Эконометрика

с Монте-Карло и эконометрессами

в задачах и упражнениях
Решебник

Дмитрий Борзых, Борис Демешев

25 февраля 2015 г.

Оглавление

1 Решения и ответы к избранным задачам 3

```
library("knitr")
library("tikzsetup")
library("tikzDevice")
tikzsetup()

library("ggplot2")
library("Hmisc")
library("lmtest")
library("apsrtable")
library("xtable")
library("MASS")
library("car")
library("texreg")
library("memisc")
library("sandwich")

library("econru")

theme_set(theme_bw())

load('pset_data.Rdata')
```

Глава 1

Решения и ответы к избранным задачам

1.1. да, да, да, нет

1.2.

1.3.

1.4.

1.5. $\hat{\alpha} + \hat{\gamma} = 0$ и $\hat{\beta} + \hat{\delta} = 1$

1.6.

1.7.

1.8. $\hat{\beta} = \sum x_i y_i / \sum x_i^2$

1.9. $\hat{\beta} = \bar{y}$

1.10. $\hat{\beta}_2 = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sum (x_i - \bar{x})^2$, $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

1.11. $\hat{\beta} = \sum x_i (y_i - 1) / \sum x_i^2$

1.12. $(300 - \hat{\beta}_1)^2 + (200 - \hat{\beta}_2)^2 + (400 - \hat{\beta}_1 - \hat{\beta}_2)^2 \to \min$

1.13. $2 \cdot (10 - \hat{\beta})^2 + (3 - \hat{\beta})^2 \to \min$

1.14.

1.15. да, возможно. Два вытянутых облачка точек. Первое облачко даёт первую регрессию, второе — вторую. Прямая, соединяющая центры облачков, — общую.

1.16. Нет. Коэффициенты можно интерпретировать только «при прочих равных», т.е. при равных x. Из-за разных x может оказаться, что у мужчин $\bar{y}$ меньше, чем $\bar{y}$ для женщин.

1.17.

1.18.

1.19. Оценки МНК линейны по объясняемой переменной. Если сложить объясняемые переменные в этих двух моделях, то получится вектор из единичек. Если строить регрессию вектора из единичек на константу и r, то получатся оценки коэффициентов 1 и 0. Значит, $\hat{\beta}_1 + \hat{\gamma}_1 = 1$, $\hat{\beta}_2 + \hat{\gamma}_2 = 0$

1.20. Увеличатся в 100 раз

1.21. да

1.22. $R^2 = 0$

1.23. $TSS_1 = TSS_2$, $R_2^2 \geqslant R_2^1$, $ESS_2 \geqslant ESS_1$, $RSS_2 \leqslant RSS_1$

1.24.

1.25. $y_i^* = 7 + 3(y_i - \bar{y})/s_y$

2.1.

2.2.

2.3. $c_i = c \cdot x_i$, где $c \neq 0$

2.4.

2.5.

2.6.

2.7.

2.8.

2.9.

2.10.

2.11.

2.12.

2.13. Через теорему Гаусса–Маркова или через условную минимизацию, $c_i = 1/n$

2.14.

2.15.

2.16.

1. $\hat{\beta} = \frac{\sum y_t t}{\sum t^2}$

2. $\mathbb{E}(\hat{\beta}) = \beta$ и $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{t=1}^T t^2}$

3. Да, состоятельна

2.17. несостоятельна

2.18.

2.19. Вроде бы равносильно переносу начала координат и применению результата для регрессии без свободного члена. Должна остаться несмещенность.

2.20.

2.21. Не прав. Ковариация $\mathrm{Cov}(y_i, \hat{y}_i)$ зависит от i, это не одно неизвестное число, для которого можно предложить одну оценку.

2.22. формула $\sum(y_i - \bar{y})^2/(n-1)$ неприменима так как $\mathbb{E}(y_i)$ не является константой

2.23. R^2 — это отношение выборочных дисперсий $\hat{y}$ и y.

2.24. Как отсутствие систематической ошибки.

2.25. нет, нет, нет

2.26. $RSS/\sigma^2 \sim \chi^2_{n-k}$, $\mathbb{E}(RSS) = (n-k)\sigma^2$, $\mathrm{Var}(RSS) = 2(n-k)\sigma^4$, $\mathbb{P}(10\sigma^2 < RSS < 30\sigma^2) \approx 0.898$

2.27.

2.28.

2.29.

2.30. Можно взять четыре наблюдения равноотстоящих по вертикали от данной прямой. Подбирая остатки, добиваемся нужного R^2.

2.31. $\hat{\beta}_1 = -4890$ и $\hat{\beta}_2 = 2.5$

$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \ldots & \ldots \\ 1 & 12 \end{bmatrix}$ — матрица исходных регрессоров; $\tilde{X} = \begin{bmatrix} 1 & 1+1994 \\ 1 & 2+1994 \\ \ldots & \ldots \\ 1 & 12+1994 \end{bmatrix}$ — матрица новых регрессоров.

$\tilde{X} = X \cdot D$, где $D = \begin{bmatrix} 1 & 1994 \\ 0 & 1 \end{bmatrix}$.

Итак, уравнение регрессии с новыми регрессорами имеет вид $y = \tilde{X}\beta + \varepsilon$ и МНК-оценки коэффициентов равны:

$$\hat{\beta} = \left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^T y = \left([XD]^T[XD]\right)^{-1}[XD]^T y = \\ D^{-1}(X^TX)^{-1}(D^T)^{-1}D^TX^Ty = D^{-1}(X^TX)^{-1}X^Ty \quad (1.1)$$

$$\hat{\beta} = D^{-1}\hat{\beta}_{old} = \begin{bmatrix} 1 & -1994 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 95 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -4890 \\ 2.5 \end{bmatrix}$$

2.32. Мы можем существенно упростить решение, воспользовавшись матричным представлением:

$$\tilde{\beta}_2^a = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i} = \frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \\ \frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix} y \quad (1.2)$$

$$\mathbb{E}\tilde{\beta}_2^a = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}y_i}{x_i} = \frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\begin{bmatrix} \mathbb{E}y_1 \\ \mathbb{E}y_2 \\ \vdots \\ \mathbb{E}y_n \end{bmatrix} = \\ \frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\begin{bmatrix} \beta_1+\beta_2x_1 \\ \beta_1+\beta_2x_2 \\ \vdots \\ \beta_1+\beta_2x_n \end{bmatrix} = \\ \frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\left[\beta_1\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_2\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right] = \frac{\beta_1}{n}\sum_{k=1}^{n}\frac{1}{x_k} + \beta_2 \quad (1.3)$$

Значит, смещение для первой оценки равно $\frac{\beta_1}{n}\sum_{k=1}^{n}\frac{1}{x_k}$.

$$\mathrm{Var}(\tilde{\beta}_2^a) = \mathrm{Var}\left(\frac{1}{n}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix} y\right) = \\ \frac{1}{n^2}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\mathrm{Var}(y)\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}^T = \\ \frac{1}{n^2}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\mathrm{Var}(\varepsilon)\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}^T = \\ \frac{1}{n^2}\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}\sigma_\varepsilon^2 I\begin{bmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{bmatrix}^T = \\ \frac{\sigma_\varepsilon^2}{n^2}\sum_{k=1}^{n}\frac{1}{x_k^2} \quad (1.4)$$

Перейдём ко второй оценке.
$\tilde{\beta}_2^b = \frac{\overline{y}}{\overline{x}} = \frac{1}{\overline{x}}\frac{1}{n}\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} y$

$$\mathbb{E}\tilde{\beta}_2^b = \frac{\overline{y}}{\overline{x}} = \frac{1}{\overline{x}}\frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\mathbb{E}y = \frac{1}{\overline{x}}\frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\begin{bmatrix}\beta_1+\beta_2 x_1\\ \beta_1+\beta_2 x_2\\ \vdots\\ \beta_1+\beta_2 x_n\end{bmatrix} =$$

$$\frac{1}{\overline{x}}\frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\left[\beta_1\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} + \beta_2\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\right] =$$

$$\frac{1}{n}\frac{\beta_1 n}{\overline{x}} + \frac{1}{n}\frac{\beta_2\sum x_i}{\overline{x}} = \frac{\beta_1}{\overline{x}} + \beta_2 \quad (1.5)$$

Значит, смещение равно $\frac{\beta_1}{\overline{x}}$.

$$\mathrm{Var}(\tilde{\beta}_2^b) = \frac{1}{\overline{x}^2}\frac{1}{n^2}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\mathrm{Var}(y)\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} =$$

$$\frac{1}{\overline{x}^2}\frac{1}{n^2}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\mathrm{Var}(\varepsilon)\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} = \frac{\sigma_\varepsilon^2}{\overline{x}^2 n} \quad (1.6)$$

2.33. Известно, что для парной регрессии $t^2_{\hat{\beta}_2} = \frac{R^2}{(1-R^2)/(n-2)}$. Поэтому из выражения $t^2_{\hat{\beta}_2} = \frac{0.05^2}{(1-0.05^2)/(n-2)} = \frac{0.05^2(n-2)}{1-0.05^2}$ становится очевидным, что при надлежащем выборе числа наблюдений можно сделать величину $t_{\hat{\beta}_2}$ сколь угодно большой.

2.34. Пусть $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$, $i = 1,\dots,n$.

Тогда $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\varepsilon}_i$

$Y_i - \overline{Y} + \overline{Y} = \hat{\beta}_1 + \hat{\beta}_2(X_i - \overline{X} + \overline{X}) + \hat{\varepsilon}_i$

$Y_i - \overline{Y} = \underbrace{\hat{\beta}_1 - \overline{Y} + \hat{\beta}_2\overline{X}}_{=0} + \hat{\beta}_2(X_i - \overline{X}) + \hat{\varepsilon}_i$

$Y_i - \overline{Y} = \hat{\beta}_2(X_i - \overline{X}) + \hat{\varepsilon}_i$

$y_i \equiv Y_i - \overline{Y}$, $i = 1,\dots,n$

$x_i \equiv X_i - \overline{X}$, $i = 1,\dots,n$

$y_i = \hat{\beta}_2 x_i + \hat{\varepsilon}_i$

$\mathbf{y} = \hat{\beta}_2\mathbf{x} + \hat{\varepsilon}$, где $\mathbf{y} = \begin{bmatrix}y_1 & \dots & y_n\end{bmatrix}^T$, $\mathbf{x} = \begin{bmatrix}x_1 & \dots & x_n\end{bmatrix}^T$, $\varepsilon = \begin{bmatrix}\varepsilon_1 & \dots & \varepsilon_n\end{bmatrix}^T$

$\mathbf{x}^T\mathbf{y} = \hat{\beta}_2\mathbf{x}^T\mathbf{x} + \underbrace{\mathbf{x}^T\hat{\varepsilon}}_{=0}$

$$\hat{\beta}_2 = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}} \quad (1.7)$$

Аналогично получаем, что в обратной регрессии $X_i = \beta_3 + \beta_4 Y_i + \xi_i$, $i = 1,\dots,n$

$$\hat{\beta}_4 = \frac{\mathbf{y}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \quad (1.8)$$

$ESS = (\hat{Y} - \overline{Y}_i)^T(\hat{Y} - \overline{Y}_i)$

Заметим, что $\hat{Y} - \overline{Y}_i = (I-\pi)(\hat{Y} - \overline{Y}_i)$.

Действительно, $(I-\pi)(P-\pi) = P-\pi$, следовательно,

$\hat{Y} - \overline{Y}_i = (P-\pi)Y = (I-\pi)(P-\pi)Y = (I-\pi)(\hat{Y} - \overline{Y}_i)$.

Далее, $\hat{Y} - \overline{Y}_i = (I-\pi)(\hat{Y} - \overline{Y}_i) = (I-\pi)(\hat{\beta}_1 + \hat{\beta}_2 X - \overline{Y}_i) = \hat{\beta}_2\mathbf{x}$

Значит, $ESS = \hat{\beta}_2^2 \mathbf{x}^T \mathbf{x}$.
Получаем:

$$R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}_2^2 \mathbf{x}^T \mathbf{x}^{(2)}}{\mathbf{y}^T \mathbf{y}} = \frac{\mathbf{x}^T \mathbf{y}^{(2)}}{(\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})} = \mathrm{Corr}^2(X, Y) \tag{1.9}$$

Заметим также, что из формул (1.7), (1.8) и (1.9) следует, что $R^2 = \hat{\beta}_2 \hat{\beta}_4$.
Если $\mathrm{Corr}^2(X,Y) = 1$, то $R^2 = \hat{\beta}_2 \hat{\beta}_4 = 1$.
Отметим также, что из $R^2 = 1$ следует, что $\hat{\varepsilon}_1 = \ldots = \hat{\varepsilon}_n = 0$ и $\hat{\xi}_1 = \ldots = \hat{\xi}_n = 0$.
Тогда $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \underbrace{\hat{\varepsilon}_i}_{=0}$ и $X_i = \hat{\beta}_3 + \hat{\beta}_4 Y_i + \underbrace{\hat{\xi}_i}_{=0}$, $i = 1, \ldots, n$.

$X_i = \hat{\beta}_3 + \hat{\beta}_4 Y_i = (\overline{X} - \hat{\beta}_4 \overline{Y}) + \hat{\beta}_4 Y_i = \left(\overline{X} - \frac{1}{\hat{\beta}_2}\overline{Y}\right) + \frac{1}{\hat{\beta}_2} Y_i$
$\hat{\beta}_2 X_i = (\hat{\beta}_2 \overline{X} - \overline{Y}) + Y_i$
$Y_i = (\overline{Y} - \hat{\beta}_2 \overline{X}) + \hat{\beta}_2 X_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
Следовательно, в случае когда $\mathrm{Corr}^2(X,Y) = 1$, линия парной регрессии Y на X совпадает с линией парной регрессии X на Y.

2.35. Да, если строить регрессию функции от y на функцию от x. А если строить регрессию просто y на x, то оценка наклона будет распределена симметрично около нуля.

2.36. Да, является. Любые, кроме констант. $\mathrm{Var}(\hat{\beta}_{2,IV}) = \sigma^2 \sum (z_i - \bar{z})^2 / \left(\sum (z_i - \bar{z}) x_i\right)^2$.

2.37.

2.38. Вспомните про t, χ^2, F распределения

2.39. $\hat{\lambda} = RSS/(n-2)$ т.к. $\mathrm{Var}(\varepsilon_i) = \lambda$. Оценка $\hat{\beta}_2$ является несмещенной, но $\mathbb{E}(\hat{\beta}_1) = \beta_1 + \lambda$. Можно предложить несмещенную оценку $\hat{\beta}_1' = \hat{\beta}_1 - RSS/(n-2)$.

2.40.

```
df1 <- data.frame(x = c(1,2,3,4), y = c(5,3,3,4) )
df2 <- data.frame(y = rep(df1$y,10), x = rep(df1$x,10))
m1 <- lm(data=df1, y~x)
m2 <- lm(data=df2, y~x)
library(memisc)
mt <- mtable(m1,m2,
  summary.stats=c("N",
    "Deviance","R-squared", "sigma", "F", "p"))
write.mtable(mt, forLaTeX=TRUE)
```

	m1	m2
(Intercept)	4.500	4.500***
	(1.313)	(0.301)
x	−0.300	−0.300**
	(0.480)	(0.110)
N	4	40
Deviance	2.300	23.000
R-squared	0.164	0.164
sigma	1.072	0.778
F	0.391	7.435
p	0.595	0.010

3.1. t-статистики

3.2.

1. Поскольку $\frac{\hat{\sigma}_\varepsilon^2 (n-k)}{\sigma_\varepsilon^2} \sim \chi^2(n-k)$, где $\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n-k}$, $k = 5$. $P(\chi_l^2 < \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} < \chi_u^2) = 0.8$. Преобразовав, получим $P(\frac{\hat{\sigma}_\varepsilon^2(n-5)}{\chi_u^2} < \sigma_\varepsilon^2 < \frac{\hat{\sigma}_\varepsilon^2(n-5)}{\chi_l^2}) = 0.8$, где $\chi_u^2 = \chi_{n-5;0.1}^2$, $\chi_l^2 = \chi_{n-5;0.9}^2$ — соответствующие квантили. По условию $\frac{\hat{\sigma}_\varepsilon^2(n-5)}{\chi_l^2} = A = 45$, $\frac{\hat{\sigma}_\varepsilon^2(n-5)}{\chi_u^2} = B = 87.942$. Поделим B на A, отсюда следует $\frac{\chi_u^2}{\chi_l^2} = 1.95426$. Перебором квантилей в таблице для хи-квадрат распределения мы находим, что $\frac{\chi_{30;0.1}^2}{\chi_{30;0.9}^2} = \frac{40.256}{20.599} = 1.95426$. Значит, $n - 5 = 30$, отсюда следует, что $n = 35$.

2. $\hat{\sigma}^2_{\varepsilon} = 45\frac{\chi^2_u}{n-5} = 45\frac{40.256}{30} = 60.384$

Решение в R:

```
df <- 1:200
a <- qchisq(0.1,df)
b <- qchisq(0.9,df)
c <- b/a
d <- 87.942/45
penalty <- (c-d)^2
df.ans <- df[which(penalty==min(penalty))]
```

Количество степеней свободы $n-5$ должно быть равно `df.ans` $= 30$.

3.3.

