

1. Пусть $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i$ — регрессионная модель, где $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$,

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}, \mathbb{E}(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 I. \text{ Для удобства расчётов даны матрицы:}$$

$$X'X = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \text{ и } (X'X)' = \begin{pmatrix} 0.3333 & -0.3333 & 0.0000 \\ -0.3333 & 1.3333 & -1.0000 \\ 0.0000 & -1.0000 & 2.0000 \end{pmatrix}$$

- (a) Укажите число наблюдений
- (b) Укажите число регрессоров в модели, учитывая свободный член
- (c) Найдите $TSS = \sum_{i=1}^n (y_i - \bar{y})^2$
- (d) Найдите $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- (e) Методом МНК найдите оценку для вектора неизвестных коэффициентов
- (f) Чему равен R^2 в модели? Прокомментируйте полученное значение с точки зрения качества оценённого уравнения регрессии
- (g) Сформулируйте основную и альтернативную гипотезы, которые соответствуют тесту на значимость переменной x_1 в уравнении регрессии
- (h) Протестируйте на значимость переменную x_1 в уравнении регрессии на уровне значимости 10%:
 - i. Приведите формулу для тестовой статистики
 - ii. Укажите распределение тестовой статистики
 - iii. Вычислите наблюдаемое значение тестовой статистики
 - iv. Укажите границы области, где основная гипотеза не отвергается
 - v. Сделайте статистический вывод о значимости переменной x_1
- (i) Найдите P –значение, соответствующее наблюдаемому значению тестовой статистики (T_{obs}) из предыдущего пункта. На основе полученного P –значения сделайте вывод о значимости переменной x_1

- (j) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 \neq 1$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (k) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 > 1$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (l) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 < 1$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (m) Сформулируйте основную гипотезу, которая соответствует тесту на значимость регрессии «в целом»
- (n) На уровне значимости 5% проверьте гипотезу о значимости регрессии «в целом»:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (o) Найдите P –значение, соответствующее наблюдаемому значению тестовой статистики (T_{obs}) из предыдущего пункта. На основе полученного P –значения сделайте вывод о значимости регрессии «в целом»

- (p) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 \neq 2$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (q) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 > 2$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (r) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 < 2$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод

Решение

(a) $n = 5$

(b) $k = 3$

(c) $TSS = 10$

(d) $RSS = 2$

(e) $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$

- (f) $R^2 = 1 - \frac{RSS}{TSS} = 0.8$. R^2 высокий, построенная эконометрическая модель «хорошо» описывает данные
- (g) Основная гипотеза — $H_0 : \beta_1 = 0$, альтернативная гипотеза — $H_a : \beta_1 \neq 0$
- (h) Проверка гипотезы
- i. $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}$; $n = 5$; $k = 3$
 - ii. $T \sim t(n - k)$; $n = 5$; $k = 3$
 - iii. $T_{obs} = \frac{\hat{\beta}_1 - 0}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-0}{\sqrt{\frac{2}{5-3}1.3333}} = 1.7321$
 - iv. Нижняя граница = -2.920 , верхняя граница = 2.920
 - v. Поскольку $T_{obs} = 1.7321$, что принадлежит промежутку от -2.920 до 2.920 , то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 10%
- (i) $p\text{-value}(T_{obs}) = \mathbb{P}(|T| > |T_{obs}|) = 2F_T(|T_{obs}|)$, где $F_T(|T_{obs}|)$ — функция распределения t -распределения с $n - k = 5 - 3 = 2$ степенями свободы в точке $|T_{obs}|$. $p\text{-value}(T_{obs}) = 2tcdf(-|T_{obs}|, n - k) = 2tcdf(-1.7321, 2) = 0.2253$. Поскольку P -значение превосходит уровень значимости 10%, то основная гипотеза — $H_0 : \beta_1 = 0$ не может быть отвергнута
- (j) Проверка гипотезы
- i. $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}$; $n = 5$; $k = 3$
 - ii. $T \sim t(n - k)$; $n = 5$; $k = 3$
 - iii. $T_{obs} = \frac{\hat{\beta}_1 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$
 - iv. Нижняя граница = -2.920 , верхняя граница = 2.920
 - v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от -2.920 до 2.920 , то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 10%
- (k) Проверка гипотезы
- i. $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}$; $n = 5$; $k = 3$
 - ii. $T \sim t(n - k)$; $n = 5$; $k = 3$
 - iii. $T_{obs} = \frac{\hat{\beta}_1 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$
 - iv. Нижняя граница = $-\infty$, верхняя граница = 1.8856
 - v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от $-\infty$ до 1.8856 , то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 10%

