

Семинар 4. Множественная регрессия: проверка гипотез о линейных ограничениях

Задача 1.

При исследовании факторов, определяющих экономический рост, по 70 странам было получено уравнение регрессии (в скобках указаны стандартные отклонения):

$$\widehat{G} = 1.5 - 0.5P + 0.2S + 12I - 0.4D + 5In, R^2 = 0.6.$$

(0.1) (0.042) (3) (0.5) (3.1)

где G – темп экономического роста, P – среднедушевой ВВП, S – бюджетный дефицит, I – объем инвестиций, D – внешний долг, In – уровень инфляции.

- 1) Проверить адекватность модели.
- 2) Согласно этой модели, при уровне значимости 5% можно утверждать, что темпы экономического роста зависят от
 - 2.1) среднедушевого ВВП
 - 2.2) бюджетного дефицита
 - 2.3) объема инвестиций
 - 2.4) внешнего долга
 - 2.5) уровня инфляции

Решение:

- 1) Проверка адекватности модели – это проверка гипотезы $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$ ($k = 6$ – число оцениваемых коэффициентов) против альтернативной гипотезы $H_1: \exists \beta_i \neq 0$ с помощью статистики

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}. \text{ В нашем случае } F = \frac{0.6}{1-0.6} \cdot \frac{70-6}{6-1} = 19.2. \text{ Т.к. 5\%-ая точка распределения Фишера}$$

$F_{5\%}(k-1, n-k) = F_{5\%}(5, 64) \approx 2.37$, значит, нулевая гипотеза $H_0: \beta_2 = \beta_3 = \dots = \beta_6 = 0$ отвергается на 5% уровне значимости.

- 2) В этом пункте нам необходимо проверить значимость каждого из факторов, входящих в модель, по отдельности. Для всех коэффициентов критическая точка t -распределения будет одна и та же: $t_{2,5\%}(n-k) = t_{2,5\%}(64) \approx 2$.

2.1) $P: t_2 = \frac{-0.5}{0.1} = -5 \Rightarrow$ коэффициент значим на 5% уровне значимости, есть влияние P на G

2.2) $S: t_3 = \frac{0.2}{0.042} = 4.76 \Rightarrow$ коэффициент значим на 5% уровне значимости, есть влияние S на G

2.3) $I: t_4 = \frac{12}{3} = 4 \Rightarrow$ коэффициент значим на 5% уровне значимости, есть влияние I на G

2.4) $D: t_5 = \frac{-0.4}{0.5} = -0.8 \Rightarrow$ коэффициент незначим на 5% уровне значимости, нет влияния D на G

2.5) $In: t_6 = \frac{5}{3.1} = 1.61 \Rightarrow$ коэффициент незначим на 5% уровне значимости, нет влияния In на G

Задача 2.

Оценивался спрос на цейлонский чай в США с помощью следующей регрессии:

$$\ln Q = \beta_1 + \beta_2 \ln P_C + \beta_3 \ln P_I + \beta_4 \ln P_B + \beta_5 \ln Y + u,$$

где Q – спрос на цейлонский чай, P_I – цена индийского чая, P_C – цена цейлонского чая, P_B – цена бразильского кофе, Y – располагаемый доход.

По 22 наблюдениям были получены следующие результаты (в скобках указаны стандартные отклонения):

$$\widehat{\ln Q} = 2.837 - 1.481 \ln P_C + 1.181 \ln P_I + 0.186 \ln P_B + 0.257 \ln Y, RSS = 0.4277$$

(2) (0.987) (0.69) (0.37) (0.37)

Эта модель была также оценена при ограничениях $\beta_2 = -1$, $\beta_3 = 0$ и получено:

$$\widehat{\ln Q + \ln P_C} = -0.738 + 0.199 \ln P_B + 0.261 \ln Y, RSS = 0.6788$$

(0.82) (0.155) (0.165)

Проверить гипотезу $\beta_2 = -1$, $\beta_3 = 0$ и дать экономическую интерпретацию полученным результатам.

Решение:

В соответствии с общей процедурой проверки нескольких линейных ограничений, необходимо рассчитать тестовую статистику $F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$. В данном случае

$$F = \frac{(0.6788 - 0.4277)/2}{0.4277/(22-5)} = 4.99. \text{ Т.к. 5\%-ая критическая точка соответствующего F-распределения есть}$$

$$F_{5\%}(2,17) = 3.59, \text{ мы отвергаем нулевую гипотезу } H_0: \begin{cases} \beta_2 = -1 \\ \beta_3 = 0 \end{cases} \text{ на 5\% уровне значимости. Таким}$$

образом, цена индийского чая влияет на спрос на цейлонский чай и/или коэффициент при цене цейлонского чая не равен -1.

Задача 3.