Упорядочим нашу выборку таким образом, чтобы наблюдения с номерами с 1 по 35 относились к мужчинам, а наблюдения с номерами с 36 по 58 относились к женщинам. Тогда уравнение

$$\ln W_i = \beta_1 + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Exp_i^2 + \\ \beta_5 Fedu_i + \beta_6 Medu_i + \varepsilon_i, i = 1, ..., 35 \quad (1.10)$$

соответствует регрессии, построенной для подвыборки из мужчин, а уравнение

$$\ln W_i = \gamma_1 + \gamma_2 Edu_i + \gamma_3 Exp_i + \gamma_4 Exp_i^2 + \\ \gamma_5 Fedu_i + \gamma_6 Medu_i + \varepsilon_i, i = 36, ..., 58 \quad (1.11)$$

соответствует регрессии, построенной для подвыборки из женщин. Введем следующие переменные:

$$d_i = \begin{cases} 1, & \text{если } i\text{–ое наблюдение соответствует мужчине,} \\ 0, & \text{в противном случае;} \end{cases}$$

$$dum_i = \begin{cases} 1, & \text{если } i\text{–ое наблюдение соответствует женщине,} \\ 0, & \text{в противном случае.} \end{cases}$$

Рассмотрим следующее уравнение регрессии:

$$\ln W_i = \beta_1 d_i + \gamma_1 dum_i + \beta_2 Edu_i d_i + \gamma_2 Edu_i dum_i + \beta_3 Exp_i d_i + \\ \gamma_3 Exp_i dum_i + \beta_4 Exp_i^2 d_i + \gamma_4 Exp_i^2 dum_i + \beta_5 Fedu_i d_i + \gamma_5 Fedu_i dum_i + \\ \beta_6 Medu_i d_i + \gamma_6 Medu_i dum_i + \varepsilon_i, i = 1, ..., 58 \quad (1.12)$$

Гипотеза, которую требуется проверить в данной задаче, имеет вид

$$H_0 : \begin{cases} \beta_1 = \gamma_1, \\ \beta_2 = \gamma_2, \quad H_1 : |\beta_1 - \gamma_1| + |\beta_2 - \gamma_2| + \dots + |\beta_6 - \gamma_6| > 0. \\ \dots \\ \beta_6 = \gamma_6 \end{cases}$$

Тогда регрессия

$$\ln W_i = \beta_1 d_i + \gamma_1 dum_i + \beta_2 Edu_i d_i + \gamma_2 Edu_i dum_i + \beta_3 Exp_i d_i + \\ \gamma_3 Exp_i dum_i + \beta_4 Exp_i^2 d_i + \gamma_4 Exp_i^2 dum_i + \beta_5 Fedu_i d_i + \\ \gamma_5 Fedu_i dum_i + \beta_6 Medu_i d_i + \gamma_6 Medu_i dum_i + \varepsilon_i, i = 1, \dots, 58 \quad (1.13)$$

по отношению к основной гипотезе H_0 является регрессией без ограничений, а регрессия

$$\ln W_i = \beta_1 + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Exp_i^2 + \beta_5 Fedu_i + \beta_6 Medu_i + \varepsilon_i, i = 1, ..., 58 \quad (1.14)$$

является регрессией с ограничениями.

Кроме того, для решения задачи должен быть известен следующий факт:

$RSS_{UR} = RSS_1 + RSS_2$, где RSS_{UR} — это сумма квадратов остатков в модели:

$$\ln W_i = \beta_1 d_i + \gamma_1 dum_i + \beta_2 Edu_i d_i + \gamma_2 Edu_i dum_i + \beta_3 Exp_i d_i + \gamma_3 Exp_i dum_i + \beta_4 Exp_i^2 d_i + \gamma_4 Exp_i^2 dum_i + \beta_5 Fedu_i d_i + \gamma_5 Fedu_i dum_i + \beta_6 Medu_i d_i + \gamma_6 Medu_i dum_i + \varepsilon_i, i = 1, ..., 58 \quad (1.15)$$

RSS_1 — это сумма квадратов остатков в модели:

$$\ln W_i = \beta_1 + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Exp_i^2 + \beta_5 Fedu_i + \beta_6 Medu_i + \varepsilon_i, i = 1, ..., 35 \quad (1.16)$$

RSS_2 — это сумма квадратов остатков в модели:

$$\ln W_i = \gamma_1 + \gamma_2 Edu_i + \gamma_3 Exp_i + \gamma_4 Exp_i^2 + \gamma_5 Fedu_i + \gamma_6 Medu_i + \varepsilon_i, i = 36, ..., 58 \quad (1.17)$$

1. Тестовая статистика:

$$T = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m)},$$

   где RSS_R – сумма квадратов остатков в модели с ограничениями;

   RSS_{UR} – сумма квадратов остатков в модели без ограничений;

   q – число линейно независимых уравнений в основной гипотезе H_0;

   n – общее число наблюдений;

   m – число коэффициентов в модели без ограничений

2. Распределение тестовой статистики при верной H_0:

$$T \sim F(q, n-m)$$

3. Наблюдаемое значение тестовой статистики:

$$T_{obs} = \frac{(70.3 - (34.4 + 23.4))/6}{(34.4 + 23.4)/(58 - 12)} = 1.66$$

4. Область, в которой H_0 не отвергается:

$$[0; T_{cr}] = [0; 2.3]$$

5. Статистический вывод:

   Поскольку $T_{obs} \in [0; T_{cr}]$, то на основе имеющихся данных мы не можем отвергнуть гипотезу H_0 в пользу альтернативной H_1. Следовательно, имеющиеся данные не противоречат гипотезе об отсутствии дискриминации на рынке труда между мужчинами и женщинами.

3.4.

3.5. Для ответа на вопрос задачи, а именно, можно или нет считать зависимость спроса на молоко от его цены и дохода единой для городской и сельской местностей, воспользуемся гипотезой о нескольких ограничениях. Тогда:

- Ограниченная («короткая») модель, то есть та модель, которая предполагает выполнение нулевой гипотезы, имеет вид :
$$R : y_i = \beta_1 + \beta_2 I_i + \beta_3 P_i + \epsilon_i$$
$$RSS_R = RSS = 8841601$$

- Для того чтобы записать спецификацию неограниченной («длинной») модели, которая пердполагает разные β_i для городской и сельской местностей, введем дополнительную переменную d_i, такую что:
$$d_i = \begin{cases} 1, \text{город}; \\ 0, \text{село} \end{cases}$$
Пусть коэффициенты для городской местности отличаются на некоторое Δ_i, тогда неограниченная модель имеет вид:
$$UR : y_i = \beta_1 + \Delta_1 d_i + (\beta_2 + \Delta_2 d_i) I_i + (\beta_3 + \Delta_3 d_i) P_i + \epsilon_i$$
$$RSS_{UR} = RSS_1 + RSS_2 = 1720236 + 7099423 = 8819659$$

- Гипотезы:
$$H_0 = \begin{cases} \Delta_1 = 0 \\ \Delta_2 = 0 \\ \Delta_3 = 0 \end{cases} \quad H_a : \Delta_1^2 + \Delta_2^2 + \Delta_3^2 > 0$$

- Тестовая статистика имеет вид:
$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m)}$$
где q — число линейно независимых уранений в нулевой гипотезе H_0;
n — общее число наблюдений;
m — число коэффициентов в неограниченной модели

- Распределение тестовой статистики при верной H_0:
$$F_{\text{кр}} \sim (F_{\alpha,q,n-m})$$

- Расчётное значение тестовой статистики $F_{\text{рас}} = 17.58$, $F_{\text{кр}} \approx 2.61$

- Так как $F_{\text{рас}} > F_{\text{кр}}$ следовательно гипотеза H_0 — отвергается.

Вывод: зависимость спроса на молоко от его цены и дохода для городской и сельской местностей нельзя считать единой.

3.6.

3.7. Задача решается аналогично предыдущем задачам, к примеру, 3.3, 3.5.

Главное отличие заключается в том, что вместо значений RSS_R и RSS_{UR} даются значения соответствующих R^2, также следует вспомнить, что $\sum_{i=1}^{n=52}(Price_i - \overline{Price})^2 = 278$ ни что иное, как TSS, которое, в свою очередь, не зависит от спецификации модели, то есть $TSS_R = TSS_{UR} = TSS$. Тогда можно выразитьRSS моделей:

$$\begin{cases} R^2 = \frac{ESS}{TSS} \\ TSS = ESS + RSS \end{cases} \to \begin{cases} RSS_R = TSS(1 - R_R^2) = 278(1 - 0.78) \approx 61.16 \\ RSS_{UR} = TSS(1 - R_{UR}^2) = 278(1 - 0.85) \approx 41.7 \end{cases}$$

Находим расчётное значение F-статистики

$$F_{\text{рас}} = \frac{(61.16 - 41.7)/5}{41.7/(52-10)} \approx 3.92$$

Находим критическое значение F-статистики

$$F_{\text{кр}} \sim F_{0.05,5,42} \approx 2.44$$

Получаем, что $F_{\text{рас}} > F_{\text{кр}}$, и, следовательно, H_0 отвергается в пользу альтернативной гипотезы на уровне значимости 5%.

Вывод: гипотеза об одинаковом ценообразовании квартир на севере и на юге отвергается на уровне значимости 5%.

3.8.

3.9. Спецификация модели :

$$\widehat{\ln Q} = \hat{\beta}_1 + \hat{\beta}_2 \ln P + \hat{\beta}_3(SPRING + SUMMER) + \hat{\beta}_5 FALL$$

Интерпретация: осень так же влияет на логарифм величины спросы, как и весна. Задача решается аналогично задачам 3.7, 3.5

$$\begin{cases} R^2 = \frac{ESS}{TSS} \\ TSS & = ESS + RSS \\ TSS_R = TSS_{UR} = TSS \end{cases}$$

Находим расчётное и наблюдаемое значение F-статистики

$$\begin{cases} F_{\text{рас}} = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/(n-m)} \approx 3.3 \\ F_{\text{кр}} = F_{0.05,1,15} \approx 4.54 \end{cases}$$

Следовательно, $F_{\text{рас}} < F_{\text{кр}}$ и H_0 не отвергается на уровне значимости 5%.

Вывод:гипотеза H_0 о равном влиянии осени и весны на логарифм спроса не отвергается на уровне значимости 5% .

3.10. Смысл гипотезы: летом и осенью одинаковая зависимость и одинаковая зависимость зимой и весной. Ограниченная модель: $\widehat{\ln Q} = \hat{\beta}_1 + \hat{\beta}_2 \cdot \ln P + \hat{\beta}_3 d$, где d равна 1 для лета и осени. Наблюдаемое значение статистики $F_{obs} = 1.375$, критическое, $F_{cr} = 3.5218933$. Гипотеза не отвергается.

3.11.

3.12.

3.13. $\hat{\beta}_1 = 1.3870 + 2.6259 = 4.0129$, $\hat{\beta}_2 = 5.2587 + 2.5955 = 7.8542$

3.14. $y_i = \beta_1 + \beta_2(x_{i1} + x_{i2} + x_{i3}) + \epsilon_i$

3.15. $y_i = \beta_1 + \beta_2(x_{i1} + x_{i2} + x_{i3}) + \epsilon_i$

3.16.

3.17. 1,2

3.18.

3.19. значим

3.20. не значим

3.21. $\alpha > 0.09$

3.22.

3.23. Из формул

$$\begin{cases} R^2 = \frac{ESS}{TSS} \\ TSS = ESS + RSS \end{cases}$$

получаем $R^2 = \frac{170.4}{(170.4+80.3)} \approx = 0.68$

Тестируемые гипотезы:

$$H_0 = \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \\ \beta_4 = 0 \end{cases} \quad H_a : \beta_2^2 + \beta_3^2 + \beta_4^2 > 0$$

Так как по условию задачи проверяем значимость модели в целом, следовательно ограниченная модель — регрессия на константу, таким образом:

$$\begin{cases} \widehat{y_i} = \overline{y} \\ RSS_R = \sum_{i=1}^n (y_i - \widehat{y_i})^2 = \sum_{i=1}^n (y_i - \overline{y_i})^2 = TSS \\ RSS_{UR} = TSS(1 - R_{UR}^2) \\ TSS_{UR} = TSS_R = TSS \end{cases}$$

Получаем, $F_{\text{рас}} = \frac{R_{UR}^2/q}{(1-R_{UR}^2)/(n-m)}$

Значения статистик:

$$\begin{cases} F_{\text{рас}} \approx 12.04 \\ F_{\text{кр}} = F_(0.01, 3, 17) \approx 5.185 \end{cases}$$

Отсюда, $F_{\text{рас}} > F_{\text{кр}}$, и H_0 отвергается на уровне значимости 1%.

Вывод: гипотеза H_0 отвергается на уровне значимости 1%, следовательно модель «в целом» значима.

3.24.

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):

$$\ln W_i = \beta + \beta_{Edu} Edu_i + \beta_{Age} Age_i + \beta_{Age^2} Age_i^2 + \beta_{Fedu} Fedu_i + \beta_{Medu} Medu_i + \varepsilon_i \quad (1.18)$$

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, $RSS_{UR} = TSS - ESS_{UR} = 150.7 - 110.3 = 40.4$

1. Спецификация:

$$\ln W_i = \beta + \beta_{Edu} Edu_i + \beta_{Age} Age_i + \beta_{Age^2} Age_i^2 + \beta_{Fedu} Fedu_i + \beta_{Medu} Medu_i + \varepsilon_i \quad (1.19)$$

2. Проверка гипотезы

   (a) $H_0 : \begin{cases} \beta_{Fedu} = 0 \\ \beta_{Medu} = 0 \end{cases} \quad H_a : |\beta_{Fedu}| + |\beta_{Medu}| > 0$

   (b) $T = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$, где $q = 2$ — число линейно независимых уравнений в основной гипотезе H_0, $n = 25$ — число наблюдений, $k = 6$ — число коэффициентов в модели без ограничения

   (c) $T \sim F(q; n-k)$

   (d) $T_{obs} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} = \frac{(60.4-40.4)/2}{40.4/(25-6)} = 4.70$

   (e) Нижняя граница = 0, верхняя граница = 3.52

   (f) 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.

3.25.

$$\widehat{Price} = \hat{\beta}_1 + \hat{\beta}_2 Hsize + 20\hat{\beta}_4 Lsize + \hat{\beta}_4 Bath + \hat{\beta}_5 BDR$$

Размер участка в 20 раз сильнее влияет на цену дома, чем число ванных комнат.

$$\begin{cases} R^2 = \frac{ESS}{TSS} \\ TSS = ESS + RSS \\ TSS_R = TSS_{UR} = TSS \end{cases}$$

$$\begin{cases} RSS_R = TSS(1 - R^2_R) \\ RSS_{UR} = TSS(1 - R^2_{UR}) \end{cases} \quad \rightarrow$$

$$\begin{cases} F_{\text{рас}} = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/(n-m)} = \frac{(0.218-0.136)/1}{(1-0.218)/15} \approx 1.573 \\ F_{\text{кр}} = F_{0.05,1,15} \approx 4.54 \end{cases}$$

$F_{\text{рас}} < F_{\text{кр}}$ и, следовательно, H_0 не отвергается на уровне значимости 5%.

Вывод: гипотеза H_0 о том, что размер участка в 20 раз сильнее влияет на цену дома, чем число ванных комнат, не отвергается на уровне значимости 5%.

3.26.

3.27. $H_0 : \beta_2 = \beta_3$ — труд и капитал вносят одинаковый вклад в выпуск фирмы.

$$\begin{cases} F_{\text{рас}} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m)} = \frac{(0.894-0.851)/1}{0.851/(27-3)} \approx 1.213 \\ F_{\text{кр}} = F_{0.05,1,24} \approx 4.26 \end{cases}$$

Получаем, что $F_{\text{рас}} < F_{\text{кр}}$, и, следовательно, H_0 не отвергается на уровне значимости 5%

Вывод: гипотеза H_0, предполагающая, что труд и капитал вносят одинаковый вклад в выпуск фирмы, не отвергается на уровне значимости 5%.

3.28.

3.29.

1.
$$\begin{cases} H_0 : \beta_2 = 0 \\ H_a : \beta_2 \neq 0 \end{cases}$$

2. (a) $t = \frac{\hat{\beta}_i - \beta_i}{se(\beta_i)}$
   (b) $t_{\alpha,n-m} = t_{0.05,47}$
   (c) $t = \frac{0.08-0}{0.0093} \approx 8.67$
   (d) $[-t_{\text{кр}}, t_{\text{кр}}]$
   (e) гипотеза H_0 отвергается, так как P-значение равно нулю; можно честно посчитать $t_{\text{кр}} = t_{0.05,47}$ или вспомнить, что при количестве наблюдений больше 30, t-распределение похоже на нормальное, для которого квантиль на уровне 5% примерно равна 1.67 и $F_{\text{наб}} > F_{\text{кр}}$. Гипотеза H_0 отвергается, следовательно коэффициент β_2 значим на уровне значимости 10%.

3. (a) $t = \frac{\hat{\beta}_i - \beta_i}{se(\beta_i)}$
   (b) $t_{\alpha,n-m} = t_{0.05,47}$
   (c) $t = \frac{-287-1}{64.92} \approx -4.42$
   (d) $(-\infty, t_{\text{кр}}]$
   (e) гипотеза H_0 отвергается, так как P-значение равно нулю; аналогично 2(e) $t_{\text{кр}} = t_{0.05,47} \approx 1.67$ и $F_{\text{наб}} > F_{\text{кр}}$. Гипотеза H_0 отвергается, следовательно коэффициент β_1 значим на уровне значимости 5%.

4.
$$H_0 = \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \\ \beta_4 = 0 \end{cases} \quad H_a : \beta_2^2 + \beta_3^2 + \beta_4^2 > 0$$

5. (a) $F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m)}$

   (b) $F_{\alpha,q,n-m} = F_{0.01,3,47}$

   (c) $F = 34.81$

   (d) $[0, F_{\text{кр}}]$

   (e) гипотеза H_0 отвергается, так как P-value $\approx 0(5.337e^{-12})$; можно вычислить $F_{\text{кр}} = F_{0.01,3,47} \approx 4.23$. Следовательно, $F_{\text{наб}} > F_{\text{кр}}$ и H_0 отвергается, и регрессия «в целом» значима на уровне значимости 1%.

6. (a) $F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m)}$

   (b) $F_{\alpha,q,n-m} = F_{0.05,3,47}$

   (c) $F \approx 9.525$

   (d) $[0, F_{\text{кр}}]$

   (e) гипотеза H_0 отвергается, так как $F_{\text{кр}} = F_{0.05,3,47} \approx 4.047$ и $F_{\text{наб}} > F_{\text{кр}}$, следовательно коэффициент β_4 значим на уровне значимости 5%.

3.30.

3.31. $0.25\hat{\beta}_1 + 0.75\hat{\beta}'_1$, $0.25\hat{\beta}_2 + 0.75\hat{\beta}'_2$ и $0.25\hat{\beta}_3 + 0.75\hat{\beta}'_3$

3.32. Сами оценки коэффициентов никак детерминистически не связаны, но при большом размере подвыборок примерно равны. А дисперсии связаны соотношением $\mathrm{Var}(\hat{\beta}_a)^{-1} + \mathrm{Var}(\hat{\beta}_b)^{-1} = \mathrm{Var}(\hat{\beta}_{tot})^{-1}$

3.33.