(l) Проверка гипотезы

i. $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}; n = 5; k = 3$

ii. $T \sim t(n - k); n = 5; k = 3$

iii. $T_{obs} = \frac{\hat{\beta}_1 - 1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$

iv. Нижняя граница = -1.8856 , верхняя граница = $+\infty$

v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от -1.8856 до $+\infty$, то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 10%

(m) Основная гипотеза — $H_0 : \beta_1 = \beta_2 = 0$, альтернативная гипотеза — $H_a : |\beta_1| + |\beta_2| > 0$

(n) Проверка гипотезы

i. $T = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k}; n = 5; k = 3$

ii. $T \sim F(n - k); n = 5; k = 3$

iii. $T_{obs} = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k} = \frac{0.8}{1-0.8} \cdot \frac{5-3}{2} = 4$

iv. Нижняя граница = 0, верхняя граница = 19

v. Поскольку $T_{obs} = 4$, что принадлежит промежутку от 0 до 19, то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 5%. Следовательно, регрессия в целом незначима. Напомним, что $R^2 = 0.8$, то есть он высокий. Но при этом регрессия «в целом» незначима. Такой эффект может возникать при малом объёме выборки, например, таком, как в данной задаче

(o) $p\text{-value}(T_{obs}) = \mathbb{P}(|T| > |T_{obs}|) = 2F_T(|T_{obs}|)$, где $F_T(|T_{obs}|)$ — функция распределения F -распределения с $k = 3$ и $n - k = 5 - 3 = 2$ степенями свободы в точке T_{obs} .
 $p\text{-value}(T_{obs}) = 1 - fcdf(-|T_{obs}|, n - k) = 1 - fcdf(4, 2) = 0.2$. Поскольку P -значение превосходит уровень значимости 10%, то основная гипотеза — $H_0 : \beta_1 = \beta_2 = 0$ не может быть отвергнута. Таким образом, регрессия «в целом» незначима

(p) Проверка гипотезы

i. $T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - (\beta_1 + \beta_2)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}}$, где $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \widehat{\text{Var}}(\hat{\beta}_1) + \widehat{\text{Var}}(\hat{\beta}_2) + 2\widehat{\text{Cov}}(\hat{\beta}_1; \hat{\beta}_2) = \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{23} + \hat{\sigma}^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$

ii. $T \sim t(n - k); n = 5; k = 3$

iii. $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 2}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$

iv. Нижняя граница = -4.3027 , верхняя граница = 4.3027

- v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от -4.3027 до 4.3027 , то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 5%

(q) Проверка гипотезы

- i. $T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - (\beta_1 + \beta_2)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}}$, где $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \widehat{\text{Var}}(\hat{\beta}_1) + \widehat{\text{Var}}(\hat{\beta}_2) + 2\widehat{\text{Cov}}(\hat{\beta}_1; \hat{\beta}_2) = \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{23} + \hat{\sigma}^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$
- ii. $T \sim t(n-k); n = 5; k = 3$
- iii. $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 2}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$
- iv. Нижняя граница $= -\infty$, верхняя граница $= 2.9200$
- v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от $-\infty$ до 2.9200 , то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 5%