По данным для 27 фирм оценили производственную функцию с помощью трех моделей:

$$\ln Y = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + u \quad (1)$$

$$\ln Y = \alpha_1 + \alpha_2 \ln(LK) + \varepsilon \quad (2)$$

$$\ln Y/K = \gamma_1 + \gamma_2 \ln L/K + \xi \quad (3)$$

Суммы квадратов остатков для оцененных функций регрессий (1), (2), (3) оказались соответственно равны:

$$RSS_1 = 8.51, RSS_2 = 8.94, RSS_3 = 9.01$$

Объяснить, почему вторая и третья модели являются ограниченными версиями первой, выписать соответствующие ограничения на коэффициенты регрессии (1) и проверить их выполнение.

Решение:

Нетрудно заметить, что предположение $\beta_2 = \beta_3$ превращает регрессию (1) в регрессию (2) в силу свойства логарифма $\ln a + \ln b = \ln ab$. Аналогично, регрессия (3) получается из регрессии (1) при предположении $\beta_2 + \beta_3 = 1$ в силу свойства логарифма $\ln a - \ln b = \ln \frac{a}{b}$.

Соответственно, нам нужно проверить две гипотезы о линейных ограничениях на параметры:

$$H_0: \beta_2 = \beta_3 \text{ и } H_0: \beta_2 + \beta_3 = 1. \text{ В первом случае тестовая статистика равна } F = \frac{(8.94 - 8.51)/1}{8.51/(27-3)} = 1.21, \text{ во}$$

$$\text{втором } F = \frac{(9.01 - 8.51)/1}{8.51/(27-3)} = 1.41.$$

В обоих случаях критическая точка F-распределения будет одной и той же в силу одинакового количества ограничений (одно ограничение в обоих случаях): $F_{5\%}(1,24) = 4.26$. Т.к. обе тестовые статистики меньше критической точки, в обоих случаях нулевая гипотеза не отвергается на 5% уровне значимости.

Задача 4.

Изучая зависимость длительности обучения S 540 индивидов от способностей IQ , характеризуемых результатами трех тестов (см. пункт в), длительности обучения матери индивида SM , длительности обучения отца индивида SF , исследователь получил следующие функции регрессии (в скобках указаны стандартные отклонения):

$$\hat{S} = \underset{(0.52)}{5} + \underset{(0.0099)}{0.115} IQ + \underset{(0.039)}{0.12} SM + \underset{(0.029)}{0.1} SF, \text{ } RSS = 2100.646, R^2 = 0.336 \quad (1)$$

$$\hat{S} = \underset{(0.48)}{6.5} + \underset{(0.009)}{0.14} IQ, \text{ } RSS = 2267.587 \quad (2)$$

- Исходя из полученных результатов, можно ли считать, что длительность обучения индивида зависит только от его способностей?

2. Была оценена также регрессия

$$\hat{S} = 5.22 + 0.115 IQ + 0.109(SM + SF), \quad RSS = 2100.962. \quad (3)$$

(0.502) (0.0099) (0.016)

Исходя из полученного результата, можно ли считать, что родители в равной степени влияют на длительность обучения индивида?

3. Значения переменной, характеризующей способности индивида, рассчитывались следующим образом:

$$\widehat{IQ} = 0.5IQ_1 + 0.25IQ_2 + 0.25IQ_3, \text{ где}$$

IQ_1 – результаты теста по арифметике,

IQ_2 – результаты теста по правописанию,

IQ_3 – результаты теста по пониманию прочитанного материала.

Исследователь оценил также регрессию

$$\hat{S} = 4.75 + 0.088 IQ_1 + 0.035 IQ_2 - 0.0013 IQ_3 + 0.12 SM + 0.1 SF, \quad R^2 = 0.352 \quad (4)$$

(0.54) (0.012) (0.015) (0.013) (0.039) (0.029)

Исходя из полученных результатов, можно ли считать, что веса в переменной IQ выбраны правильно?

Решение:

1. Предположение, что длительность обучения индивида зависит только от IQ, эквивалентно нулевой гипотезе о линейных ограничениях вида $H_0: \beta_3 = \beta_4 = 0$ против альтернативы в виде $H_1: \beta_3^2 + \beta_4^2 > 0$. В соответствии с общей процедурой проверки нескольких линейных ограничений, необходимо рассчитать тестовую статистику $F_o = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$. В данном случае

$F_o = \frac{(2267.587 - 2100.646)/2}{2100.646/(540-4)} = 21.298$. Т.к. 5%-ая критическая точка соответствующего F-распределения есть $F_{5\%}(2, 536) \approx F_{5\%}(2, 500) = 3.01$, мы отвергаем нулевую гипотезу на 5% уровне значимости.