3.34.

3.35.

3.36.

3.37.

3.38.

Из оценки ковариационной матрицы находим, что $se(\hat{\beta}_{totsp} = \hat{\beta}_{livesp}) = 0.269606$.

Исходя из $Z_{crit} = 1.96$ получаем доверительный интервал, $[-0.822083; 0.2347725]$.

Вывод: при уровне значимости 5% гипотеза о равенстве коэффициентов не отвергается.

3.39.

3.40.

3.41.

1. $\mathbb{P}(\hat{\beta}_3 > se(\hat{\beta}_3)) = \mathbb{P}(t_{17} > 1) = 0.1656664$

2. $\mathbb{P}(\hat{\beta}_3 > \sigma_{\hat{\beta}_3}) = \mathbb{P}(N(0,1) > 1) = 0.1586553$

3.42. В обоих случаях можно так подобрать коэффициенты $\hat{\beta}$, что $kr_i = \widehat{kr}_i$. А именно, идеальные прогнозы достигаются при $\hat{\beta}_{p1} = 1$, $\hat{\beta}_{p2} = 1$, $\hat{\beta}_{p3} = 1$, $\hat{\beta}_{p4} = 1$, $\hat{\beta}_{p5} = 1$ и (в первой модели) $\hat{\beta}_1 = 0$. Отсюда $RSS = 0$, $ESS = TSS$, поэтому $R^2 = 1$ даже в модели без свободного члена. Получаем $\hat{\sigma}^2 = 0$, поэтому строго говоря t статистики и P-значения не существуют из-за деления на ноль.

На практике при численной минимизации RSS оказывается, что t-статистики коэффициентов при задачах принимают очень большие значения, а соответствующие P-значения крайне близки к нулю. В первой модели особенной на практике будет t статистика свободного члена. В силу неопределенности вида $0/0$ свободный коэффициент на практике может оказаться незначим.

3.43.

3.44. $\hat{\beta}_2 = 0.41$, $\hat{\beta}_3 = 0.3$, $\hat{\beta}_4 = -0.235$, переменная x значима

3.45. $\hat{\beta}_2 = 0.75$, $\hat{\beta}_3 = 0.625$, $\hat{\beta}_4 = 0.845$, переменные z и w значимы

3.46. $RSS_1 > RSS_2 = RSS_3$, в моделях два и три, ошибка прогноза равна $\hat{\beta}_4$

3.47.

3.48. $RSS/\sigma^2 \sim \chi^2_{n-k}$, $\mathbb{E}(RSS) = (n-k)\sigma^2$, $\mathrm{Var}(RSS) = 2(n-k)\sigma^4$, $\mathbb{P}(5\sigma^2 < RSS < 10\sigma^2) \approx 0.451$

3.49. $\mathbb{P}(\hat{\sigma}_3^2 > \hat{\sigma}_1^2) = 0.5$, $\mathbb{P}(\hat{\sigma}_1^2 > 2\hat{\sigma}_2^2) = 0.5044$, $\mathbb{E}(\hat{\sigma}_2^2/\hat{\sigma}_1^2) = 1.25$, $\mathrm{Var}(\hat{\sigma}_2^2/\hat{\sigma}_1^2) = 4.6875$

3.50. 90% во всех пунктах

3.51. Поскольку $\hat{\mu}$, $\hat{\nu}$, $\hat{\gamma}$ и $\hat{\delta}$ являются МНК-коэффициентами в регрессии $y_i = \mu + \nu x_i + \gamma d_i + \delta x_i d_i + \varepsilon_i$, $i = 1, \ldots, n$, то для любых μ, ν, γ и δ имеет место

$$\sum_{i=1}^{n} (y_i - \hat{\mu} - \hat{\nu} x_i - \hat{\gamma} d_i - \hat{\delta} x_i d_i - \varepsilon_i)^2 \leqslant \sum_{i=1}^{n} (y_i - \mu - \nu x_i - \gamma d_i - \delta x_i d_i - \varepsilon_i)^2 \quad (1.20)$$

Перепишем неравенство (1.20) в виде

$$\sum_{i=1}^{m} (y_i - (\hat{\mu} + \hat{\gamma}) - (\hat{\nu} + \hat{\delta}) x_i)^2 + \sum_{i=m+1}^{n} (y_i - \hat{\mu} - \hat{\nu} x_i)^2 \leqslant \sum_{i=1}^{m} (y_i - (\mu + \gamma) - (\nu + \delta) x_i)^2 + \sum_{i=m+1}^{n} (y_i - \mu - \nu x_i)^2 \quad (1.21)$$

Учитывая, что неравенство (1.21) справедливо для всех μ, ν, γ и δ, то оно останется верным для $\mu = \hat{\mu}$, $\nu = \hat{\nu}$ и произвольных γ и δ. Имеем

$$\sum_{i=1}^{m} (y_i - (\hat{\mu} + \hat{\gamma}) - (\hat{\nu} + \hat{\delta}) x_i)^2 + \sum_{i=m+1}^{n} (y_i - \hat{\mu} - \hat{\nu} x_i)^2 \leqslant \sum_{i=1}^{m} (y_i - (\hat{\mu} + \gamma) - (\hat{\nu} + \delta) x_i)^2 + \sum_{i=m+1}^{n} (y_i - \hat{\mu} - \hat{\nu} x_i)^2 \quad (1.22)$$

Следовательно

$$\sum_{i=1}^{m} (y_i - (\hat{\mu} + \hat{\gamma}) - (\hat{\nu} + \hat{\delta}) x_i)^2 \leqslant \sum_{i=1}^{m} (y_i - (\hat{\mu} + \gamma) - (\hat{\nu} + \delta) x_i)^2 \quad (1.23)$$

Обозначим $\tilde{\beta}_1 := \hat{m u} + \gamma$ и $\tilde{\beta}_2 := \hat{\nu} + \delta$. В силу произвольности γ и δ коэффициенты $\tilde{\beta}_1$ и $\tilde{\beta}_2$ также произвольны. тогда для любых $\tilde{\beta}_1$ и $\tilde{\beta}_2$ выполнено неравенство:

$$\sum_{i=1}^{m} (y_i - (\hat{\mu} + \hat{\gamma}) - (\hat{\nu} + \hat{\delta}) x_i)^2 \leqslant \sum_{i=1}^{m} (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2$$

которое означает, что $\hat{\mu}+\hat{\gamma}$ и $\hat{\nu}+\hat{\delta}$ являются МНК-оценками коэффициентов β_1 и β_2 в регрессии $y_i=\beta_1+\beta_2 x_i+\varepsilon_i$, оцененной по наблюдениям $i=1,\ldots,m$, то есть $\hat{\beta}_1=\hat{\mu}+\hat{\gamma}$ и $\hat{\beta}_2=\hat{\nu}+\hat{\delta}$.

3.52. не верно, поскольку R^2_{adj} может принимать отрицательные значения, а $F(n-k,n-1)$ — не может.

3.53. Сгенерировать сильно коррелированные x и z

3.54.

3.55.

3.56.

3.57. bootstrap, дельта-метод.

3.58.

3.59.

3.60. При наличии ошибок в измерении зависимой переменной оценки остаются несмещенными, их дисперсия растет. Однако оценка дисперсии может случайно оказаться меньше. Например, могло случиться, что ошибки u_i случайно компенсировали ε_i.

3.61.

3.62.

3.63. 0

3.64. Проводим тест Чоу

3.65. Несмещенной является $\hat{\sigma}^2$, поэтому $\hat{\sigma}$ смещена, но состоятельна.

4.1.

1. В случае нестохастических регрессоров:

   Пусть дана регрессионная модель $y=X\beta+\epsilon$ с k регрессорами, включая свободный член и n наблюдениями, тогда если

   (a) регрессионная модель правильно специфицирована
   (b) $\mathrm{rang}(X)=k$
   (c) X не являются стохастическими
   (d) $\mathbb{E}(\epsilon)=0$
   (e) $\mathrm{Var}(\epsilon)=\sigma^2 I$

   то $\hat{\beta}=(X'X)^{-1}X'y$ являются лучшими оценками в классе линейных несмещённых оценок, то есть BLUE-оценками.

   В случае стохастических регрессоров:

   Пусть дана регрессионная модель $y=X\beta+\epsilon$ с k регрессорами, включая свободный член и n наблюдениями, тогда если

   (a) регрессионная модель правильно специфицирована
   (b) $\mathrm{rang}(X)=k$
   (c) $\mathbb{E}(\epsilon|X)=0$
   (d) $\mathrm{Var}(\epsilon|X)=\sigma^2 I$

   то $\hat{\beta}=(X'X)^{-1}X'y$ являются лучшими оценками в классе линейных несмещённых оценок, то есть BLUE-оценками.

2. Да, верно. В самом деле,

$$\mathbb{E}(\hat{\beta})=\mathbb{E}((X'X)^{-1}X'y)=(X'X)^{-1}X'\mathbb{E}(y)=(X'X)^{-1}X'\mathbb{E}(X\beta+\epsilon)=$$
$$=(X'X)^{-1}X'X\mathbb{E}(\beta)+(X'X)^{-1}X'\underbrace{\mathbb{E}(\epsilon)}_{=0}=\beta$$

3.

$$\operatorname{Var}(\hat{\beta}) = \operatorname{Var}((X'X)^{-1}X'y) = (X'X)^{-1}X' \operatorname{Var}(y)((X'X)^{-1}X')' = \\ = (X'X)^{-1}X' \operatorname{Var}(X\beta + \epsilon)X(X'X)^{-1}$$

Так как β является константой, то $\operatorname{Var}(X\beta) = 0$. Тогда

$$\operatorname{Var}(\hat{\beta}) = (X'X)^{-1}X' \operatorname{Var}(X\beta + \epsilon)X(X'X)^{-1} = \\ = (X'X)^{-1}X' \operatorname{Var}(\epsilon)X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} = \\ = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

4.2. Да, в общем случае (кроме случая $\beta = 0$ это верно. Так как $\tilde{\beta}$ является несмещённой, то $\mathbb{E}(\tilde{\beta}) = \beta$.

$$\mathbb{E}(\tilde{\beta}) = \mathbb{E}(((X'X)^{-1}X' + A)y) = \mathbb{E}[((X'X)^{-1}X' + A)(X\beta + \epsilon)] = \\ = \mathbb{E}(((X'X)^{-1}X' + A)X\beta) + ((X'X)^{-1}X' + A)\underbrace{\mathbb{E}(\epsilon)}_{=0} = \\ = \mathbb{E}((X'X)^{-1}X'X\beta + AX\beta) = \beta + AX\beta$$

$$\mathbb{E}(\tilde{\beta}) = \beta$$

$$\beta + AX\beta = \beta$$

$$AX\beta = 0$$

Значит, либо $AX = 0$, либо $\beta = 0$.

Заметим, что при $\beta = 0$ при любом AX оценка $\tilde{\beta}$ будет несмещённой.

4.3.

$$X'X = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 1.5 \end{pmatrix}$$

$$\operatorname{Var}(\hat{\beta}_1) = \operatorname{Var}(\hat{\beta})_{[1,1]} = \sigma^2$$

$$\operatorname{Var}(\hat{\beta}_2) = \operatorname{Var}(\hat{\beta})_{[2,2]} = 1.5\sigma^2$$

$$\operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \operatorname{Var}(\hat{\beta})_{[1,2]} = -\sigma^2$$

$$\operatorname{Corr}(\hat{\beta}_1, \hat{\beta}_2) = \frac{\operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{\sqrt{\operatorname{Var}(\hat{\beta}_1)\operatorname{Var}(\hat{\beta}_2)}} = \frac{\sigma^2}{\sigma \cdot \sqrt{1.5}\sigma} = \sqrt{\frac{2}{3}} = \frac{\sqrt{6}}{3}$$

Показательно, что значения y здесь не используются.

Ответ: $\frac{\sqrt{6}}{3}$.

4.4.

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$X'X\hat{\beta} = X'y$$

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y$$

поэтому

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y = (D'X'XD)^{-1}D'X'X\hat{\beta} = (D'X'XD)^{-1}D'X'XDD^{-1}\hat{\beta} = D^{-1}\hat{\beta}$$

$$D^{-1} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{cases} \hat{\alpha}_1 = \beta_1 - \beta_2 + \beta_3 \\ \hat{\alpha}_2 = \beta_2 - \beta_3 \\ \hat{\alpha}_3 = \beta_3 \end{cases}$$

4.5.

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$X'X\hat{\beta} = X'y$$

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y$$

поэтому

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y = (D'X'XD)^{-1}D'X'X\hat{\beta} = (D'X'XD)^{-1}D'X'XDD^{-1}\hat{\beta} = D^{-1}\hat{\beta}$$

$$D^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{cases} \hat{\alpha}_1 = \beta_1 - \beta_2 \\ \hat{\alpha}_2 = \beta_2 - \beta_3 \\ \hat{\alpha}_3 = \beta_3 \end{cases}$$

4.6.

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$X'X\hat{\beta} = X'y$$

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y$$

поэтому

$$\hat{\alpha} = (D'X'XD)^{-1}D'X'y = (D'X'XD)^{-1}D'X'X\hat{\beta} = (D'X'XD)^{-1}D'X'XDD^{-1}\hat{\beta} = D^{-1}\hat{\beta}$$

$$D^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{cases} \hat{\alpha}_1 = \beta_1 \\ \hat{\alpha}_2 = \beta_2 - \beta_3 \\ \hat{\alpha}_3 = \beta_3 \end{cases}$$

4.7. Да, верно.

$$\begin{aligned} \hat{\varepsilon}'\hat{y} &= (y - \hat{y})'\hat{y} = (y - X\hat{\beta})'X\hat{\beta} = (y - X(X'X)^{-1}X'y)'X(X'X)^{-1}X'y = \\ &= ((I - X(X'X)^{-1}X')y)'X(X'X)^{-1}X'y = y'(I - X(X'X)^{-1}X')X(X'X)^{-1}X'y = \\ &= y'(X(X'X)^{-1}X' - X(X'X)^{-1}X'X(X'X)^{-1}X')y = y'(X(X'X)^{-1}X' - X(X'X)^{-1}X')y = 0 \end{aligned}$$

Да, верно.

$$\hat{y}'\hat{\varepsilon} = (\hat{\varepsilon}'\hat{y})' = 0$$

так как выше доказано, что $he'\hat{y} = 0$.

4.8.

1.

$$\hat{\gamma} = (Z'Z)^{-1}Z'y = A^{-1}(X'X)^{-1}(A')^{-1}A'X'y = \\ A^{-1}(X'X)^{-1}X'y = A^{-1}\hat{\beta} \quad (1.24)$$

2. $\hat{u} = y - Z\hat{\gamma} = y - XAA^{-1}\hat{\beta} = y - X\hat{\beta} = \hat{\varepsilon}$

3. Пусть $z^0 = \begin{pmatrix} 1 & z_1^0 & \ldots & z_{k-1}^0 \end{pmatrix}$ — вектор размера $1 \times k$ и $x^0 = \begin{pmatrix} 1 & x_1^0 & \ldots & x_{k-1}^0 \end{pmatrix}$ — вектор размера $1 \times k$. Оба эти вектора представляют собой значения факторов. Тогда $z^0 = x^0 A$ и прогнозное значение для регрессии с преобразованными факторами равно $z^0\hat{\gamma} = x^0 AA^{-1}\hat{\beta} = x^0\hat{\beta}$ прогнозному значению для регрессии с исходными факторами.

4.9.

1.

$$\mathbb{E}(\tilde{\beta}) = ((X'X)^{-1} + \gamma I)X'\mathbb{E}(y) = \\ ((X'X)^{-1} + \gamma I)X'X\beta = \beta + \gamma X'X\beta \quad (1.25)$$

2.

$$\begin{aligned} \operatorname{Var}(\tilde{\beta}) = \operatorname{Var}(((X'X)^{-1} + \gamma I)X'y) = \\ \operatorname{Var}(((X'X)^{-1} + \gamma I)X'\varepsilon) = \\ (((X'X)^{-1} + \gamma I)X') \operatorname{Var}(\varepsilon)(((X'X)^{-1} + \gamma I)X')' = \\ (((X'X)^{-1} + \gamma I)X')\sigma_\varepsilon^2 I(((X'X)^{-1} + \gamma I)X')' = \\ \sigma_\varepsilon^2((X'X)^{-1} + \gamma I)X'X((X'X)^{-1} + \gamma I) = \\ \sigma_\varepsilon^2((X'X)^{-1} + \gamma I)(I + \gamma X'X) = \\ \sigma_\varepsilon^2((X'X)^{-1} + 2\gamma I + \gamma^2 X'X) \quad (1.26) \end{aligned}$$

4.10. Да, верно.

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

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2 = y'\left(I - \frac{\overrightarrow{1}'}{\overrightarrow{1}^2}\right)y$$

не зависит от X.

$$RSS_a = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = y'(I - X(X'X)^{-1}X')y$$

$$RSS_b = y'(I - Z(Z'Z)^{-1}Z')y = y'(I - XD(D'X'XD)^{-1}D'X')y =$$

так как D является квадратной и невырожденной, то используя формулу $(AB)^{-1} = B^{-1}A^{-1}$, получим:

$$RSS_b = y'(I - XD(D'X'XD)^{-1}D'X')y = y'(I - XDD^{-1}(X'X)^{-1}D'^{-1}D'X')y = \\ = y'(I - X(X'X)^{-1}X')y = RSS_a$$

Значит,

$$R_a^2 = 1 - \frac{RSS_a}{TSS_a} = 1 - \frac{RSS_b}{TSS_b} = R_b^2$$

4.11. Да, верно.

$$RSS_1 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = y'(I - X(X'X)^{-1}X')y$$

$$RSS_2 = y'(I - Z(Z'Z)^{-1}Z')y = y'(I - XD(D'X'XD)^{-1}D'X')y =$$

так как D является квадратной и невырожденной, то используя формулу $(AB)^{-1} = B^{-1}A^{-1}$, получим:

$$RSS_2 = y'(I - XD(D'X'XD)^{-1}D'X')y = y'(I - XDD^{-1}(X'X)^{-1}D'^{-1}D'X')y =$$
$$= y'(I - X(X'X)^{-1}X')y = RSS_1$$

4.12.
4.13.

