Конспектировали: Марк Любимов Галим Бикбов.

Упражнение:
Пусть даны факторы x, y, z, тогда обозначим матрицу, состоящую из данных факторов, как:

$$X = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix}$$

$\widetilde{X}$ — центрированная матрица для X.
Дана ковариационная матрица:

$$\frac{\widetilde{X}^T \cdot \widetilde{X}}{(n-1)} = \begin{pmatrix} 28 & 110 & 844 \\ 110 & 644 & 3429 \\ 844 & 3429 & 26585 \end{pmatrix}$$

$$\bar{x} = 15.4;$$

$$\bar{y} = 43;$$

$$\bar{z} = 265;$$

$$n = 50.$$

Рассматривается регрессионная модель:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_x x_i + \hat{\beta}_z z_i$$

Найти:

1. $\hat{\beta}_x, \hat{\beta}_x, \hat{\beta}_1$;
2. RSS, TSS, ESS;
3. При условии, что наблюдений много и $\alpha = 0.05$, проверить гипотезу H_0 против H_1:
   а) $H_0 : \beta_x = 0$
   б) $H_1 : \beta_x \neq 0$

Решение:
Зная ковариационную матрицу, можно найти W и M

$$W = \widetilde{X}^T \cdot \widetilde{X} = \begin{pmatrix} 1372 & 5390 & 41356 \\ 5390 & 31556 & 168021 \\ 41356 & 168021 & 1302665 \end{pmatrix}$$

$$M = W^{-1} = \begin{pmatrix} 0.01702223 & -0.0000961 & -0.00052801 \\ -0.0000961 & 0.00010171 & -0.00001007 \\ -0.00052801 & -0.00001007 & 0.00001883 \end{pmatrix}$$

Вспомним,что предскавляют собой элементы матрицы M

$$M = W^{-1} = \begin{pmatrix} \dots & \dots & \dots \\ \frac{-\hat{\beta}_x}{RSS} & \frac{1}{RSS} & \frac{-\hat{\beta}_z}{RSS} \\ \dots & \dots & \dots \end{pmatrix}$$

Тогда:

$$RSS = \frac{1}{0.00010171} = 9831.87$$

следовательно

$$\hat{\beta}_x = 0.0000961 \cdot RSS = 0.945$$

и

$$\hat{\beta}_z = 0.00001007 \cdot RSS = 0.099$$

Чтобы найти $\hat{\beta}_1$, просуммируем все $\hat{y}_i$ и поделим на n

$$\bar{y}_i = \hat{\beta}_1 + \hat{\beta}_x \bar{x}_i + \hat{\beta}_z \bar{z}_i$$

Отсюда найдем $\hat{\beta}_1$:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_x \bar{x} - \hat{\beta}_z \bar{z} = 43 - 0.945 \cdot 15.4 - 0.099 \cdot 265 = 2.21$$

TSS известно из матрицы W:

$$TSS = 31556$$

Теперь легко посчитать R_2:

$$R^2 = 1 - \frac{RSS}{TSS} = 0.688$$

Перейдем к проверке гипотезы. Для этого необходимо посчитать статистику t:

$$t = \frac{\hat{\beta}_x - 0}{se(\hat{\beta}_x)}$$

Дисперсии регрессионных коэффициентов равны:

$$\mathrm{Var}\begin{pmatrix} \hat{\beta}_x \\ \hat{\beta}_z \end{pmatrix} = \frac{RSS}{n-k} \cdot (\widetilde{X}_y^T \cdot \widetilde{X}_y)^{-1}$$

где

$$X_y = \begin{pmatrix} x_1 & z_1 \\ x_2 & z_2 \\ \vdots & \vdots \\ x_n & z_n \end{pmatrix}$$

$$(\widetilde{X}_y^T \cdot \widetilde{X}_y)^{-1} = \begin{pmatrix} 451.2 & 12651.8 \\ 12651.8 & 407874.1 \end{pmatrix}$$

Тогда стандартная ошибка равна:

$$se(\hat{\beta}_x) = \sqrt{\frac{451.2 \cdot 9831.87}{47}} = 307.2$$

Посчитаем статистику t:

$$t = \frac{0.945}{307.2} = 0.003 < 1.68$$

Таким обращом гипотеза не отвергается. Решение для $\hat{\beta}_z$ аналогично.

Утверждение: Выразить

$$\text{Var}(\hat{\beta}_x)$$

Пусть даная центрированная матрица X_y. Рассмотрим регрессию

$$\hat{y} = X_y \cdot \hat{\beta}$$

Выведем формулу для частного случая: 7-го коэффциента.

$$\text{Var}(\hat{\beta}_7) = \sigma^2 \cdot (X_y^T \cdot X_y)^{-1} = \frac{\sigma^2}{RSS_7} = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \cdot \frac{1}{\frac{RSS_7}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \cdot \frac{1}{1 - R_7^2}$$

то есть

$$\text{Var}(\hat{\beta}_x^{mult}) = \text{Var}(\hat{\beta}_x^{simple}) \cdot \frac{1}{1 - R_x^2} = \text{Var}(\hat{\beta}_x^{simple}) \cdot \frac{1}{VIF_x}$$

Головоломка:

Злобный Дракон поймал двух принцесс, Читу и Ниту, и посадил их в разные башни своего замка. Затем Злобный Дракон подбросил правильную монетку бесконечное число раз. Все результаты чётных бросков он сообщил Чите, а все результаты нечётных — Ните. Далее Дракон предлагает каждой из принцесс назвать номер любого подбрасывания, результат которого ей не известен. То есть Чита должна назвать нечётный номер, а Нита — чётный.