(r) Проверка гипотезы

- i. $T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - (\beta_1 + \beta_2)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}}$, где $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \widehat{\text{Var}}(\hat{\beta}_1) + \widehat{\text{Var}}(\hat{\beta}_2) + 2\widehat{\text{Cov}}(\hat{\beta}_1; \hat{\beta}_2) = \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{23} + \hat{\sigma}^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$
- ii. $T \sim t(n-k); n = 5; k = 3$
- iii. $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 2}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$
- iv. Нижняя граница $= -2.9200$, верхняя граница $= +\infty$
- v. Поскольку $T_{obs} = 0.8660$, что принадлежит промежутку от -2.9200 до $+\infty$, то на основе имеющихся данных нельзя отвергнуть основную гипотезу на уровне значимости 5%

2. На основе 100 наблюдений была оценена функция спроса:

$$\widehat{\ln Q} = 0.87 - 1.23 \ln P$$

(s.e.) (0.04) (0.02)

Значимо ли коэффициент эластичности спроса по цене отличается от -1 ? Рассмотрите уровень значимости 5%.

3. На основе 100 наблюдений была оценена функция спроса:

$$\widehat{\ln Q} = 2.87 - 1.12 \ln P$$

(s.e.) (0.04) (0.02)

На уровне значимости 5% проверьте гипотезу $H_0 : \beta_{\ln P} = -1$ против альтернативной $H_a : \beta_{\ln P} < -1$. Дайте экономическую интерпретацию проверяемой гипотезе и альтернативе.

4. Используя годовые данные с 1960 по 2005 г., была построена кривая Филлипса, связывающая уровень инфляции Inf и уровень безработицы $Unem$:

$$\widehat{Inf} = 2.34 - 0.23Unem$$

$$\sqrt{\widehat{Var}(\hat{\beta}_{Unem})} = 0.04, R^2 = 0.12$$

На уровне значимости 1% проверьте гипотезу $H_0 : \beta_{Unem} = 0$ против альтернативной $H_a : \beta_{Unem} \neq 0$.

5. Была оценена функция Кобба-Дугласа с учётом человеческого капитала H (K — физический капитал, L — труд):

$$\widehat{\ln Q} = 1.4 + 0.46 \ln L + 0.27 \ln H + 0.23 \ln K$$

$$ESS = 170.4, RSS = 80.3, n = 21$$

- (a) Чему равен коэффициент R^2 ?
- (b) На уровне значимости 1% проверьте гипотезу о значимости регрессии «в целом»
6. На основе опроса 25 человек была оценена следующая модель зависимости логарифма зарплаты $\ln W$ от уровня образования Edu (в годах) и возраста Age :

$$\widehat{\ln W} = 1.7 + 0.5Edu + 0.06Age - 0.0004Age^2$$

$$ESS = 90.3, RSS = 60.4$$

Когда в модель были введены переменные $Fedu$ и $Medu$, учитывающие уровень образования родителей, величина ESS увеличилась до 110.3.

- (a) Напишите спецификацию уравнения регрессии с учётом образования родителей
- (b) Сформулируйте и на уровне значимости 5% проверьте гипотезу о значимом влиянии уровня образования родителей на заработную плату:
- i. Сформулируйте гипотезу
 - ii. Приведите формулу для тестовой статистики

- iii. Укажите распределение тестовой статистики
- iv. Вычислите наблюдаемое значение тестовой статистики
- v. Укажите границы области, где основная гипотеза не отвергается
- vi. Сделайте статистический вывод

Решение

Ограниченная модель (Restricted model):

$$\ln W_i = \beta + \beta_{Edu} Edu_i + \beta_{Age} Age_i + \beta_{Age^2} Age_i^2 + \varepsilon_i$$

Неограниченная модель (Unrestricted model):

$$\ln W_i = \beta + \beta_{Edu} Edu_i + \beta_{Age} Age_i + \beta_{Age^2} Age_i^2 + \beta_{Fedu} Fedu_i + \beta_{Medu} Medu_i + \varepsilon_i$$