1. $n = 5$
2. $k = 3$
3. $TSS = 10$
4. $RSS = 2$
5. $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$
6. $R^2 = 1 - \frac{RSS}{TSS} = 0.8$. R^2 высокий, построенная эконометрическая модель «хорошо» описывает данные
7. Основная гипотеза — $H_0 : \beta_2 = 0$, альтернативная гипотеза — $H_a : \beta_2 \neq 0$
8. Проверка гипотезы
   (a) $T = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}; n = 5; k = 3$
   (b) $T \sim t(n-k); n = 5; k = 3$
   (c) $T_{obs} = \frac{\hat{\beta}_2 - 0}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - 0}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-0}{\sqrt{\frac{2}{5-3}1.3333}} = 1.7321$
   (d) Нижняя граница $= -2.920$, верхняя граница $= 2.920$
   (e) 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%
9. $p - value(T_{obs}) = \mathbb{P}(|T| > |T_{obs}|) = 2F_T(|T_{obs}|)$, где $F_T(|T_{obs}|)$ — функция распределения $t-$распределения с $n - k = 5 - 3 = 2$ степенями свободы в точке $|T_{obs}|$. $p - value(T_{obs}) = 2tcdf(-|T_{obs}|, n-k) = 2tcdf(-1.7321, 2) = 0.2253$. Поскольку $P-$значение превосходит уровень значимости 10%, то основная гипотеза — $H_0 : \beta_2 = 0$ не может быть отвергнута
10. Проверка гипотезы
   (a) $T = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}; n = 5; k = 3$
   (b) $T \sim t(n-k); n = 5; k = 3$
   (c) $T_{obs} = \frac{\hat{\beta}_2 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$
   (d) Нижняя граница $= -2.920$, верхняя граница $= 2.920$
   (e) 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%
11. Проверка гипотезы
   (a) $T = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}; n = 5; k = 3$

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

(c) $T_{obs} = \frac{\hat{\beta}_2 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$

(d) Нижняя граница $= -\infty$, верхняя граница $= 1.8856$

(e) 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%

12. Проверка гипотезы

(a) $T = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}}; n=5; k=3$

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

(c) $T_{obs} = \frac{\hat{\beta}_2 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - 1}{\sqrt{\frac{RSS}{n-k}[(X'X)^{-1}]_{22}}} = \frac{2-1}{\sqrt{\frac{2}{5-3}1.3333}} = 0.8660$

(d) Нижняя граница $= -1.8856$, верхняя граница $= +\infty$

(e) 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%

13. Основная гипотеза — $H_0 : \beta_2 = \beta_3 = 0$, альтернативная гипотеза — $H_a : |\beta_2| + |\beta_3| > 0$

14. Проверка гипотезы

(a) $T = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k}; n=5; k=3$

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

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

(d) Нижняя граница $= 0$, верхняя граница $= 19$

(e) 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

15. $p - value(T_{obs}) = \mathbb{P}(|T| > |T_{obs}|) = 2F_T(|T_{obs}|)$, где $F_T(|T_{obs}|)$ — функция распределения $F-$распределения с $k=3$ и $n-k = 5-3 = 2$ степенями свободы в точке T_{obs}. $p - value(T_{obs}) = 1 - fcdf(-|T_{obs}|, n-k) = 1 - fcdf(4, 2) = 0.2$. Поскольку $P-$значение превосходит уровень значимости 10%, то основная гипотеза — $H_0 : \beta_2 = \beta_3 = 0$ не может быть отвергнута. Таким образом, регрессия «в целом» незначима

16. Проверка гипотезы

(a) $T = \frac{\hat{\beta}_2 + \hat{\beta}_3 - (\beta_2 + \beta_3)}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3)}}$, где $\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3) = \widehat{\mathrm{Var}}(\hat{\beta}_2) + \widehat{\mathrm{Var}}(\hat{\beta}_3) + 2\widehat{\mathrm{Cov}}(\hat{\beta}_2; \hat{\beta}_3) = \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{23} + \hat{\sigma}^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$

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

(c) $\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 2}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$

(d) Нижняя граница $= -4.3027$, верхняя граница $= 4.3027$

(e) 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%

17. Проверка гипотезы

(a) $T = \frac{\hat{\beta}_2 + \hat{\beta}_3 - (\beta_2 + \beta_3)}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3)}}$, где $\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3) = \widehat{\mathrm{Var}}(\hat{\beta}_2) + \widehat{\mathrm{Var}}(\hat{\beta}_3) + 2\widehat{\mathrm{Cov}}(\hat{\beta}_2; \hat{\beta}_3) = \hat{\sigma}^2[(X'X)^{-1}]_{22} + 2\hat{\sigma}^2[(X'X)^{-1}]_{23} + \hat{\sigma}^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$

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

(c) $\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat\beta_2+\hat\beta_3-2}{\sqrt{\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$

(d) Нижняя граница $=-\infty$, верхняя граница $=2.9200$

(e) 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%

18. Проверка гипотезы

(a) $T = \frac{\hat\beta_2+\hat\beta_3-(\beta_2+\beta_3)}{\sqrt{\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3)}}$, где $\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3) = \widehat{\mathrm{Var}}(\hat\beta_2) + \widehat{\mathrm{Var}}(\hat\beta_3) + 2\widehat{\mathrm{Cov}}(\hat\beta_2;\hat\beta_3) = \hat\sigma^2[(X'X)^{-1}]_{22} + 2\hat\sigma^2[(X'X)^{-1}]_{23} + \hat\sigma^2[(X'X)^{-1}]_{33} = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33})$

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

(c) $\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3) = \frac{RSS}{n-k}([(X'X)^{-1}]_{22} + 2[(X'X)^{-1}]_{23} + [(X'X)^{-1}]_{33}) = \frac{2}{5-3}(1.3333 + 2(-1.0000) + 2.0000) = 1.3333$. Тогда $T_{obs} = \frac{\hat\beta_2+\hat\beta_3-2}{\sqrt{\widehat{\mathrm{Var}}(\hat\beta_2+\hat\beta_3)}} = \frac{2+1-2}{\sqrt{1.3333}} = 0.8660$

(d) Нижняя граница $=-2.9200$, верхняя граница $=+\infty$

(e) 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%

4.14.

4.15.

4.16. $\mathrm{Var}(\hat\beta) = \sigma^2(X'X)^{-1}$

4.17. $(n-1)\sigma^2$, $(n-k)\sigma^2$

4.18. $TSS = y'(I-\pi)y$, $RSS = y'(I-P)y$, $ESS = y'(P-\pi)y$

4.19. $\mathbb{E}(TSS) = (n-1)\sigma^2 + \beta'X'(I-\pi)X\beta$

4.20. $(n-1)\sigma^2$, $(n-k)\sigma^2$, $(k-1)\sigma^2$

4.21.

4.22. $\mathbb{E}(\varepsilon)=0$, $\mathbb{E}(\hat\varepsilon)=0$, $\sum\varepsilon_i$ может оказаться равной нулю только случайно, в нормальной модели это происходит с вероятностью 0, $\sum\hat\varepsilon_i = 0$ в модели со свободным членом

4.23. $\sum y_i^2 = \sum \hat y_i^2 + \sum \hat\varepsilon_i^2$, $TSS = ESS + RSS$,

4.24.

4.25. Спроецируем единичный столбец на «плоскость», обозначим его $1'$. Делаем проекцию y на «плоскость» и на $1'$. Далее аналогично.

4.26. Проекция y на $\hat y$ это $\hat y$, поэтому оценки коэффициентов будут 0 и 1. Оценка дисперсии $\frac{RSS}{(n-2)ESS}$. Нарушены предпосылки теоремы Гаусса-Маркова, например, ошибки новой модели в сумме дают 0, значит коррелированы.

4.27. Либо в регрессию включена константа, либо единичный столбец (тут была опечатка, столбей) можно получить как линейную комбинацию регрессоров, например, включены дамми-переменные для каждого возможного значения качественной переменной.

4.28. Сами оценки коэффициентов никак детерминистически не связаны, но при большом размере подвыборок примерно равны. А ковариационные матрицы связаны соотношением $\mathrm{Var}(\hat\beta_a)^{-1} + \mathrm{Var}(\hat\beta_b)^{-1} = \mathrm{Var}(\hat\beta_{tot})^{-1}$

4.29.

4.30.

4.31.

4.32. Подсказка: запишите матрицу X как блочную и, пользуясь матричным выражением для $\hat\beta$ и формулой Фробениуса, найдите $\hat\beta_2$.

1. Да, верно. $X = (X_1 X_2)$ — блочная матрица. Аналогично, $\hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix}$ — блочная матрица (хотя на самом деле вектор).

$$\hat{\beta} = (X'X)^{-1}X'y = ((X_1X_2)'(X_1X_2))^{-1}(X_1X_2)'y = (\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}(X_1X_2))^{-1}(\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}y =$$
$$= \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y$$

Запишем и докажем формулу Фробениуса для обращения блочных матриц.

Формула Фробениуса:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ -H^{-1}CA^{-1} & H^{-1} \end{pmatrix}$$

где $H = D - CA^{-1}B$.

Докажем формулу, обращая матрицу методом Гаусса. Умножим слева на $\begin{pmatrix} A^{-1} & 0 \\ 0 & I \end{pmatrix}$

$$\left(\begin{array}{cc|cc} A & B & I & 0 \\ C & D & 0 & I \end{array}\right) = \left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ C & D & 0 & I \end{array}\right) =$$

вычтем из второй строки первую, умноженную на C

$$= \left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & D - CA^{-1}B & -CA^{-1} & I \end{array}\right) =$$

умножим слева на $\begin{pmatrix} I & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{pmatrix}$

$$= \left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & I & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array}\right) =$$

вычтем из первой строки вторую, умноженную на $A^{-1}B$.

$$= \left(\begin{array}{cc|cc} I & 0 & A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ 0 & I & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{array}\right)$$

Значит,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} =$$
$$= \begin{pmatrix} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ -H^{-1}CA^{-1} & H^{-1} \end{pmatrix}$$

По формуле Фробениуса получим, что

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} = \begin{pmatrix} (X_1'X_1)^{-1} + (X_1'X_1)^{-1}X_1'X_2H^{-1}X_2'X_1(X_1'X_1)^{-1} & -(X_1'X_1)^{-1}X_1'X_2H^{-1} \\ -H^{-1}X_2'X_1(X_1'X_1)^{-1} & H^{-1} \end{pmatrix},$$

где $H = X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$. Верхняя строка в данном пункте не важна, и сейчас её опустим. Заметим, что

$$H = X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 = X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2 = X_2'M_1X_2$$

Итак,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y =$$

$$= \begin{pmatrix} ? & ? \\ -H^{-1}X_2'X_1(X_1'X_1)^{-1} & H^{-1} \end{pmatrix} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y =$$

$$= \begin{pmatrix} ? \\ -H^{-1}X_2'X_1(X_1'X_1)^{-1}X_1' + H^{-1}X_2' \end{pmatrix} y =$$

$$= \begin{pmatrix} ? \\ H^{-1}X_2'(I - X_1(X_1'X_1)^{-1}X_1') \end{pmatrix} y =$$

$$= \begin{pmatrix} ? \\ H^{-1}X_2'M_1y \end{pmatrix} = \begin{pmatrix} ? \\ (X_2'M_1X_2)^{-1}X_2'M_1y \end{pmatrix}$$

$$\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1y$$

Заметим свойства матрицы-проектора M_1.

$$M_1' = (I - X_1(X_1'X_1)^{-1}X_1')' = I - X_1(X_1'X_1)^{-1}X_1' = M_1$$

$$(M_1)^2 = (I - X_1(X_1'X_1)^{-1}X_1')^2 = I - 2X_1(X_1'X_1)^{-1}X_1' + X_1(X_1'X_1)^{-1}X_1' \cdot X_1(X_1'X_1)^{-1}X_1' =$$
$$= I - X_1(X_1'X_1)^{-1}X_1' = M_1$$

Значит,

$$\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1y = (X_2'M_1M_1X_2)^{-1}X_2'M_1M_1y = (X_2'M_1'M_1X_2)^{-1}X_2'M_1'M_1y =$$
$$= ((M_1X_2)'M_1X_2)^{-1}(M_1X_2)'M_1y$$

но ведь и

$$\hat{\gamma}_2 = ((M_1X_2)'M_1X_2)^{-1}(M_1X_2)'M_1y$$

Значит, $\hat{\beta}_2 = \hat{\gamma}_2 = ((M_1X_2)'M_1X_2)^{-1}(M_1X_2)'M_1y$, что и требовалось доказать.

2. Да, верно.

$$\hat{y} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2$$

$$M_1\hat{\varepsilon} = M_1y - M_1\hat{y} = M_1y - M_1(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) = M_1y - M_1X_2\hat{\beta}_2 - M_1X_1\hat{\beta}_1$$
$$M_1X_1 = (I - X_1(X_1'X_1)^{-1}X_1')X_1 = X_1 - X_1(X_1'X_1)^{-1}X_1'X_1 = 0$$
$$M_1\hat{\varepsilon} = M_1y - M_1X_2\hat{\beta}_2 = M_1y - M_1X_2\hat{\gamma}_2 = \hat{u}$$

$$M_1\hat{\varepsilon} = M_1(y - \hat{y}) = M_1(I - X(X'X)^{-1}X')y = (I - X(X'X)^{-1}X')y$$

так как M_1 ортогональное дополнение к X_1, а $(I - X(X'X)^{-1}X')y$ уже лежит в ортогональном дополнении к X_1, так как $I - X(X'X)^{-1}X'$ ортогональное дополнение к к прямой сумме пространств X_1 и X_2 —$X_1 \oplus X_2$.

4.33.

4.34. Докажем несмещенность МНК-оценок.

$$\mathbb{E}\hat{\beta} = \mathbb{E}\left((X^TX)^{-1}X^Ty\right) = (X^TX)^{-1}X^T\mathbb{E}(y) =$$

$$= (X^TX)^{-1}X^T\mathbb{E}(X\beta + \varepsilon) = (X^TX)^{-1}X^TX\beta = \beta$$

Обозначим $\varphi(X, y) = (X^T X)^{-1} X^T y$. Тогда $\hat{\beta} = \varphi(X, y)$. Покажем, что функция φ линейна по переменной y.

1. $\varphi(X, \lambda \cdot y) = (X^T X)^{-1} X^T (\lambda \cdot y) = \lambda (X^T X)^{-1} X^T y = \lambda \cdot \varphi(X, y)$

2. $\varphi(X, y + z) = (X^T X)^{-1} X^T (y + z) = (X^T X)^{-1} X^T y + (X^T X)^{-1} X^T z = \varphi(X, y) + \varphi(X, z)$

Что и требовалось доказать.

4.35. Нет, так как для функции $\varphi(X, y) = (X^T X)^{-1} X^T y$ не выполнено, например, свойство однородности по переменной X. Действительно,

$$\varphi(X, \lambda \cdot y) = ((\lambda \cdot X)^T (\lambda \cdot X))^{-1} (\lambda \cdot X)^T y = \frac{1}{\lambda} \cdot (X^T X)^{-1} X^T y = \frac{1}{\lambda} \varphi(X, y)$$

.

4.36. $\tilde{\beta} = (X^T C X)^{-1} X^T C y$, где

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & n \end{bmatrix}$$

4.37. $Pi = i \Leftrightarrow P\pi = \pi$ поскольку, если матрицу π записать по столбцам $\pi = \frac{1}{n} \begin{bmatrix} i & i & \ldots & i \end{bmatrix}$, то можно записать следующую цепочку равенств $P\pi = P \frac{1}{n} \begin{bmatrix} i & i & \ldots & i \end{bmatrix} = \frac{1}{n} \begin{bmatrix} Pi & Pi & \ldots & Pi \end{bmatrix} = \frac{1}{n} \begin{bmatrix} i & i & \ldots & i \end{bmatrix} \Leftrightarrow Pi = i$.

Свойство $P^2 = P$ имеет место независимо от выполнимости условия $Pi = i$. Действительно, $P^2 = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = P$.

Рассмотрите пример $y = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}^T$, $x = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$. Постройте регрессию $y = \beta x + \varepsilon$ без свободного члена. Убедитесь, что $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ и $\overline{Y} = \overline{\hat{Y}} = 0$, но $Pi \neq i$.

Ответ: $P\pi = \pi$

4.38. (1), (2) $\Leftrightarrow$ (3), (5)

4.39.

$$\mathbb{E}(\varepsilon^T \pi \varepsilon) = \mathbb{E}(\mathrm{tr}[\varepsilon^T \pi \varepsilon]) = \mathbb{E}(\mathrm{tr}[\pi \varepsilon \varepsilon^T]) = \mathrm{tr}[\pi \mathbb{E}(\varepsilon \varepsilon^T)] =$$

$$\mathrm{tr}[\pi \operatorname{Var}(\varepsilon)] = \mathrm{tr} \left[\frac{1}{n} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \sigma_n^2 \end{bmatrix} \right] =$$

$$\frac{1}{n} tr \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_n^2 \\ \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_n^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_n^2 \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \qquad (1.27)$$

4.40.

