S_{PP}/(n-1)} \cdot \sqrt{S_{QQ}/(n-1)}} \cdot \frac{\sqrt{S_{QQ}/(n-1)}}{\sqrt{S_{PP}/(n-1)}} = r_{PQ} \cdot \frac{s_Q}{s_P}.$$

Получаем:

$$b = -0.856 \cdot \frac{40.76}{0.377} = -92.55, \quad a = \bar{Q} - b \cdot \bar{P} = 673.65.$$

Окончательно, получаем уравнение

$$Q_t = 673.56 - 92.55 \cdot P_t + e_t,$$

где e_t – остаток.

(б) Найдем сумму квадратов остатков $\sum_{t=1}^n e_t^2$, $n = 12$. Из теории известно, что

$$\sum_{t=1}^n e_t = 0, \quad \sum_{t=1}^n P_t \cdot e_t = 0. \quad (1)$$

Имеем: $Q_t = a + bP_t + e_t$, $t = 1, \dots, n$. Вычитая почленно равенство $\bar{Q} = a + b\bar{P}$, получаем:

$Q_t - \bar{Q} = b(P_t - \bar{P}) + e_t$. Возводя в квадрат, получаем:

$(Q_t - \bar{Q})^2 = b^2(P_t - \bar{P})^2 + e_t^2 + 2b(P_t - \bar{P})e_t$, $t = 1, \dots, n$. Суммируя по t и учитывая (1), получаем:

$$RSS = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Q_t - \bar{Q})^2 - b^2 \sum_{t=1}^n (P_t - \bar{P})^2 = S_{QQ} - b^2 S_{PP},$$

так как $\sum_{t=1}^n 2b(P_t - \bar{P})e_t = 2b \sum_{t=1}^n P_t e_t - 2b\bar{P} \sum_{t=1}^n e_t = 0$. Таким образом,

$$RSS = \sum_{t=1}^n e_t^2 = 18275.15 - (92.55)^2 \cdot 1.56 = 4884.29. \quad (2)$$

Наконец,

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{RSS}{S_{QQ}} = 0.733.$$

Замечание. Те, кто хорошо знают теорию, могут воспользоваться равенством $r_{PQ}^2 = R^2$. Однако, в дальнейшем формула (2) нам все равно понадобится.

(в) Обозначим $P_0 = 1.2$, $Q_0 = \alpha + \beta P_0 + \varepsilon_0$, $m_0 = E(Q_0) = \alpha + \beta P_0$. Тогда точечная оценка параметра m_0 есть $\hat{m}_0 = a + b P_0 = 673.65 - 92.55 \cdot 1.2 = 562.60$. Как известно, 95%-ный доверительный интервал для m_0 имеет вид:

$$m_0 = \hat{m}_0 \pm t_{0.025}(n-2) \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(P_0 - \bar{P})^2}{\sum_{t=1}^n (P_t - \bar{P})^2}},$$

где $n = 12$, $s = \sqrt{\frac{RSS}{n-2}}$. Из таблиц получаем $t_{0.025}(10) = 2.228$. С учетом (2) окончательно получаем
 $m_0 = 562.60 \pm 27.24$, $m_0 \in (535.36, 589.84)$

Задача 4

Ответы на вопросы должны быть краткими и исчерпывающими

1. Перечислите основные статистические свойства МНК–оценок в классической линейной модели множественной регрессии.

Ответ. Классическая линейная модель множественной регрессии:

$y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 I$. МНК–оценка: $\hat{\beta} = b = (X'X)^{-1}X'y$. Свойства:

- $E(b) = \beta$, т.е. b – несмещенная оценка, $V(b) = \sigma^2(X'X)^{-1}$;
- b – состоятельная оценка;
- в классе всех линейных (по y) и несмещенных оценок оценка b эффективна, т.е. если $\tilde{\beta} = Cy$, $E(\tilde{\beta}) = \beta$, то $V(\tilde{\beta}) \leq V(b)$ (теорема Гаусса–Маркова).

2. Известно, что в стандартной линейной модели $V(\varepsilon) = \Omega(\theta)$, т.е. матрица Ω известна с точностью до параметра. Какие свойства имеет МНК–оценка $\hat{\beta}$ в этом случае? Как бы вы оценили коэффициенты β , если бы параметр θ вам был известен? если бы вы не знали точное значение θ , но могли бы его оценить?

Ответ. В этом случае МНК–оценка $\hat{\beta}$ является несмещенной, состоятельной, но теряет эффективность. Если параметр θ известен, то эффективной оценкой является оценка обобщенного метода наименьших квадратов:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}(\theta)X)^{-1}X'\Omega^{-1}y. \quad (3)$$

Если параметр θ неизвестен, но есть его «хорошая» оценка $\hat{\theta}$, то надо реализовать формулу (3) с заменой θ на $\hat{\theta}$.

3. У Оксаны Блэк есть данные по 1000 индивидуумов, половина из которых живет в Москве, а половина – в Новосибирске. Для каждого индивидуума известны уровень образования (число лет, потраченных на обучение) и зарплата. Какую регрессию следует осуществить Оксане, чтобы ответить на вопрос: есть ли существенная разница между уровнем зарплат в Москве и в Новосибирске.

Ответ. Самый простой способ – ввести фиктивную (dummy) переменную N_i : $N_i = 1$, если индивидуум i живет в Новосибирске, и $N_i = 0$, если индивидуум i живет в Москве, оценить регрессию $w_i = \alpha + \beta_{EDU} \cdot EDU_i + \beta_N \cdot N_i + \varepsilon_i$ и протестировать гипотезу $H_0: \beta_N = 0$ против альтернативы $H_a: \beta_N \neq 0$. Если нулевая гипотеза будет отвергнута, то это может служить эмпирическим подтверждением того, что уровни зарплат в Москве и Новосибирске отличаются, в предположении, что отдача от образования в обоих городах одинаковая.