Если результаты бросков, названных Читой и Нитой, одинаковые, то Злобный Дракон дарит каждой принцессе тортик, розового плюшевого зайца и отпускает на свободу. Если же результаты бросков отличаются, то Злобный Дракон съедает Читу и Ниту с клюквенным вареньем. Дракон обожает принцесс и клюквенное варенье!

Принцессы знают о повадках Злобного Дракона и могли заранее до похищения договориться о стратегиях.

1. Каковы шансы принцесс спастись, если Нита называет число первой, а Чита называет своё число, зная что выбрала Нита?

2. Каковы шансы принцесс спастись, если Чита и Нита после похищения лишены возможности передавать информацию друг другу?

Решение находится после Гетероскедастичности

1. Гетероскедатичность

Среди условий теоремы Гаусса-Маркова есть предпосылка $\text{Var}(u_i(x)) = \sigma^2$. Теперь $\text{Var}(u_i|x)) = h(x_i)$

Теорма Гаусса-Маркова неприменима, тогда что случится с оценкой $\hat{\beta} = (X^T X)^{-1} X^{-1} y$?

Сохраняется:

1. единственность
2. линейность по y
3. несмещенность $E(u_i|x) = 0$

Состоятельность сохраняется при некоторых дополнительных предпосылках.
Пропало:

1. У $\hat{\beta}$ наименьшая дисперсия среди других линейно несмещенных оценок
2. теорема о распределении разных статистик

$$\frac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim t_{n-k}$$

$$\frac{(RSS_R - RSS_{UR})/(k_{UR} - k_R)}{RSS_{UR}/(n - k_{UR})} \sim F_{k_{UR} - k_R, n - k_{UR}}$$

где k_R — число регрессоров в регрессии с ограничениями, k_{UR} — число регрессоров в регрессии без ограничений

Сами оценки ничего, они не плохие, но нельзя проверять гипотезы.

Упр. Хотим скорректировать формулы:

1.

$$y_i = \beta x_i + u_i$$

$$\text{Var}(u_i|x_i) = {\sigma_i}^2$$

$$\text{Var}(\hat{\beta}|x) = \text{Var}\left(\frac{\sum x_i y_i}{\sum {x_i}^2}|x\right) = \frac{\sum {x_i}^2 \text{Var}(y_i|x_i)}{\sum {x_i}^2} = \frac{\sum {x_i}^2 {\sigma_i}^2}{(\sum x_i)^2}$$

2.

$$y = X \cdot \beta + u$$

$$\text{Var}(u|x) = \Omega$$

$$\text{Var}(\hat{\beta}|x) = \text{Var}((X^T X)^{-1} X^{-1} y|X) = X^T X^{-1} X^T \text{Var}(y|X) X (X^T X)^{-1} =$$
$$= (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

Раньше, в случае гомоскедастичности, мы предполагали, что $\Omega_{ii} = \sigma^2$, $\text{Var}(\hat{\beta}|x) = \frac{RSS}{n-k}(X^T X)^{-1}$

Как оценить ${\sigma_i}^2$?

$${\sigma_i}^2 = \text{Var}(u_i|X) = E({u_i}^2|X)$$

Оценки стандартных ошибок, полученные в предположении гомоскедатичности, в случае гетероскедатичности становятся несостоятельны. Для того, чтобы получить состоятельные оценки, используются различные корректировки.

Корректировка Уайта, HC_0

$$E({u_i}^2|x) \approx {u_i}^2 \approx {\hat{u}_i}^2$$

$$\hat{\text{Var}}_{HC_0}(\hat{\beta}|X) = (X^T X)^{-1} X^T \hat{\Omega} X (X^T X)^{-1}$$

Утверждение. Если :

1. $y = X \cdot \beta + u$
2. $\mathrm{Var}(u|X) = \Omega$
3. β — константа, ${\sigma_i}^2 = f(x_i)$
4. Строки x_i независимы и одинаково распределены
5. $E({x^4}_{ij}) < \infty$

То :

$$\frac{\hat{\beta} - \beta}{se_{HC}(\hat{\beta})} \xrightarrow[n \longrightarrow \infty]{} N(0;1)$$

HC_1 :

$$\sum_{i=1}^{n} E({u_i}^2|X) = n \cdot \sigma^2$$

$$\sum_{i=1}^{n} E(\hat{u}_i^2|X) = E(RSS|X) = (n-k) \cdot \sigma^2$$

$$\hat{\Omega}_{HC_1} = \frac{n}{n-k} \hat{\Omega}_{HC_0}$$

HC_2:

$$\sum_{i=1}^{n} E({u_i}^2|X) = n \cdot \sigma^2$$

$$\sum_{i=1}^{n} E(\hat{u}_i^2|X) = \mathrm{Var}(\hat{u}_i|X)$$

$$\hat{u} = (I - H)y = y - \hat{y} = y - Hy$$

$$\mathrm{Var}(u_i|X) = (1 - h_{ii}) \cdot \sigma^2$$

Корректировка HC_3 основана на кроссвалидации.
Экспериментальное исследование:
HC_0 плохо
HC_1 лучше HC_0
HC_2 лучше HC_1
HC_3 немного лучше HC_2

Список литературы

[1] Achim Zeileis, Econometric Computing with HC and HAC Covariance Matrix Estimators

Возможное решение задачи про Дракона:

1. Нита называет такой номер, до которого был орёл, а после — Решка. Чита, зная, что находится на названном Нитой месте, однозначно угадывает.
2. Приведём пример стратегии, гарантирующей вероятность спасения 2/3. Чита отступает от первого орла на бросок назад, а Нита — на бросок вперёд.