1.

$$RSS = \hat{\varepsilon}^T \hat{\varepsilon} y^T (I - P) y = y^T y - y^T P y = y^T y - y^T X (X^T X)^{-1} X^T y; \qquad (1.28)$$

При этом $y^T y = 3924$, а

$$y^T X(X^T X)^{-1} X^T y = \begin{bmatrix} 460 & 810 & 615 & 712 \end{bmatrix} \begin{bmatrix} 0.038 & -0.063 & -0.063 & 0.100 \\ -0.063 & 1.129 & 1.107 & -2.192 \\ -0.063 & 1.107 & 1.110 & -2.170 \\ 0.100 & -2.192 & -2.170 & 4.292 \end{bmatrix} \cdot \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix} = 3051.2 \quad (1.29)$$

Итого, $RSS = 3924 - 3051.2 = 872.8$

$\hat{\sigma}^2_\varepsilon = \frac{RSS}{n-k} = \frac{872.8}{100-4} = 9.0917$

$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2_\varepsilon (X^T X)^{-1} \Rightarrow \widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.56939$, $\widehat{\mathrm{Var}}(\hat{\beta}_1) = 0.34251$, $\widehat{\mathrm{Var}}(\hat{\beta}_2) = 10.269$

$\widehat{\mathrm{Corr}}(\hat{\beta}_1, \hat{\beta}_2) = \frac{\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2)}} = -0.30361$

2. (указание) $\widehat{\mathrm{Corr}}(x_2, x_3) = \frac{\sum (x_{i2} - \overline{x}_2)(x_{i3} - \overline{x}_3)}{\sqrt{\sum (x_{i2} - \overline{x}_2)}\sqrt{\sum (x_{i3} - \overline{x}_3)}}$. Все необходимые величины можно извлечь из матрицы $X^T X$ — это величины $\sum x_{i2}$ и $\sum x_{i3}$, а остальное – из матрицы $X^T(I - \pi)X = X^T X - X^T \pi X = X^T X - (\pi X)^T \pi X$. При этом имейте в виду, что $\pi X = \begin{bmatrix} 1 & \overline{x}_1 & \overline{x}_2 & \overline{x}_3 \\ \dots & \dots & \dots & \dots \\ 1 & \overline{x}_1 & \overline{x}_2 & \overline{x}_3 \end{bmatrix}$ и $\overline{x}_1 = 1.23$, $\overline{x}_2 = 0.96$, $\overline{x}_3 = 1.09$

3. $$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix} = \begin{bmatrix} 0.03767 & -0.06263 & -0.06247 & 0.1003 \\ -0.06263 & 1.129 & 1.107 & -2.192 \\ -0.06247 & 1.107 & 1.110 & -2.170 \\ 0.1003 & -2.192 & -2.170 & 4.292 \end{bmatrix} \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix} = \begin{bmatrix} -0.40221 \\ 6.1234 \\ 5.9097 \\ -7.5256 \end{bmatrix}$$

   $t = \frac{\hat{\beta}_2}{\sqrt{\mathrm{Var}(\hat{\beta}_2)}} \sim t_{100-4}$

   $t = \frac{\hat{\beta}_2}{\sqrt{\mathrm{Var}(\hat{\beta}_2)}} = \frac{6.1234}{\sqrt{10.269}} = 1.9109 \Rightarrow \hat{\beta}_2$ — не значим.

4.41.

1. $\widehat{\mathrm{Cov}} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \hat{\sigma}^2_\varepsilon (X^T X)^{-1}$ — несмещённая оценка для ковариационной матрицы МНК-коэффициентов. Действительно, $\mathbb{E}\widehat{\mathrm{Cov}} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \mathbb{E}\hat{\sigma}^2_\varepsilon (X^T X)^{-1} = \sigma^2_\varepsilon (X^T X)^{-1} = \mathrm{Cov} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$.

   Поэтому искомая оценка $\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2_\varepsilon \left[(X^T X)^{-1}\right]_{23}$, где $\left[(X^T X)^{-1}\right]_{23}$ — элемент матрицы $(X^T X)^{-1}$, расположенный во второй строке, 3-м столбце.

   Заметим, что $\hat{\sigma}^2_{\hat{\beta}_2} = \hat{\sigma}^2_\varepsilon \left[(X^T X)^{-1}\right]_{22} \Rightarrow 0.7^2 = \hat{\sigma}^2_\varepsilon \cdot (3030) \Rightarrow \hat{\sigma}^2_\varepsilon = 0.00016172$

   Значит, $\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = 0.00016172 \cdot (-589) = -0.095253$.

2. $t = \frac{\hat{\beta}_2 + \hat{\beta}_3 - \beta_2 - \beta_3}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3)}} \sim t_{n-k}$

   Требуется проверить $H_0 : \beta_2 + \beta_3 = 1$.

   $\widehat{\mathrm{Var}}(\hat{\beta}_2 + \hat{\beta}_3) = \widehat{\mathrm{Var}}(\hat{\beta}_2) + \widehat{\mathrm{Var}}(\hat{\beta}_3) + 2\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = 0.7^2 + 0.138^2 + 2 \cdot 0.095253 = 0.319044$

$t = \frac{\hat{\beta}_2+\hat{\beta}_3-\beta_2-\beta_3}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2+\hat{\beta}_3)}} = \frac{0.76+0.19-1}{\sqrt{0.319044}} = -0.088520674$

Значит, гипотеза не отвергается на любом «разумном» уровне значимости.

3. Мы знаем, что $\frac{\hat{\beta}_2+\hat{\beta}_3-\beta_2-\beta_3}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2+\hat{\beta}_3)}} \sim t_{n-k} = t_{15-3}$, поэтому построить доверительный интервал для $\beta_2+\beta_3$ не составляет труда. $\mathbb{P}\left(\left|\frac{\hat{\beta}_2+\hat{\beta}_3-\beta_2-\beta_3}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2+\hat{\beta}_3)}}\right| < t^*\right) = 0.95$

   Обозначим $se = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2+\hat{\beta}_3)}$, тогда:

$$\mathbb{P}\left(\left|\frac{\hat{\beta}_2+\hat{\beta}_3-\beta_2-\beta_3}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2+\hat{\beta}_3)}}\right| < t^*\right) = \\ \mathbb{P}\left(-t^*se < \hat{\beta}_2+\hat{\beta}_3-\beta_2-\beta_3 < t^*se\right) = \\ \mathbb{P}\left(-t^*se-(\hat{\beta}_2+\hat{\beta}_3) < -\beta_2-\beta_3 < -(\hat{\beta}_2+\hat{\beta}_3)+t^*se\right) = \\ \mathbb{P}\left((\hat{\beta}_2+\hat{\beta}_3)+t^*se > \beta_2+\beta_3 > (\hat{\beta}_2+\hat{\beta}_3)-t^*se\right) \quad (1.30)$$

   Отсюда получаем доверительный интервал

$$\beta_2+\beta_3 \in \\ [(0.76+0.19)-2.16\cdot 0.319; (0.76+0.19)+2.16\cdot 0.319] \quad (1.31)$$

   Или $0.26 < \beta_2+\beta_3 < 1.639$

4.42.

Метод наименьших квадратов:

$$\hat{\varepsilon}'\hat{\varepsilon} \to \min_{\hat{\beta}}$$

$$(y-\hat{y})'(y-\hat{y}) \to \min_{\hat{\beta}}$$

$$(y-X\hat{\beta})'(y-X\hat{\beta}) \to \min_{\hat{\beta}}$$

$$y'y-\hat{\beta}'X'y-y'X\hat{\beta}+\hat{\beta}'X'X\hat{\beta} \to \min_{\hat{\beta}}$$

Воспользуемся тем, что : $\frac{\partial x'A}{\partial x'} = A'$, $\frac{\partial Ax}{\partial x'} = A$, $\frac{\partial x'Ax}{\partial x'} = x'(A'+A)$ Условие первого порядка:

$$-2(X'y)'+(X'X+(X'X)')\hat{\beta}' = 0$$

$$-2X'y+2\hat{\beta}X'X = 0$$

и

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}((X'X)^{-1}X'y) = (X'X)^{-1}X'\,\mathrm{Var}(y)((X'X)^{-1}X')' = \\ = (X'X)^{-1}X'\,\mathrm{Var}(X\beta+\epsilon)X(X'X)^{-1}$$

Так как β является константой, то $\operatorname{Var}(X\beta) = 0$. Тогда

$$\operatorname{Var}(\hat{\beta}) = (X'X)^{-1}X' \operatorname{Var}(X\beta + \epsilon)X(X'X)^{-1} = \\ = (X'X)^{-1}X' \operatorname{Var}(\epsilon)X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} = \\ = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

4.43. Находим $X'X$, её элементы и есть то, что нужно.
4.44.

$$\widehat{Var}(\hat{\beta}_1 + \hat{\beta}_2 - \hat{\beta}_3) = \operatorname{Var}(\hat{\beta}1) + \operatorname{Var}(\hat{\beta}2) + \operatorname{Var}(\hat{\beta}3) + 2\operatorname{Cov}(\hat{\beta}1, \hat{\beta}2) - 2\operatorname{Cov}(\hat{\beta}1, \hat{\beta}3) - 2\operatorname{Cov}(\hat{\beta}2, \hat{\beta}3)$$

$$\widehat{Var}(\hat{\beta}_1 + \hat{\beta}_2 - \hat{\beta}_3) = 1/3 + 4/3 + 2 - 2/3 + 2 = 5$$

4.45. Из того, что $\widehat{Var}(\hat{\beta}) = \sigma^2 \cdot (X'X)^{-1}$ видно, что $\sigma^2 = \frac{1}{3}$.
$RSS = \sigma^2 \cdot (n-k) = 1/3 \cdot 2 = 2/3$

$$R^2 = 49\frac{1}{3}/50 = 148/150$$

$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{148/2}{2/2} = 74$ $F^{crit}_{0.05,2,2} = 19 < 74$ гипотеза отвергается, регрессия значима.
4.46.
5.1.
(1) Пусть X_i — количество чатлов, заработанное в i-ый день. Запишем функцию правдоподобия:

$$\mathrm{L}(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} \exp(-\lambda)$$

Возьмем от неё логарифм и решим задачу максимизации по параметру λ:

$$l(\lambda) = \ln\lambda \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \ln X_i! - n\lambda \to \max_{\lambda}$$

$$\frac{\sum_{i=1}^{n} X_i}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} = 2.5$$

Таким образом, оценкой максимального правдоподобия для параметра пуассоновского распределения λ является среднее от наблюдаемых величин.

(2) Требуется найти такое n, при котором выполненяется следующее равенство:

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant 2200\right) = 0.99$$

Из ЦПТ знаем, что центрированная и нормированная сумма случайных величин X_i является стандартной нормальной случайной величиной. Пользуемся тем фактом, что для пуассоновского распределения $\mathbb{E}(X_i) = \operatorname{Var}(X_i) = \lambda$.

$$\frac{\sum_{i=1}^{n} X_i - n\hat{\lambda}}{\sqrt{n\hat{\lambda}}} \sim N(0,1)$$

$$\hat{\mathbb{P}}\left(Z_{0,1} \geqslant \frac{2200 - 2.5n}{\sqrt{2.5n}}\right) = 0.99$$

Пользуемся таблицей стандартного нормального распределения, чтобы узнать значение правой части неравенства (-2.33) и решаем полученное уравнение относительно n. Получается, что им необходимо потратить 922 дня.

(3) Для построения доверительного интервала так же пользуемся центрированной нормированной величиной:

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}}\right| \leqslant a\right) = 0.95$$

Из таблицы для стандартного нормального распределения узнаем, что порог a равен 1.96, решаем полученные неравенства относительно параметра λ и получаем искомый доверительный интервал:

$$2.19 \leqslant \lambda \leqslant 2.82$$

(4) Проверим гипотезу $H_0 : \lambda = 2$ с помощью трех тестов.

- Тест отношения правдоподобия. Воспользуемся функцией l, найденной в первом пункте, для того чтобы рассчитать LR-статистику:

$$\mathrm{LR} = -2(l(2) - l(2.5)) = 11.57$$

  Табличное значение $\chi^2_1 = 3.84$, рассчитанная LR статистика больше табличного значения, поэтому можно сделать вывод о том, что нулевая гипотеза отвергается.

- Тест Вальда. Так как основная гипотеза подразумевает только одно линейное ограничение $\lambda - 2 = 0$, то статистика Вальда принимает следующий вид:

$$\mathrm{W} = \left(\hat{\lambda}_{UR} - 2\right)^2 \left(-E\left(\frac{\partial^2 l}{\partial \lambda^2}\right)_{UR}\right)^{-1} = 10$$

  Табличное значение остается тем же, можно сделать вывод о том, что нулевая гипотеза отвергается.

- Тест множителей Лагранжа. Для заданного ограничения значение статистики рассчитывается по формуле:

$$\mathrm{LM} = \left(\frac{\partial l}{\partial \lambda}\right)^2_R \left(\left(\frac{\partial^2 l}{\partial \lambda^2}\right)_R\right)^{-1} = 15.625$$

  Нулевая гипотеза отвергается и с помощью последнего теста.

5.2. Эта задача решается в том же ключе, что и предыдущая!

(1) Пусть X_i — продолжительность i-го сеанса связи. Запишем функцию правдоподобия для экспоненциально распределенных случайных величин, возьмем логарифм, найдем максимум полученной функции (в минутах):

$$\mathrm{L}(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda X_i)$$

$$l(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i \to \max_{\lambda}$$

$$\frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0 \Rightarrow \hat{\lambda} = \frac{1}{\bar{X}} = 0.15$$

(2) Построим доверительный интервал, пользуясь тем, что для экспоненциального распределения $\mathbb{E}(X_i) = 1/\lambda, \mathrm{Var}(X_i) = 1/\lambda^2$:

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^{n} X_i - n/\lambda}{\sqrt{n/\lambda^2}}\right| \leqslant a\right) = 0.95$$

$$0.12 \leqslant \lambda \leqslant 0.18$$

(3) Проверяем гипотезу $H_0 : 1/\lambda = 5$.

- Тест отношения правдоподобия. Нулевая гипотеза отвергается.

$$\mathrm{LR} = -2(l(0.2) - l(0.15)) = 8 > 3.84$$

- Тест Вальда. Нулевая гипотеза отвергается.

$$\mathrm{W} = \left(\hat{\lambda}_{UR} - 0.2\right)^2 \left(-E\left(\frac{\partial^2 l}{\partial \lambda^2}\right)_{UR}\right)^{-1} = 11.11 > 3.84$$

- Тест множителей Лагранжа. Нулевая гипотеза упорно отвергается.

$$\mathrm{LM} = \left(\frac{\partial l}{\partial \lambda}\right)^2_R \left(\left(\frac{\partial^2 l}{\partial \lambda^2}\right)_R\right)^{-1} = 10.24 > 3.84$$

5.3.
5.4. $\hat{\theta} = 1/\bar{Y}$, $\hat{\beta} = \bar{X}/\bar{Y}$, $\hat{a} = 1/(1+\bar{X})$
5.5. В данном примере мы имеем
$\theta = \begin{bmatrix}\mu & \nu\end{bmatrix}'$ — вектор неизвестных параметров
$\Theta = \mathbb{R} \times (0; +\infty)$ — множество допустимых значений вектора неизвестных параметров
Функция правдоподобия имеет вид:

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\nu}} \cdot \exp\left\{-\frac{(x_i - \mu)^2}{2\nu}\right\} =$$
$$(2\pi)^{-n/2} \cdot \nu^{-n/2} \cdot \exp\left\{-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\nu}\right\} \quad (1.32)$$

Логарифмическая функция правдоподобия:

$$l(\theta) := \ln \mathrm{L}(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\nu - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\nu}$$

$\Theta_{UR} = \Theta$
$\Theta_R = \{(0, 1)\}$
Из системы уравнений

$$\begin{cases} \frac{\partial l}{\partial \mu} = \frac{\sum_{i=1}^{n}(x_i - \mu)}{\nu} = 0 \\ \frac{\partial l}{\partial \nu} = -\frac{n}{2\nu} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\nu^2} = 0 \end{cases}$$

находим
$\hat{\theta}_{UR} = (\hat{\mu}_{UR}, \hat{\nu}_{UR})$, где $\hat{\mu}_{UR} = \overline{x} = -1.5290$, $\hat{\nu}_{UR} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 1.0603$
$\hat{\theta}_R = (\hat{\mu}_R, \hat{\nu}_R) = (0, 1)$
По имеющимся данным находим
$l(\hat{\theta}_R) = -\frac{10}{2}\ln(2\pi) - \frac{10}{2}\ln 1 - \frac{\sum_{i=1}^{n}(x_i - 0)^2}{2\cdot 1} = -26.1804$
$l = -\frac{10}{2}\ln(2\pi) - \frac{10}{2}\ln(1.0603) - \frac{\sum_{i=1}^{n}(x_i + 1.5290)^2}{2\cdot 1.0603} = -14.4824$
$LR_{\text{набл}} = -2(l(\hat{\theta}_R) - l) = -2\cdot(-26.1804 + 14.4824) = 23.3959$

Критическое значение χ^2 распределения с двумя степенями свободы, отвечающее уровню значимости 5%, равно 5.9915. Следовательно, тест отношения правдоподобия говорит о том, что гипотеза H_0 должна быть отвергнута.

Для выполнения тестов Вальда и множителей Лагранжа нам понадобится информационная матрица Фишера
$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{v}$, $\frac{\partial^2 l}{\partial \nu \partial \mu} = -\frac{\sum_{i=1}^{n}(x_i - \mu)}{\nu^2}$, $\frac{\partial^2 l}{\partial \nu^2} = \frac{n}{2\nu^2} - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\nu^3}$

$\mathbb{E}\frac{\partial^2 l}{\partial\nu\partial\mu} = -\frac{\sum_{i=1}^n \mathbb{E}(x_i-\mu)}{\nu^2} = 0$, $\mathbb{E}\frac{\partial^2 l}{\partial\nu^2} = \frac{n}{2\nu^2} - \frac{\sum_{i=1}^n \mathbb{E}(x_i-\mu)^2}{\nu^3} = \frac{n}{2\nu^2} - \frac{n\nu}{nu^3} = -\frac{n}{2\nu^2}$

$I(\theta) = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 l}{\partial\mu^2} & \frac{\partial^2 l}{\partial\nu\partial\mu} \\ \frac{\partial^2 l}{\partial\nu\partial\mu} & \frac{\partial^2 l}{\partial\nu^2} \end{bmatrix} = \begin{bmatrix} \frac{n}{\nu} & 0 \\ 0 & \frac{n}{2\nu^2} \end{bmatrix}$

$I(\hat\theta_{UR}) = \begin{bmatrix} \frac{n}{\hat\nu_{UR}} & 0 \\ 0 & \frac{n}{2\cdot\hat\nu_{UR}^2} \end{bmatrix} = \begin{bmatrix} \frac{10}{1.0603} & 0 \\ 0 & \frac{10}{2\cdot 1.0603^2} \end{bmatrix} = \begin{bmatrix} 9.4307 & 0 \\ 0 & 4.4469 \end{bmatrix}$

$g(\hat\theta_{UR}) = \begin{bmatrix} \hat\mu_{UR} - 0 \\ \hat\nu_{UR} - 1 \end{bmatrix} = \begin{bmatrix} -1.5290 - 0 \\ 1.0603 - 1 \end{bmatrix} = \begin{bmatrix} -1.5290 \\ 0.0603 \end{bmatrix}$

$\frac{\partial g}{\partial\theta'} = \begin{bmatrix} \frac{\partial g_1}{\partial\mu} & \frac{\partial g_1}{\partial\nu} \\ \frac{\partial g_2}{\partial\mu} & \frac{\partial g_2}{\partial\nu} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\frac{\partial g'}{\partial\theta} = \begin{bmatrix} \frac{\partial g_1}{\partial\mu} & \frac{\partial g_2}{\partial\mu} \\ \frac{\partial g_1}{\partial\nu} & \frac{\partial g_2}{\partial\nu} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$W_{\text{набл}} = g'(\hat\theta_{UR}) \cdot \left[\frac{\partial g}{\partial\theta'}(\hat\theta_{UR}) \cdot I^{-1}(\hat\theta_{UR}) \cdot \frac{\partial g'}{\partial\theta}(\hat\theta_{UR})\right]^{-1} g(\hat\theta_{UR}) =$

$\begin{bmatrix} -1.5290 & 0.0603 \end{bmatrix} \cdot \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 9.4307 & 0 \\ 0 & 4.4469 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]^{-1} \cdot \begin{bmatrix} -1.5290 \\ 0.0603 \end{bmatrix} = 22.0635$

Тест Вальда также говорит о том, что на основании имеющихся наблюдений гипотеза H_0 должна быть отвергнута.