По условию $ESS_R = 90.3$, $RSS_R = 60.4$, $TSS = ESS_R + RSS_R = 90.3 + 60.4 = 150.7$. Также сказано, что $ESS_{UR} = 110.3$. Значит, $RSS_{UR} = TSS - ESS_{UR} = 150.7 - 110.3 = 40.4$

(a) Спецификация:

$$\ln W_i = \beta + \beta_{Edu} Edu_i + \beta_{Age} Age_i + \beta_{Age^2} Age_i^2 + \beta_{Fedu} Fedu_i + \beta_{Medu} Medu_i + \varepsilon_i$$

(b) Проверка гипотезы

- i. $H_0 : \begin{cases} \beta_{Fedu} = 0 \\ \beta_{Medu} = 0 \end{cases} \quad H_a : |\beta_{Fedu}| + |\beta_{Medu}| > 0$
- ii. $T = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$, где $q = 2$ — число линейно независимых уравнений в основной гипотезе H_0 , $n = 25$ — число наблюдений, $k = 6$ — число коэффициентов в модели без ограничения
- iii. $T \sim F(q; n - k)$
- iv. $T_{obs} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} = \frac{(60.4 - 40.4)/2}{40.4/(25-6)} = 4.70$
- v. Нижняя граница = 0, верхняя граница = 3.52
- vi. Поскольку $T_{obs} = 4.70$, что не принадлежит промежутку от 0 до 3.52, то на основе имеющихся данных можно отвергнуть основную гипотезу на уровне значимости 5%. Таким образом, образование родителей существенно влияет на заработную плату.

7. Рассмотрим следующую модель зависимости цены дома $Price$ (в тысячах долларов) от его площади $Hsize$ (в квадратных метрах), площади участка $Lsize$ (в квадратных метрах), числа ванных комнат $Bath$ и числа спален BDR :

$$\widehat{Price} = \hat{\beta}_1 + \hat{\beta}_2 Hsize + \hat{\beta}_3 Lsize + \hat{\beta}_4 Bath + \hat{\beta}_5 BDR$$

$$R^2 = 0.218, n = 23$$

Напишите спецификацию регрессии с ограничениями для проверки статистической гипотезы $H_0 : \beta_3 = 20\beta_4$. Дайте интерпретацию проверяемой гипотезе. Для регрессии с ограничением был вычислен коэффициент $R_R^2 = 0.136$. На уровне значимости 5% проверьте нулевую гипотезу.

8. Рассмотрим следующую модель зависимости почасовой оплаты труда W от уровня образования $Educ$, возраста Age , уровня образования родителей $Fathedu$ и $Mothedu$:

$$\widehat{\ln W} = \hat{\beta}_1 + \hat{\beta}_2 Educ + \hat{\beta}_3 Age + \hat{\beta}_4 Age^2 + \hat{\beta}_5 Fathedu + \hat{\beta}_6 Mothedu$$

$$R^2 = 0.341, n = 27$$

Напишите спецификацию регрессии с ограничениями для проверки статистической гипотезы $H_0 : \beta_5 = 2\beta_4$. Дайте интерпретацию проверяемой гипотезе. Для регрессии с ограничением был вычислен коэффициент $R_R^2 = 0.296$. На уровне значимости 5% проверьте нулевую гипотезу.

9. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$,

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}, \mathbb{E}(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I.$$

На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 \neq 2$:

- (a) Приведите формулу для тестовой статистики
- (b) Укажите распределение тестовой статистики
- (c) Вычислите наблюдаемое значение тестовой статистики
- (d) Укажите границы области, где основная гипотеза не отвергается
- (e) Сделайте статистический вывод

10. Пусть $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и $i = 1, \dots, 18$ — классическая регрессионная модель, где $\mathbb{E}(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$. Также имеются следующие данные: $\sum_{i=1}^{18} y_i^2 = 4256$, $\sum_{i=1}^{18} x_i^2 = 185$, $\sum_{i=1}^{18} x_i y_i = 814.25$, $\sum_{i=1}^{18} y_i = 225$, $\sum_{i=1}^{18} x_i = 49.5$. Используя эти данные, оцените эту регрессию и на уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 = 3.5$ против альтернативной $H_a : \beta_1 > 3.5$:

- (a) Приведите формулу для тестовой статистики
- (b) Укажите распределение тестовой статистики
- (c) Вычислите наблюдаемое значение тестовой статистики
- (d) Укажите границы области, где основная гипотеза не отвергается
- (e) Сделайте статистический вывод

11. По данным для 27 фирм исследователь оценил зависимость объёма выпуска y от труда l и капитала k с помощью двух моделей:

$$\ln y_i = \beta_1 + \beta_2 \ln l_i + \beta_3 \ln k_i + \varepsilon_i$$

$$\ln y_i = \beta_1 + \beta_2 \ln(l_i \cdot k_i) + \varepsilon_i$$

Он получил для этих двух моделей суммы квадратов остатков $RSS_1 = 0.851$ и $RSS_2 = 0.894$ соответственно. Сформулируйте гипотезу, которую хотел проверить исследователь. На уровне значимости 5% проверьте эту гипотезу и дайте экономическую интерпретацию.

12. Пусть задана линейная регрессионная модель:

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \beta_4 x_{3i} + \beta_5 x_{4i} + \varepsilon_i, i = 1, \dots, 20$$

По имеющимся данным оценены следующие регрессии:

$$\underset{(s.e.)}{\hat{y}_i} = 10.01 + \underset{(0.15)}{1.05}x_1 + \underset{(0.06)}{2.06}x_2 + \underset{(0.04)}{0.49}x_3 - \underset{(0.06)}{1.31}x_4, RSS = 6.85$$

$$y_i - \widehat{x_1 - 2x_2} = 10.00 + 0.50x_3 - 1.32x_4, RSS = 8.31$$

(s.e.) (0.15) (0.07) (0.06)

$$y_i + \widehat{x_1 + 2x_2} = 9.93 + 0.56x_3 - 1.50x_4, RSS = 4310.62$$

(s.e.) (3.62) (1.48) (1.42)

$$y_i - \widehat{x_1 + 2x_2} = 10.71 + 0.09x_3 - 1.28x_4, RSS = 3496.85$$

(s.e.) (3.26) (1.33) (1.28)

$$y_i + \widehat{x_1 - 2x_2} = 9.22 + 0.97x_3 - 1.54x_4, RSS = 516.23$$

(s.e.) (1.25) (0.51) (0.49)

На уровне значимости 5% проверьте гипотезу $H_0 : \begin{cases} \beta_2 = 1 \\ \beta_3 = 2 \end{cases}$ против альтернативной гипотезы $H_a : |\beta_2 - 1| + |\beta_3 - 2| \neq 0$.

13. Рассмотрим следующую модель зависимости расходов на образование на душу населения от дохода на душу населения, доли населения в возрасте до 18 лет, а также доли городского населения:

$$education_i = \beta_1 + \beta_2 income_i + \beta_3 young_i + \beta_4 urban_i + \varepsilon_i$$

Ниже приведены результаты оценивания уравнения этой линейной регрессии:

$$\widehat{education}_i = -\frac{287}{(64.9199)} + 0.0807 \cdot income_i + \frac{0.817}{(0.1598)} \cdot young_i - \frac{0.106}{(0.0343)} \cdot urban_i$$

	Estimate	St.Error	t value	P-value
Intercept	-286.84	64.92	-4.42	0.00
Income	0.08	0.01	8.67	0.00
Young	0.82	0.16	5.12	0.00
Urban	-0.11	0.03	-3.09	0.00

- (a) Сформулируйте основную и альтернативную гипотезы, которые соответствуют тесту на значимость коэффициента при переменной доход на душу населения в уравнении регрессии
- (b) На уровне значимости 10% проверьте гипотезу о значимости коэффициента при переменной доход на душу населения в уравнении регрессии:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод

- (с) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 > 1$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (d) Сформулируйте основную гипотезу, которая соответствует тесту на значимость регрессии «в целом»
- (е) На уровне значимости 1% проверьте гипотезу о значимости регрессии «в целом», если известно, что F –статистика равна 34.81 со степенями свободы 3 и 47, P –значение равно $5.337e^{-12}$:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (f) Далее приведены результаты оценивания уравнения регрессии без переменной, отражающей долю городского населения:

$$\widehat{education}_i = - \frac{301}{(70.27134)} + \frac{0.0612}{(0.00741)} \cdot income_i + \frac{0.836}{(0.17327)} \cdot young_i$$

	Estimate	St.Error	t value	P-value
Intercept	-301.09	70.27	-4.28	0.00
Income	0.06	0.01	8.25	0.00
Young	0.84	0.17	4.83	0.00

Также известно, что RSS для первой модели равен 33489.35, а для второй модели — 40276.61. На уровне значимости 5% проверьте гипотезу $H_0 : \beta_4 = 0$ против альтернативной $H_0 : \beta_4 \neq 0$:

- Приведите формулу для тестовой статистики
- Укажите распределение тестовой статистики
- Вычислите наблюдаемое значение тестовой статистики
- Укажите границы области, где основная гипотеза не отвергается

v. Сделайте статистический вывод

14. Рассмотрим следующую модель зависимости расходов на образование на душу населения от дохода на душу населения, доли населения в возрасте до 18 лет, а также доли городского населения:

$$education_i = \beta_1 + \beta_2 income_i + \beta_3 young_i + \beta_4 urban_i + \varepsilon_i$$

Ниже приведены результаты оценивания уравнения этой линейной регрессии:

$$\widehat{education}_i = -\frac{287}{(64.9199)} + \frac{0.0807}{(0.0093)} \cdot income_i + \frac{0.817}{(0.1598)} \cdot young_i - \frac{0.106}{(0.0343)} \cdot urban_i$$

	Estimate	St.Error	t value	P-value
Intercept	-286.84	64.92	-4.42	0.00
Income	0.08	0.01	8.67	0.00
Young	0.82	0.16	5.12	0.00
Urban	-0.11	0.03	-3.09	0.00

- (a) Сформулируйте основную и альтернативную гипотезы, которые соответствуют тесту на значимость коэффициента при переменной доля населения в возрасте до 18 лет в уравнении регрессии
- (b) На уровне значимости 10% проверьте гипотезу о значимости коэффициента при переменной доля населения в возрасте до 18 лет в уравнении регрессии:
- Приведите формулу для тестовой статистики
 - Укажите распределение тестовой статистики
 - Вычислите наблюдаемое значение тестовой статистики
 - Укажите границы области, где основная гипотеза не отвергается
 - Сделайте статистический вывод
- (c) Далее приведены результаты оценивания уравнения регрессии без переменной, отражающей долю населения в возрасте до 18 лет:

$$\widehat{education}_i = \frac{25.3}{(27.3827)} + \frac{0.0762}{(0.0114)} \cdot income_i - \frac{0.112}{(0.0423)} \cdot urban_i$$

	Estimate	St.Error	t value	P-value
Intercept	25.25	27.38	0.92	0.36
Income	0.08	0.01	6.67	0.00
Urban	-0.11	0.04	-2.66	0.01

Также известно, что RSS для первой модели равен 33489.35, а для второй модели — 52132.29. На уровне значимости 5% проверьте гипотезу $H_0 : \beta_3 = 0$ против альтернативной $H_0 : \beta_3 \neq 0$:

- i. Приведите формулу для тестовой статистики
- ii. Укажите распределение тестовой статистики
- iii. Вычислите наблюдаемое значение тестовой статистики
- iv. Укажите границы области, где основная гипотеза не отвергается
- v. Сделайте статистический вывод