2. Регрессия (3) была получена из регрессии (1) исходя из нулевой гипотезы $H_0: \beta_3 = \beta_4$. В соответствии с общей процедурой проверки нескольких линейных ограничений, необходимо рассчитать тестовую статистику $F_o = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$. В данном случае

$F_o = \frac{(2100.962 - 2100.646)/1}{2100.646/(540-4)} = 0.08$. Т.к. 5%-ая критическая точка соответствующего F-распределения есть $F_{5\%}(1, 536) \approx F_{5\%}(1, 500) = 3.86$, мы не отвергаем нулевую гипотезу на 5% уровне значимости о равном вкладе образований матери и отца.

3. Нам нужно проверить, будут ли реальные коэффициенты зависимости от отдельных тестов равны весам этих тестов в обобщенном показателе. Для того чтобы регрессия (1) представляла собой ограниченную версию регрессии (4) $S = \beta_1 + \beta_2 IQ_1 + \beta_3 IQ_2 + \beta_4 IQ_3 + \beta_5 SM + \beta_6 SF + u$, необходимо проверить гипотезу о выполнении следующих ограничений: $H_0: \beta_2 = 2\beta_3 = 2\beta_4$. Проверим, так ли это. Если мы подставим наши два ограничения в модель (4), мы получим:

$$S = \beta_1 + \beta_2 \left(IQ_1 + \frac{1}{2} IQ_2 + \frac{1}{2} IQ_3 \right) + \beta_5 SM + \beta_6 SF + u$$

Т.е. с точностью до нормировки (сумма весов должна составлять 1) мы учли то, что IQ_1 весит в два раза больше, чем IQ_2 и IQ_3 .

Соответственно, (4) - это регрессия без ограничений, (1) - это регрессия с ограничениями.

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} = \frac{(2100.646 - RSS_{UR})/2}{RSS_{UR}/(540-6)}. \text{ Осталось только посчитать } RSS_{UR}, \text{ то есть } RSS_{(4)}:$$

$$RSS_{(4)} = (1 - R_{(4)}^2) \cdot TSS_{(4)} = (1 - R_{(4)}^2) \cdot TSS_{(1)} = (1 - R_{(4)}^2) \cdot \frac{RSS_{(1)}}{(1 - R_{(1)}^2)} = (1 - 0.352) \cdot \frac{2100.646}{(1 - 0.336)} = 2050.028$$

(здесь мы воспользовались тем фактом, что TSS во всех моделях одинаковый ($TSS_R = TSS_{UR}$), т.к. TSS - это сумма квадратов отклонений зависимой переменной, а она во все трех регрессиях одинакова). Таким образом, $F = \frac{(2100.646 - 2050.028)/2}{2050.028/(540-6)} \approx 6.5$. Т.к. 5%-ая критическая точка

соответствующего F-распределения есть $F_{5\%}(2, 534) \approx F_{5\%}(2, 500) = 3.01$, мы отвергаем нулевую гипотезу на 5% уровне значимости.

Здесь мы вспомнили полезное свойство $TSS_R = TSS_{UR}$. Также можно пользоваться готовой формулой:

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n-k)} = \frac{(0.352 - 0.336)/2}{(1 - 0.336)/(540-6)} \approx 6.5.$$

Задача 5.

С помощью модели

$$\ln Y = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + u$$

по данным для 30 фирм была оценена зависимость выпуска Y от труда L и капитала K:

$$\widehat{\ln Y} = 1.2 + \underset{(0.3)}{0.6} \ln L + \underset{(0.12)}{0.4} \ln K, F_{\text{statistic}} = 200.24$$

В скобках указаны значения стандартных ошибок. На уровне значимости 5 % отвергаются гипотезы:

- а) $H_0: \beta_2 = 0$
- б) $H_0: \beta_3 = 0$
- в) $H_0: \beta_2 = \beta_3 = 0$
- г) $H_0: \beta_2 = 0.5$
- д) $H_0: \beta_3 = 0.5$

Решение:

а) Проверяем второй коэффициент на значимость: $t_{b_2} = \frac{0.6}{0.12} = 5$, $t_{2,5\%}(30-3) = 2.052$, т.о. коэффициент значим на 5% уровне.

б) Проверяем третий коэффициент на значимость: $t_{b_3} = \frac{0.4}{0.08} = 5$, $t_{2,5\%}(30-3) = 2.052$, т.о. коэффициент значим на 5% уровне.

в) Проверяем модель на адекватность: $F_0 = 200.24$, $F_{(2,27)}^{5\%} = 3.35$, т.о. модель адекватна, H_0 отвергается на 5% уровне значимости.

г) $t_{b_2} = \frac{0.6 - 0.5}{0.12} = 0.833$, $t_{2,5\%}(27) = 2.052$, т.о. $H_0: \beta_2 = 0.5$ не отвергается на 5% уровне значимости.

д) $t_{b_3} = \frac{0.4 - 0.5}{0.12} = -0.833$, $t_{2,5\%}(27) = 2.052$, т.о. $H_0: \beta_3 = 0.5$ не отвергается на 5% уровне значимости.