$I(\hat\theta_R) = \begin{bmatrix} \frac{n}{\hat\nu_R} & 0 \\ 0 & \frac{n}{2\cdot\hat\nu_R^2} \end{bmatrix} = \begin{bmatrix} \frac{10}{1} & 0 \\ 0 & \frac{10}{2\cdot 1^2} \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}$

$\frac{\partial l}{\partial\theta}(\hat\theta_R) = \begin{bmatrix} \frac{\sum_{i=1}^n (x_i - \hat\mu_R)}{\hat\nu_R} \\ -\frac{n}{2\cdot\hat\nu_R} + \frac{\sum_{i=1}^n (x_i-\hat\mu_R)^2}{2\cdot\hat\nu_R^2} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^n (x_i - 0)}{1} \\ -\frac{10}{2\cdot 1} + \frac{\sum_{i=1}^n (x_i-0)^2}{2\cdot 1^2} \end{bmatrix} = \begin{bmatrix} -15.29 \\ 11.9910 \end{bmatrix}$

$LM_{\text{набл}} = \left[\frac{\partial l}{\partial\theta}(\hat\theta_R)\right]' \cdot I^{-1}(\hat\theta_R) \cdot \left[\frac{\partial l}{\partial\theta}(\hat\theta_R)\right] = \begin{bmatrix} -15.29 & 11.9910 \end{bmatrix} = \cdot \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}^{-1} \cdot \begin{bmatrix} -15.29 \\ 11.9910 \end{bmatrix} = 52.1354$

Тест множителей Лагранжа также указывает на то, что гипотеза H_0 должна быть отвергнута.

5.6. В данной задаче мы имеем:

$\theta = p$ — вектор неизвестных параметров

$\Theta = (0, 1)$ — множество допустимых значений вектора неизвестных параметров

Функция правдоподобия имеет вид:

$$\mathrm{L}(\theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n p^{x_i} \cdot (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i}$$

Логарифмическая функция правдоподобия:

$$l(\theta) := \ln \mathrm{L}(\theta) = \left(\sum_{i=1}^n x_i\right) \cdot \ln p + \left(n - \sum_{i=1}^n x_i\right) \cdot \ln(1-p)$$

$\Theta_{UR} = \Theta$

$\Theta_R = \{0.5\}$

Решая уравнение правдоподобия

$$\frac{\partial l}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0$$

получаем

$\hat\theta_{UR} = \hat p_{UR}$, где $\hat p_{UR} = \overline{x} = 0.42$

$\hat\theta_R = \hat p_R = 0.5$

По имеющимся данным находим

$l(\hat\theta_R) = 42 \cdot \ln(0.5) + (100 - 42) \cdot \ln(1 - 0.5) = -69.3147$

$l(\hat\theta_{UR} = 42 \cdot \ln(0.42) + (100 - 42) \cdot \ln(1 - 0.42) = -68.0292$

$LR_{\text{набл}} = -2(l(\hat\theta_R) - l) = -2 \cdot (-69.3147 + 68.0292) = 2.5710$

Критическое значение χ^2 распределения с одной степенью свободы, отвечающее за 5% уровень значимости, равно 3.8414. Следовательно, тест отношения правдоподобия говорит о том, что на основании имеющихся данных, основная гипотеза $H_0 : p = 0.5$ не может быть отвергнута.

Для выполнения тестов Вальда и множителей Лагранжа нам понадобится информационная матрица Фишера

$\frac{\partial^2 l}{\partial p^2} = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}$

$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 l}{\partial p^2}\right] = -\mathbb{E}\left[-\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}\right] = -\left(-\frac{np}{p^2} - \frac{n-np}{(1-p)^2}\right) = \frac{n}{p(1-p)}$

$I(\hat{\theta}_{UR}) = \frac{n}{\hat{p}_{UR}(1-\hat{p}_{UR})} = \frac{100}{0.42\times(1-0.42)} = 172.4138$

$g(\hat{\theta}_{UR}) = \hat{\theta}_{UR} - 0.5 = 0.42 - 0.5 = -0.08$

$\frac{\partial g}{\partial \theta'} = 1'$, $\frac{\partial g'}{\partial \theta} = 1$

$W_{\text{набл}} = g'(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g'}{\partial \theta}(\hat{\theta}_{UR})\right]^{-1} g(\hat{\theta}_{UR}) = [-0.08]' \cdot [1' \cdot 172.4138^{-1} \cdot 1]^{-1} \cdot [-0.08] = 2.6272$

Тест Вальда также говорит о том, что гипотеза H_0 не отвергается.

$I(\hat{\theta}_R) = \frac{n}{\hat{p}_R(1-\hat{p}_R)} = \frac{100}{0.5\times(1-0.5)} = 400$

$\frac{\partial l}{\partial \theta}(\hat{\theta}_R) = \frac{\sum_{i=1}^n x_i}{\hat{p}_R} - \frac{n-\sum_{i=1}^n x_i}{1-\hat{p}_R} = \frac{42}{0.5} - \frac{100-42}{1-0.5} = -32$

$LM_{\text{набл}} = \left[\frac{\partial l}{\partial \theta}(\hat{\theta}_R)\right]' \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial l}{\partial \theta}(\hat{\theta}_R)\right] = [-32]' \cdot [400]^{-1} \cdot [-32] = 2.56$

Согласно тесту множителей Лагранжа, основная гипотеза H_0 не может быть отвергнута.

5.7. В данной задаче мы имеем

$\theta = \lambda$ — вектор неизвестных параметров

$\Theta = (0, +\infty)$ — множество допустимых значений вектора неизвестных параметров

Функция правдоподобия имеет вид:

$$\mathrm{L}(\theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! \dots x_n!} e^{-\lambda n}$$

Логарифмическая функция правдоподобия:

$$l(\theta) := \ln \mathrm{L}(\theta) = \left(\sum_{i=1}^n x_i\right) \cdot \ln \lambda - \sum_{i=1}^n \ln(x_i!) - \lambda n$$

$\Theta_{UR} = \Theta$

$\Theta_R = \{2\}$

Решая уравнение правдоподобия

$$\frac{\partial l}{\partial p} = \frac{\sum_{i=1}^n x_i}{\lambda} - n = 0$$

получаем

$\hat{\theta}_{UR} = \hat{\lambda}_{UR}$, где $\hat{\lambda}_{UR} = \overline{x} = 1.7$

$\hat{\theta}_R = \hat{p}_R = 2$

По имеющимся данным находим

$l(\hat{\theta}_R) = (80 \cdot 1.7) \cdot \ln(2) - \sum_{i=1}^n \ln(x_i!) - 2 \cdot 80 = -65.7319$

$l(\hat{\theta}_{UR} = (80 \cdot 1.7) \cdot \ln(1.7) - \sum_{i=1}^n \ln(x_i!) - 1.7 \cdot 80 = -63.8345$

$LR_{\text{набл}} = -2(l(\hat{\theta}_R) - l) = -2 \cdot (-65.7319 + 63.8345) = 3.7948$

Критическое значение χ^2 распределения с одной степенью свободы, отвечающее за 5% уровень значимости, равно 3.8414. Следовательно, тест отношения правдоподобия говорит о том, что на основании имеющихся данных, основная гипотеза $H_0 : \lambda = 2$ не может быть отвергнута.

Для выполнения тестов Вальда и множителей Лагранжа нам понадобится информационная матрица Фишера

$\frac{\partial^2 l}{\partial p^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2}$

$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 l}{\partial p^2}\right] = -\mathbb{E}\left[-\frac{\sum_{i=1}^n x_i}{\lambda^2}\right] = -\left(-\frac{n\lambda}{\lambda^2}\right) = \frac{n}{\lambda}$

$I(\hat{\theta}_{UR}) = \frac{n}{\hat{\lambda}_{UR}} = \frac{80}{1.7} = 47.0588$

$g(\hat{\theta}_{UR}) = \hat{\theta}_{UR} - 2 = 1.7 - 2 = -0.3$

$\frac{\partial g}{\partial \theta'} = 1'$, $\frac{\partial g'}{\partial \theta} = 1$

$W_{\text{набл}} = g'(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g'}{\partial \theta}(\hat{\theta}_{UR})\right]^{-1} g(\hat{\theta}_{UR}) = [-0.3]' \cdot [1' \cdot 47.0588^{-1} \cdot 1]^{-1} \cdot [-0.3] = 4.2352$

Поскольку наблюдаемое значение статистики Вальда превосходит критическое значение 3.8414, то гипотеза H_0 должна быть отвергнута.

$I(\hat{\theta}_R) = \frac{n}{\hat{\lambda}_R} = \frac{80}{2} = 40$

$\frac{\partial l}{\partial \theta}(\hat{\theta}_R) = \frac{\sum_{i=1}^n x_i}{\hat{\lambda}_R} - n = \frac{80 \cdot 1.7}{2} - 80 = -12$

$LM_{\text{набл}} = \left[\frac{\partial l}{\partial \theta}(\hat{\theta}_R)\right]' \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial l}{\partial \theta}(\hat{\theta}_R)\right] = [-12]' \cdot [40]^{-1} \cdot [-12] = 3.6$

Согласно тесту множителей Лагранжа, основная гипотеза H_0 не может быть отвергнута.

5.8. Для того чтобы записать функцию правдоподобия, найдем параметры распределения случайных величин y_i. Величины распределены нормально, так как задаются линейной функцией только от нормально распределенных случайных величин ϵ_i. Посчитаем математическое ожидание и дисперсию:

$$\mathbb{E}(y_i) = \mathbb{E}(\beta_1 + \beta_2 x_i + \epsilon_i) = \beta_1 + \beta_2 x_i$$
$$\text{Var}(y_i) = \text{Var}(\beta_1 + \beta_2 x_i + \epsilon_i) = \text{Var}(\epsilon_i) = \sigma^2 x_i^2$$
$$y_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2 x_i^2)$$

Теперь, используя формулу для функции плостности нормального распределения, можем в явном виде записать функцию правдоподобия:

$$\text{L} = \prod_{i=1}^n \frac{1}{\sigma |x_i| \sqrt{2\pi}} \exp\left(\frac{-(y_i - \beta_1 - \beta_2 x_i)^2}{2\sigma^2 x_i^2}\right)$$

5.9. Посчитаем математическое ожидание и дисперсию:

$$\mathbb{E}(y_i) = \mathbb{E}(\beta_1 + \beta_2 x_i + \epsilon_i) = \beta_1 + \beta_2 x_i$$
$$\text{Var}(y_i) = \text{Var}(\beta_1 + \beta_2 x_i + \epsilon_i) = \text{Var}(\epsilon_i) = \sigma^2 |x_i|$$
$$y_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2 |x_i|)$$

Функция правдоподобия:

$$\text{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2 |x_i|}} \exp\left(\frac{-(y_i - \beta_1 - \beta_2 x_i)^2}{2\sigma^2 |x_i|}\right)$$

5.10. (1) Для того, чтобы найти оценки для β и σ^2 методом максимального правдоподобия, составим сначала функцию правдоподобия:

$$\mathbb{E}(y_i) = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$
$$\text{Var}(y_i) = \sigma^2$$

$$\text{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - x_i'\beta)^2}{2\sigma^2}\right)$$

Логарифмическая функция правдоподобия:

$$l = n \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i'\beta)^2 \to \max_{\beta, \sigma^2}$$

Решим задачу максимизации:

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}(y_i - x_i'\beta)x_i = 0 \Rrightarrow \hat{\beta}_{ML} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{-n\sqrt{2\pi\sigma^2}}{2\sqrt{2\pi(\sigma^2)^3}} + \frac{\sum_{i=1}^n (y_i - x_i\beta)^2}{2(\sigma^2)^2} = 0 \Rightarrow \hat{\sigma}^2_{ML} = \frac{\text{RSS}}{n}$$

(2) Проверим, являются ли полученные оценки несмещенными.

$$\mathbb{E}(\hat{\beta}_{ML}) = \mathbb{E}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}\right) = \frac{\sum_{i=1}^n x_i \mathbb{E}(y_i)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i \mathbb{E}(x_i'\beta + \epsilon)}{\sum_{i=1}^n x_i^2} = \frac{\beta \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \beta$$

Таким образом, оценки максимального правдоподобия $\hat{\beta}_{ML}$ совпадают с полученными с помощью метода наименьших квадратов и являются несмещенными.

Можно заметить, что полученная оценка дисперсии $\hat{\sigma}^2_{ML}$ не совпадает с несмещенной оценкой метода наименьших квадратов и является смещенной:

$$\mathbb{E}(\hat{\sigma}^2_{ML}) = \mathbb{E}\left(\frac{\text{RSS}}{n}\right) = \frac{n-k}{n}\mathbb{E}(\hat{\sigma}^2_{OLS}) = \frac{n-k}{n}s^2$$

5.11. (1) Оценим параметр μ с помощью метода максимального правдоподобия. Функция правдоподобия:

$$\text{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(X_i-\mu)^2}{2}\right)$$

Логарифмическая функция правдоподобия:

$$l = n \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^n (X_i - \mu)^2 \to \max_{\mu}$$

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X} = 2$$

(2) Построим доверительный интервал:

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}}\right| \leqslant a\right) = 0.95$$

Из таблицы для стандартного нормального распределения получаем $a = 1.96$, решаем неравенство относительно μ:

$$1.8 \leqslant \mu \leqslant 2.2$$

(3) Проверим гипотезу $H_0 : \mu = 3$.

- LR-тест отвергает основную гипотезу:

$$\text{LR} = -2\left(-0.5\sum_{i=1}^n (X_i - 3)^2 + 0.5\sum_{i=1}^n (X_i - 2)^2\right) = 100 > \chi^2_1 = 3.84$$

- Тест Вальда при ограничении $\mu - 3 = 0$ отвергает основную гипотезу:

$$\text{W} = (-1)^2\left(-\mathbb{E}\left(\frac{\partial^2 l}{\partial \mu^2}\right)_{UR}\right) = n = 100 > \chi^2_1 = 3.84$$

- LM-тест также отвергает основную гипотезу:

$$\text{LM} = \left(\sum_{i=1}^n (X_i - 3n)^2\right)\left(-\mathbb{E}\left(\frac{\partial^2 l}{\partial \mu^2}\right)^{-1}_R\right) = 100 > \chi^2_1 = 3.84$$

(4) Построим доверительный интервал для вероятности $\mathbb{P}(X_i > 2.5)$:

$$\mathbb{P}(X_i > 2.5) = 1 - \mathbb{P}(X_i < 2.5) = 1 - \mathbb{P}(X_i - \mu < 2.5 - \mu) = 1 - \Phi_{0,1}(2.5 - \mu)$$

В предыдущем пункте мы нашли доверительный интервал для $\mu : 1.8 \leqslant \mu \leqslant 2.2$.
Таким образом, с 95% уверенностью искомая вероятность лежит в пределах:

$$0.24 \leqslant \mathbb{P}(X_i > 2.5) \leqslant 0.38$$

5.12. (1) Оценим параметр σ^2 с помощью метода максимального правдоподобия. Функция правдоподобия:

$$\mathrm{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-X_i^2}{2\sigma^2}\right)$$

Логарифмическая функция правдоподобия:

$$l = n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} X_i^2 \to \max_{\sigma}^{2}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} = 9$$

(2) Построим доверительный интервал. Воспользуемся тем фактом, что:

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Тогда, пользуясь табличными значениями для хи-квадрат распределения, получаем следующее неравенство:

$$\mathbb{P}\left(74.22 \leqslant \frac{\sum_{i=1}^{n} X_i^2}{\sigma^2} \leqslant 129.56\right) = 0.95$$

Решаем неравенство относительно σ^2:

$$6.95 \leqslant \sigma^2 \leqslant 12.13$$

(3) Проверим гипотезу $H_0 : \sigma^2 = 4$. Посчитаем статистики для трех тестов.

- $\mathrm{LR} = -2\left(\ln(4) - \ln(9)\right) = 43.9$
- $\mathrm{W} = \frac{900}{9^2}(5)^2 = 277.78$
- $\mathrm{LM} = 0.018\left(\frac{900}{4} - 100\right)^2 = 281.25$

Все три рассчитанных статистики превышают табличное значение для одного ограничения: $\chi_1^2 = 3.84$, поэтому все три теста отвергают нулевую гипотезу.

(4) Построим доверительный интервал для вероятности $\mathbb{P}(X_i > 2.5)$:

$$\mathbb{P}(X_i > 2.5) = 1 - \mathbb{P}(X_i < 2.5) = 1 - \mathbb{P}\left(\frac{X_i}{\sqrt{\sigma^2}} < \frac{2.5}{\sqrt{\sigma^2}}\right) = 1 - \Phi_{0,1}\left(\frac{2.5}{\sqrt{\sigma^2}}\right)$$

В предыдущем пункте мы нашли доверительный интервал для $\sigma : 6.95 \leqslant \sigma^2 \leqslant 12.13$.
Таким образом, с 95% уверенностью искомая вероятность лежит в пределах:

$$0.17 \leqslant \mathbb{P}(X_i > 2.5) \leqslant 0.24$$

5.13. (1) Оценим параметры μ, σ^2 с помощью метода максимального правдоподобия. Функция правдоподобия:

$$\mathrm{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\left(X_i-\mu\right)^2}{2\sigma^2}\right)$$

Логарифмическая функция правдоподобия:

$$l = n\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \to \max_{\mu,\sigma^2}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = 2$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(X_i-\hat{\mu}\right)^2}{n} = 5$$

(2) Построим доверительные интервалы для параметров μ и σ^2. Доверительный интервал для математического ожидания в случае неизвестной дисперсии строится, используя статистику:

$$\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{S} \sim t_{n-1}$$

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1} = 5.05$$

Тогда, пользуясь табличными значениями для распределения Стьюдента, получаем следующее неравенство:

$$\mathbb{P}\left(-1.98 \leqslant \frac{2-\mu}{0.225} \leqslant 1.98\right) = 0.95$$

Решаем неравенство относительно μ:

$$1.55 \leqslant \mu \leqslant 2.45$$

Приступим к интервалу для дисперсии. Построение доверительного интервала для дисперсии в случае неизвестного математического ожидания использует статистику:

$$\frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}$$

В качестве границ используем табличные значения хи-квадрат распределения, получаем следующее неравенство:

$$\mathbb{P}\left(73.36 \leqslant \frac{99\cdot 5.05}{\sigma^2} \leqslant 128.42\right) = 0.95$$

Решаем неравенство относительно σ^2:

$$3.84 \leqslant \sigma^2 \leqslant 6.8$$

5.14. $\hat{p}_1 = X_1/n$, $\hat{p}_2 = X_2/n$.

5.15. Функция правдоподобия:

$$\mathcal{L} = \prod_{i=1}^{n} \frac{\theta(\ln X_i)^{\theta-1}}{X_i}$$

Логарифм функции правдоподобия:

$$\ln\mathcal{L} = \sum_{i=1}^{n}\left[\ln\left(\theta(\ln X_i)^{\theta-1}\right) - \ln X_i\right] = \sum_{i=1}^{n}\left[\ln\theta + (\theta-1)\ln\ln X_i - \ln X_i\right] = n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln\ln X_i - \sum_{i=1}^{n}\ln X_i$$

$$\frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\ln\ln X_i$$

FOC:

$$\frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}\theta} = 0 \Rightarrow \frac{n}{\theta} = -\sum_{i=1}^{n}\ln\ln X_i \Rightarrow \hat{\theta}_{\mathrm{ML}} = -\frac{n}{\sum\limits_{i=1}^{n}\ln\ln X_i}$$

Подставим имеющиеся данные: $-\frac{n}{\sum\ln\ln X_i} = -\frac{100}{-30} = \frac{10}{3}$.

(b) (3 балла) Так как оценки ММП асимптотически нормальны, то для нахождения доверительного интервала достаточно найти стандартное отклонение параметра и домножить на квантиль двухстороннего распределения: $\mathbb{P}\left(\left\{|\hat{\theta}-\theta| \leqslant z_{0,025}\sqrt{\mathrm{Var}(\hat{\theta})}\right\}\right) = 0{,}95$. Известно, что $\widehat{\mathrm{Var}}(\hat{\theta}) = -\boldsymbol{H}^{-1}|_{\hat{\theta}}$. Матрица $\boldsymbol{H}$ — это матрица вторых производных логарифма функции правдоподобия.

$$\boldsymbol{H} = \frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\theta^2} = -\frac{n}{\theta^2} \Rightarrow -\boldsymbol{H}^{-1} = \frac{\theta^2}{n} \Rightarrow \widehat{\mathrm{Var}}(\hat{\theta}) = \frac{100/9}{100} = \frac{1}{9} \Rightarrow \hat{\sigma}_\theta = \frac{1}{3}$$

Следовательно, с вероятность 0,95 θ лежит в интервале $\frac{10}{3} \pm 1{,}96\cdot\frac{1}{3} \approx \frac{10}{3} \pm \frac{2}{3}$, или [2,680; 3,987].

(c) (3 × 3 = 9 баллов) Тест Вальда выглядит следующим образом:

$$W = \big(\mathbf{c}(\boldsymbol{\theta}) - \boldsymbol{q}\big)'(\boldsymbol{C}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{C}')^{-1}\big(\mathbf{c}(\boldsymbol{\theta}) - \boldsymbol{q}\big) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

За $\boldsymbol{C}$ обозначено $\frac{\mathbb{P}\mathbf{c}(\boldsymbol{\theta})}{\mathbb{P}\boldsymbol{\theta}}$, за $\boldsymbol{I}$ — информационная матрица Фишера ($\boldsymbol{I}(\theta) = -\mathbb{E}\left(\frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\boldsymbol{\theta}^2}\right)$). В данном случае $\boldsymbol{\theta} = \theta$, и нулевая гипотеза $\mathbf{c}(\theta) = \boldsymbol{q}$ выглядит как $\theta = 1$ ($\mathbf{c}(\theta) = \theta$) — одномерный случай, одна степень свободы хи-квадрата, $W \overset{\mathcal{H}_0}{\sim} \chi^2_1$. $\mathbf{c}'(\theta) = 1$, поэтому расчётная статистика выглядит следующим образом:

$$W = (\hat{\theta} - \theta_0)\frac{n}{\theta^2}(\hat{\theta} - \theta_0) = \left(\frac{10}{3} - 1\right)\cdot\frac{100}{100/9}\cdot\left(\frac{10}{3} - 1\right) = 49$$

Тест отношения правдоподобия:

$$LR = -2\big(\ln\mathcal{L}_{\mathrm{R}} - \ln\mathcal{L}_{\mathrm{UR}}\big) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

$$LR = -2\left(\left[n\ln\theta_0 + (\theta_0 - 1)\sum_{i=1}^{n}\ln\ln X_i - \cancel{\sum_{i=1}^{n}\ln X_i}\right] - \left[n\ln\hat{\theta} + (\hat{\theta} - 1)\sum_{i=1}^{n}\ln\ln X_i - \cancel{\sum_{i=1}^{n}\ln X_i}\right]\right) =$$

$$= -2\left(\cancel{100\ln 1} + \cancel{(1-1)(-30)} - 100\ln\frac{10}{3} - \left(\frac{10}{3} - 1\right)(-30)\right) = -2\left(-100\ln\frac{10}{3} + \frac{7}{3}\cdot 30\right) \approx 100{,}8$$

Тест множителей Лагранжа:

$$LM = \boldsymbol{S}(\theta_0)'\boldsymbol{I}^{-1}(\theta_0)\boldsymbol{S}(\theta_0) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

$\boldsymbol{S} = \frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}\theta}\big|_{\theta_0}$. В точке θ_0 значение частной производной логарифма функции правдоподобия равно $\frac{100}{1} - 30 = 70$, $\boldsymbol{I}^{-1}(\theta_0) = \frac{\theta_0^2}{n} = \frac{1}{100}$, откуда

$$LM = 70\cdot\frac{1}{100}\cdot 70 = 49$$

Для уровня значимости 5 % критическое значение χ^2_1 равно $\approx 3{,}84$, поэтому во всех трёх тестах гипотеза $\mathcal{H}_0\colon \theta = 1$ отвергается.

5.16.

$$\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \quad \mathcal{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right), \quad \ln\mathcal{L} = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\text{FOC:} \quad \frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu), \quad \frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\mu^2} = -\frac{n}{\sigma^2}, \quad \frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\mu\,\mathbb{P}(\sigma^2)} = -\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(X_i-\mu), \quad \frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(X_i-\mu)^2$$

Т. к. даны $\sum X_i$ и $\sum X_i^2$, то можно вывести, что $\sum (X_i-\mu)^2 = \sum X_i^2 - \sum 2\mu X_i + \sum \mu^2 = \sum X_i^2 - 2\mu\sum X_i + n\mu^2$.

Из условий первого порядка следует, что ММП-оценка матожидания $\hat{\mu}_{\text{ML}}$ — это выборочное среднее, а дисперсии $\hat{\sigma}^2_{\text{ML}}$ — выборочная дисперсия (без коррекции на одну степень свободы):

$$\hat{\mu}_{\text{ML}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{100}{100} = 1, \quad \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{100}(900 - 2\cdot 1\cdot 100 + 100\cdot 1^2) = \frac{800}{100} = 8$$

(b) (2 + 2 = 4 балла)

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\boldsymbol{\theta}^2}\right), \quad \boldsymbol{I}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2(\hat{\sigma}^2)^2} \end{pmatrix} = \begin{pmatrix} \frac{100}{8} & 0 \\ 0 & \frac{100}{128} \end{pmatrix}, \quad \boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{2}{25} & 0 \\ 0 & \frac{32}{25} \end{pmatrix}$$

Так как ММП-оценки асимптотически нормальны, то 95%-й доверительный интервал для вектора неизвестных параметров выглядит как

$$\begin{pmatrix} \hat{\mu} \pm z_{\frac{\alpha}{2}}\sqrt{\widehat{\text{Var}}(\hat{\mu})} \\ \hat{\sigma}^2 \pm z_{\frac{\alpha}{2}}\sqrt{\widehat{\text{Var}}(\hat{\mu})} \end{pmatrix} \approx \begin{pmatrix} 1 \pm 1{,}96\sqrt{\frac{2}{25}} \\ 8 \pm 1{,}96\sqrt{\frac{32}{25}} \end{pmatrix} \approx \begin{pmatrix} [0{,}446; 1{,}554] \\ [5{,}783; 10{,}217] \end{pmatrix}$$

(c) (3 × 3 = 9 баллов)

Тест Вальда:

$$W = \big(c(\sigma^2) - \sigma_0^2\big)'(\boldsymbol{C}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{C}')^{-1}\big(c(\sigma^2) - \sigma_0^2\big) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

$\frac{\mathbb{P}c}{\mathbb{P}\sigma^2} = 1$, поэтому

$$W = \big(8-1\big)^2\frac{n}{2(\sigma^2)^2} = 49\cdot\frac{100}{128} \approx 38{,}28$$

Тест отношения правдоподобия:

$$LR = -2\big(\ln\mathcal{L}_{\text{R}} - \ln\mathcal{L}_{\text{UR}}\big) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

$$LR = -2\big(\ln\mathcal{L}(\sigma_0^2) - \ln\mathcal{L}(\hat{\sigma}^2)\big) = -2\left(-\frac{n}{2}\ln\sigma_0^2 - \frac{1}{2\sigma_0^2}\cdot 800 + \frac{n}{2}\ln\hat{\sigma}^2 + \frac{1}{2\hat{\sigma}^2}\cdot 800\right) =$$

$$-2\left(\cancel{-50\ln 1} - \frac{1}{2}\cdot 800 + 50\ln 8 + \frac{1}{16}\cdot 800\right) \approx 492$$

Тест множителей Лагранжа:

$$LM = \boldsymbol{S}(\sigma_0^2)'\boldsymbol{I}^{-1}(\sigma_0^2)\boldsymbol{S}(\sigma_0^2) \overset{\mathcal{H}_0}{\sim} \chi^2_r$$

$$\boldsymbol{I}(\sigma_0^2) = \frac{n}{2(\sigma_0^2)^2} = 50, \quad \boldsymbol{S}(\sigma_0^2) = \left.\frac{\mathbb{P}\ln\mathcal{L}}{\mathbb{P}(\sigma^2)}\right|_{\sigma_0^2} = -\frac{100}{2} + \frac{1}{2}\cdot 800 = 350$$

$$LM = 350^2 \cdot \frac{1}{50} = 2450$$

Для уровня значимости 5 % критическое значение χ_1^2 равно $\approx 3{,}84$, поэтому во всех трёх тестах гипотеза $\mathcal{H}_0\colon \sigma^2 = 1$ отвергается.

(d) (3 × 3 = 9 баллов)

Тест Вальда выглядит следующим образом:

$$W = \big(\mathbf{c}(\hat{\boldsymbol{\theta}}) - \boldsymbol{q}\big)'(\boldsymbol{C}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{C}')^{-1}\big(\mathbf{c}(\hat{\boldsymbol{\theta}}) - \boldsymbol{q}\big) \overset{\mathcal{H}_0}{\sim} \chi_r^2$$

За $\boldsymbol{C}$ обозначено $\frac{\mathbb{P}\mathbf{c}(\boldsymbol{\theta})}{\mathbb{P}\boldsymbol{\theta}}$, за $\boldsymbol{I}$ — информационная матрица Фишера ($\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\mathbb{P}^2\ln\mathcal{L}}{\mathbb{P}\boldsymbol{\theta}^2}\right)$). В данном случае нулевая гипотеза $\mathbf{c}(\boldsymbol{\theta}) = \boldsymbol{q}$ записывается как $\mathbf{c}(\boldsymbol{\theta}) = \begin{pmatrix}\mu\\ \sigma^2\end{pmatrix} = \begin{pmatrix}2\\ 1\end{pmatrix}$, поэтому все статистики имеют две степени свободы хи-квадрата. $\boldsymbol{C} = \frac{\mathbb{P}\mathbf{c}}{\mathbb{P}\boldsymbol{\theta}} = \begin{pmatrix}\frac{\mathbb{P}c_1}{\mathbb{P}\mu} & \frac{\mathbb{P}c_2}{\mathbb{P}\mu}\\ \frac{\mathbb{P}c_1}{\mathbb{P}\sigma^2} & \frac{\mathbb{P}c_2}{\mathbb{P}\sigma^2}\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, $\mathbf{c}(\boldsymbol{\theta}) - \boldsymbol{q} = \begin{pmatrix}1\\ 8\end{pmatrix} - \begin{pmatrix}2\\ 1\end{pmatrix} = \begin{pmatrix}-1\\ 7\end{pmatrix}$, поэтому расчётная статистика выглядит следующим образом:

$$W = \begin{pmatrix}-1 & 7\end{pmatrix}\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\frac{100}{8} & 0\\ 0 & \frac{100}{128}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right]\begin{pmatrix}-1\\ 7\end{pmatrix} = \begin{pmatrix}-1 & 7\end{pmatrix}\begin{pmatrix}\frac{25}{2} & 0\\ 0 & \frac{25}{32}\end{pmatrix}\begin{pmatrix}-1\\ 7\end{pmatrix} = 50{,}78$$

Тест отношения правдоподобия:

$$LR = -2\big(\ln\mathcal{L}_{\mathrm{R}} - \ln\mathcal{L}_{\mathrm{UR}}\big) \overset{\mathcal{H}_0}{\sim} \chi_r^2$$

$$\begin{aligned}LR &= -2\big(\ln\mathcal{L}(\boldsymbol{q}) - \ln\mathcal{L}(\hat{\boldsymbol{\theta}})\big) = \\ &= -2\left(-\frac{n}{2}\ln\sigma_0^2 - \frac{1}{2\sigma_0^2}\left(\sum X_i^2 - 2\mu_0\sum X_i + n\mu_0^2\right) + \frac{n}{2}\ln\hat{\sigma}^2 + \frac{1}{2\hat{\sigma}^2}\left(\sum X_i^2 - 2\hat{\mu}\sum X_i + n\hat{\mu}^2\right)\right) = \\ &= -2\left(\cancel{-\frac{100}{2}\ln 1} - \frac{1}{2}(900\cancel{-2\cdot 2\cdot 100 + 100\cdot 2^2}) + \frac{100}{2}\ln 8 + \frac{1}{16}(900 - 2\cdot 1\cdot 100 + 100\cdot 1)\right) \approx 592\end{aligned}$$

Тест множителей Лагранжа:

$$LM = \boldsymbol{S}(\boldsymbol{\theta}_0)'\boldsymbol{I}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{S}(\boldsymbol{\theta}_0) \overset{\mathcal{H}_0}{\sim} \chi_r^2$$

$$\boldsymbol{I}(\boldsymbol{\theta}_0) = \begin{pmatrix}\frac{n}{\sigma_0^2} & 0\\ 0 & \frac{n}{2(\sigma_0^2)^2}\end{pmatrix} = \begin{pmatrix}100 & 0\\ 0 & 50\end{pmatrix}, \quad \boldsymbol{I}^{-1}(\boldsymbol{\theta}_0) = \begin{pmatrix}\frac{1}{100} & 0\\ 0 & \frac{1}{50}\end{pmatrix}$$

$$\boldsymbol{S}(\boldsymbol{\theta}_0) = \begin{pmatrix}\frac{1}{\sigma_0^2}(100 - 100\mu_0)\\ -\frac{100}{2\sigma_0^2} + \frac{1}{2(\sigma_0^2)^2}(900 - 200\mu_0 + 100\mu_0^2)\end{pmatrix} = \begin{pmatrix}-100\\ 400\end{pmatrix}$$

$$LM = \begin{pmatrix}-100 & 400\end{pmatrix}\begin{pmatrix}\frac{1}{100} & 0\\ 0 & \frac{1}{50}\end{pmatrix}\begin{pmatrix}-100\\ 400\end{pmatrix} = 3300$$

Для уровня значимости 5 % критическое значение χ^2_2 равно $\approx 5{,}99$, поэтому во всех трёх тестах гипотеза $\mathcal{H}_0\colon \boldsymbol{\theta} = \boldsymbol{\theta}_0$ отвергается.

5.17. Можно решать перебором вариантов.

6.1. $f(x)$ чётная, $\mathbb{E}(X) = 0$, $\mathrm{Var}(X) = \pi^2/3$, логистическое похоже на $N(0, \pi^2/3)$

6.2. $\ln\left(\frac{\mathbb{P}(y_i=1)}{\mathbb{P}(y_i=0)}\right) = \beta_1 + \beta_2 x_i$.

6.3.

6.4.

6.5.

6.6. Для краткости введем следующие обозначения: $y_i = honey_i$, $d_i = bee_i$[1].

1. Функция правдоподобия имеет следующий вид:

$$\begin{aligned}
\mathrm{L}(\beta_1,\beta_2) = \prod_{i=1}^{n} \mathbb{P}_{\beta_1,\beta_2}\left(\{Y_i = y_i\}\right) = \\
\prod_{i:y_i=0} \mathbb{P}_{\beta_1,\beta_2}\left(\{Y_i = 1\}\right) \cdot \prod_{i:y_i=1} \mathbb{P}_{\beta_1,\beta_2}\left(\{Y_i = 0\}\right) = \\
\prod_{i:y_i=1} \Lambda(\beta_1 + \beta_2 d_i) \cdot \prod_{i:y_i=0} [1 - \Lambda(\beta_1 + \beta_2 d_i)] = \\
\prod_{i:y_i=1,d_i=1} \Lambda(\beta_1 + \beta_2) \cdot \prod_{i:y_i=1,d_i=0} \Lambda(\beta_1) \cdot \prod_{i:y_i=0,d_i=1} [1 - \Lambda(\beta_1 + \beta_2)] \cdot \\
\cdot \prod_{i:y_i=0,d_i=0} [1 - \Lambda(\beta_1)] = \\
\Lambda(\beta_1+\beta_2)^{\#\{i:y_i=1,d_i=1\}} \cdot \Lambda(\beta_1)^{\#\{i:y_i=1,d_i=0\}} \cdot [1 - \Lambda(\beta_1+\beta_2)]^{\#\{i:y_i=0,d_i=1\}} \cdot \\
\cdot [1 - \Lambda(\beta_1)]^{\#\{i:y_i=0,d_i=0\}}
\end{aligned} \tag{1.33}$$

где

$$\Lambda(x) = \frac{e^x}{1 + e^x} \tag{1.34}$$

логистическая функция распределения, $\#A$ означает число элементов множества A.

2. Введём следующие обозначения:

$$a := \Lambda(\beta_1) \tag{1.35}$$

$$b := \Lambda(\beta_1 + \beta_2) \tag{1.36}$$

Тогда с учетом имеющихся наблюдений функция правдоподобия принимает вид:

$$\mathrm{L}(a,b) = b^{12} \cdot a^{32} \cdot [1-b]^{36} \cdot [1-a]^{20}$$

Логарифмическая функция правдоподобия:

$$l(a,b) = \ln \mathrm{L}(a,b) = 12\ln b + 32\ln a + 36\ln[1-b] + 20\ln[1-a]$$

Решая систему уравнений правдоподобия

$$\begin{cases} \frac{\partial l}{\partial a} = \frac{32}{a} - \frac{20}{1-a} = 0 \\ \frac{\partial l}{\partial b} = \frac{12}{b} - \frac{36}{1-b} = 0 \end{cases}$$

[1] Y_i — случайный Мёд, y_i — реализация случайного Мёда (наблюдаемый Мёд)

получаем $\hat{a} = \frac{8}{13}$, $\hat{b} = \frac{1}{4}$. Из формул (1.34) и (1.35), находим $\hat{\beta}_{1,UR} = \ln\left(\frac{\hat{a}}{1-\hat{a}}\right) = \ln\left(\frac{8}{5}\right) = 0.47$. Далее, из (1.34) и (1.36) имеем $\hat{\beta}_{1,UR}+\hat{\beta}_{2,UR} = \ln\left(\frac{\hat{b}}{1-\hat{b}}\right)$. Следовательно, $\hat{\beta}_{2,UR} = \ln\left(\frac{\hat{b}}{1-\hat{b}}\right) - \hat{\beta}_{1,UR} = \ln\left(\frac{1}{3}\right) - \ln\left(\frac{8}{5}\right) = -1.57$.

3. Гипотеза, состоящая в том, что «правильность Мёда не связана с правильностью пчёл» формализуется как $H_0 : \beta_2 = 0$. Протестируем данную гипотезу при помощи теста отношения правдоподобия. Положим в функции правдоподобия $\mathrm{L}(\beta_1, \beta_2)$ $\beta_2 = 0$. Тогда с учетом (1.35) и (1.36) получим

$$\mathrm{L}(a, b = a) = a^{32+12} \cdot [1-a]^{20+36}$$

В этом случае логарифмическая функция правдоподобия имеет вид:

$$l(a, b = a) := \mathrm{L}(a, b = a) = 44 \ln a + 56 \ln[1-a]$$

Решаем уравнение правдоподобия

$$\frac{\partial l}{\partial a} = \frac{44}{a} - \frac{56}{1-a} = 0$$

и получаем $\hat{a} = \frac{11}{25}$. Следовательно, согласно (1.34) и (1.35), $\hat{\beta}_{1,R} = -0.24$ и $\hat{\beta}_{2,R} = 0$.

Статистика отношения правдоподобия имеет вид:

$$LR = -2(l(\hat{\beta}_{1,R}, \hat{\beta}_{2,R}) - l(\hat{\beta}_{1,UR}, \hat{\beta}_{2,UR}))$$

и имеет асимптотическое χ^2 распределение с числом степеней свободы, равным числу ограничений, составляющих гипотезу H_0, т.е. в данном случае $LR \overset{a}{\sim} \chi^2_1$.

Находим наблюдаемое значение статистики отношения правдоподобия:

$$l(\hat{\beta}_{1,R}, \hat{\beta}_{2,R}) = l(\hat{a}_R, \hat{b}_R = \hat{a}_R) = 44 \ln \hat{a}_R + 56 \ln[1-\hat{a}_R] = 44 \ln\left[\frac{11}{25}\right] + 56 \ln\left[1 - \frac{11}{25}\right] = -68.59 \quad (1.37)$$

$$l(\hat{\beta}_{1,UR}, \hat{\beta}_{2,UR}) = l(\hat{a}_{UR}, \hat{b}_{UR}) = 12 \ln \hat{b}_{UR} + 32 \ln \hat{a}_{UR} + 36 \ln[1-\hat{b}_{UR}] + 20 \ln[1-\hat{a}_{UR}] = 12 \ln\left[\frac{1}{4}\right] + 32 \ln\left[\frac{8}{13}\right] + 36 \ln\left[1-\frac{1}{4}\right] + 20 \ln\left[1-\frac{8}{13}\right] = -61.63 \quad (1.38)$$

Следовательно, $LR_{\text{набл}} = -2(-68.59 + 61.63) = 13.92$, при этом критическое значение χ^2 распределения с одной степенью свободы для 5% уровня значимости равна 3.84. Значит, на основании теста отношения правдоподобия гипотеза $H_0 : \beta_2 = 0$ должна быть отвергнута. Таким образом, данные показывают, что, в действительности, правильность мёда связана с правильностью пчёл.

4.

$$\hat{\mathbb{P}}\{honey = 0 | bee = 0\} = 1 - \hat{\mathbb{P}}\{honey = 1 | bee = 0\} = 1 - \frac{\exp\{\hat{\beta}_{1,UR} + \hat{\beta}_{2,UR} \cdot 0\}}{1 + \exp\{\hat{\beta}_{1,UR} + \hat{\beta}_{2,UR} \cdot 0\}} = 1 - \frac{\exp\{\ln\left(\frac{8}{5}\right)\}}{1 + \exp\{\ln\left(\frac{8}{5}\right)\}} = 1 - 0.62 = 0.38 \quad (1.39)$$

6.7. в теории оценки не существуют, в R получатся некие точечные оценки, достаточно далеко лежащие от нуля с огромными стандартными ошибками и P-значением близким к 1.

7.1.

7.2. увеличить количество наблюдений, уменьшить дисперсию случайной ошибки

7.3.

7.4.

7.5.

7.6.

7.7.

7.8. $r^* = -1/2$

7.9. $r^* = -1/3$

8.1.

8.2. Поделить зависимую переменную и каждый регрессор, включая единичный столбец, на $|x_i|$.

8.3. Поделить зависимую переменную и каждый регрессор, включая единичный столбец, на $\sqrt{|x_i|}$.

8.4. $\mathrm{Var}(\varepsilon_i) = cx_i^4$

8.5. $\mathrm{Var}(\varepsilon_i) = cx_i$

8.6. По графику видно, что с увеличением общей площади увеличивается разброс цены. Поэтому разумно, например, рассмотреть следующие подходы:

1. Перейти к логарифмам, т.е. оценивать модель $\ln price_i = \beta_1 + \beta_2 \ln totsp_i + \varepsilon_i$

2. Оценивать квантильную регрессию. В ней угловые коэффициенты линейной зависимости будут отличаться для разных квантилей переменной $price$.

3. Обычную модель линейной регрессии с гетероскедастичностью вида $Var(\varepsilon_i) = \sigma^2 totsp_i^2$

8.7.

8.8. Протестируем гетероскедастичность ошибок при помощи теста Голдфельда- Квандта. H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2$, H_a : $\mathrm{Var}(\varepsilon_i) = f(x_i)$

1. Тестовая статистика $GQ = \frac{RSS_3/(n_3-k)}{RSS_1/(n_1-k)}$, где $n_1 = 11$ — число наблюдений в первой подгруппе, $n_3 = 11$ — число наблюдений в последней подгруппе, $k = 3$ — число факторов в модели, считая единичный столбец.

2. Распределение тестовой статистики при верной H_0: $GQ \sim F_{n_3-k,n_1-k}$

3. Наблюдаемое значение $GQ_{obs} = 1.41$

4. Область, в которой H_0 не отвергается: $GQ \in [0; 3.44]$

5. Статистический вывод: поскольку $GQ_{obs} \in [0; 3.44]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 не может быть отвергнута. Таким образом, тест Голдфельда-Квандта не выявил гетероскедастичность.

8.9. Протестируем гетероскедастичность ошибок при помощи теста Голдфельда- Квандта. H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2$, H_a : $\mathrm{Var}(\varepsilon_i) = f(x_i)$

1. Тестовая статистика $GQ = \frac{RSS_3/(n_3-k)}{RSS_1/(n_1-k)}$, где $n_1 = 21$ — число наблюдений в первой подгруппе, $n_3 = 21$ — число наблюдений в последней подгруппе, $k = 3$ — число факторов в модели, считая единичный столбец.

2. Распределение тестовой статистики при верной H_0: $GQ \sim F_{n_3-k,n_1-k}$

3. Наблюдаемое значение $GQ_{obs} = 6.49$

4. Область, в которой H_0 не отвергается: $GQ \in [0; 3.12]$

5. Статистический вывод: поскольку $GQ_{obs} \notin [0; 3.12]$, то на основании имеющихся наблюдений на уровне значимости 1% основная гипотеза H_0 отвергается. Таким образом, тест Голдфельда-Квандта выявил гетероскедастичность.

8.10. Протестируем гетероскедастичность ошибок при помощи теста Голдфельда- Квандта. H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2$, H_a : $\mathrm{Var}(\varepsilon_i) = f(x_i)$

1. Тестовая статистика $GQ = \frac{RSS_3/(n_3-k)}{RSS_1/(n_1-k)}$, где $n_1 = 11$ — число наблюдений в первой подгруппе, $n_3 = 11$ — число наблюдений в последней подгруппе, $k = 3$ — число факторов в модели, считая единичный столбец.

2. Распределение тестовой статистики при верной H_0: $GQ \sim F_{n_3-k,n_1-k}$

3. Наблюдаемое значение $GQ_{obs} = 2.88$

4. Область, в которой H_0 не отвергается: $GQ \in [0; 3.44]$

5. Статистический вывод: поскольку $GQ_{obs} \in [0; 3.44]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 не может быть отвергнута. Таким образом, тест Голдфельда-Квандта не выявил гетероскедастичность.

8.11. Протестируем гетероскедастичность ошибок при помощи теста Голдфельда- Квандта. H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2$, H_a : $\mathrm{Var}(\varepsilon_i) = f(x_i)$

1. Тестовая статистика $GQ = \frac{RSS_3/(n_3-k)}{RSS_1/(n_1-k)}$, где $n_1 = 21$ — число наблюдений в первой подгруппе, $n_3 = 21$ — число наблюдений в последней подгруппе, $k = 3$ — число факторов в модели, считая единичный столбец.

2. Распределение тестовой статистики при верной H_0: $GQ \sim F_{n_3-k,n_1-k}$

3. Наблюдаемое значение $GQ_{obs} = 5.91$

4. Область, в которой H_0 не отвергается: $GQ \in [0; 2.21]$

5. Статистический вывод: поскольку $GQ_{obs} \notin [0; 2.21]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 отвергается. Таким образом, тест Голдфельда-Квандта выявил гетероскедастичность.

8.12. Протестируем гетероскедастичность ошибок при помощи теста Уайта. H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2$, H_a : $\mathrm{Var}(\varepsilon_i) = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i$.

1. Тестовая статистика $W = n \cdot R^2_{aux}$, где n — число наблюдений, R^2_{aux} — коэффициент детерминации для вспомогательной регрессии.

2. Распределение тестовой статистики при верной H_0: $W \sim \chi^2_{k_{aux}-1}$, где $k_{aux} = 6$ — число регрессоров во вспомогательной регрессии, считая константу.

3. Наблюдаемое значение тестовой статистики: $W_{obs} = 18$

4. Область, в которой H_0 не отвергается: $W \in [0; W_{crit}] = [0; 11.07]$

5. Статистический вывод: поскольку $W_{obs} \notin [0; 11.07]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 отвергается. Таким образом, тест Уайта выявил гетероскедастичность.

8.13.
8.14.
8.15.
8.16.
8.17.
8.18.
8.19.
8.20.
8.21.
8.22.
8.23. 0.0752, 5, 10
8.24. $k(k+1)/2$
8.25.
8.26. Известно, что оценки параметров, получаемые по обобщённому методу наименьших квадратов, являются наилучшими, поэтому: $\delta^2 \begin{bmatrix} x_1 & 0\ldots0 & 0 \\ 0 & & 0 \\ & \ldots & \\ 0 & & 0 \\ 0 & 0\ldots0 & x_n \end{bmatrix}$

8.27.

$$\begin{aligned}\operatorname{Cov}(\hat{\beta}_{GLS}, \varepsilon) = \operatorname{Cov}\left((X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y, \varepsilon\right) = \\ \operatorname{Cov}\left((X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\varepsilon, \varepsilon\right) = \\ = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\operatorname{Cov}(\varepsilon, \varepsilon) = \\ (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\Sigma = (X^T\Sigma^{-1}X)^{-1}X^T\end{aligned} \tag{1.40}$$

8.28. Для нахождения эффективной оценки воспользуемся взвешенным методом наименьших квадратов. Разделим каждое из уравнений $y_i = \beta_1 + \varepsilon$ на корень из дисперсии ε_i с тем, чтобы ошибки в полученных уравнениях имели равные дисперсии (в этом случае можно будет сослаться на т. Гаусса-Маркова). Итак, после деления i-го уравнения на величину $\sqrt{x_i}/\sigma_\varepsilon$, мы получаем:

$$\begin{bmatrix} y_1\sqrt{x_1}/\sigma_\varepsilon \\ y_2\sqrt{x_2}/\sigma_\varepsilon \\ \ldots \\ y_n\sqrt{x_n}/\sigma_\varepsilon \end{bmatrix} = \beta_1 \begin{bmatrix} \sqrt{x_1}/\sigma_\varepsilon \\ \sqrt{x_2}/\sigma_\varepsilon \\ \ldots \\ \sqrt{x_n}/\sigma_\varepsilon \end{bmatrix} + \begin{bmatrix} \varepsilon_1\sqrt{x_1}/\sigma_\varepsilon \\ \varepsilon_2\sqrt{x_2}/\sigma_\varepsilon \\ \ldots \\ \varepsilon_n\sqrt{x_n}/\sigma_\varepsilon \end{bmatrix}$$

Поскольку условия т. Гаусса-Маркова для последней модели выполнены, то МНК-оценка для последней модели будет наиболее эффективной. Поэтому

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i\sqrt{x_i}/\sigma_\varepsilon)(\sqrt{x_i}/\sigma_\varepsilon)}{\sum_{i=1}^n (\sqrt{x_1}/\sigma_\varepsilon)} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$$

8.29.
8.30. В предположении о гомоскедастичности, $\gamma_2 = 0$, оценка правдоподобия совпадает с МНК-оценкой, значит $\hat{\beta} = \sum y_i x_i / \sum x_i^2$. И $\hat{\sigma}_i^2 = RSS/n$, значит $\hat{\gamma}_1 = \ln(RSS/n)$.
8.31. Решение средствами пакета `sandwich`

```
df <- data.frame(y=c(1,3,3), x=c(0,0,1))

model <- lm(data=df, y~x)
coef(model)

## (Intercept)           x
##           2           1
```

```
# residuals
resid(model)
```

```
##             1             2             3
## -1.000000e+00  1.000000e+00  2.220446e-16
```

```
vcov(model)
```

```
##             (Intercept)  x
## (Intercept)           1 -1
## x                    -1  3
```

```
vcovHC(model) # should fail
```

```
##             (Intercept)   x
## (Intercept)         NaN NaN
## x                   NaN NaN
```

```
vcovHC(model, type ="HC0" )
```

```
##             (Intercept)    x
## (Intercept)         0.5 -0.5
## x                  -0.5  0.5
```

```
# help(vcovHC)
```

Решение с ручным подсчётом матриц

```
y <- c(1,3,3)
X <- cbind(rep(1,3), c(0,0,1))

hat_beta <- solve(t(X) %*% X) %*% t(X) %*% y
hat_beta
```

```
##      [,1]
## [1,]    2
## [2,]    1
```

```
# by hand Var(hat_beta)
y_hat <- X %*% hat_beta
e_hat <- y-y_hat
RSS <- sum((y-y_hat)^2)
vcov_ols <- RSS/(3-2)*solve(t(X) %*% X)
vcov_ols
```

```
##      [,1] [,2]
## [1,]    1   -1
## [2,]   -1    3
```

```
# crossprod(X) is just synonym fort(X) %*% X

H <- X %*% solve(crossprod(X)) %*% t(X)
H
```

```
##              [,1]         [,2]         [,3]
## [1,] 5.000000e-01 5.000000e-01 1.110223e-16
## [2,] 5.000000e-01 5.000000e-01 1.110223e-16
## [3,] 1.110223e-16 1.110223e-16 1.000000e+00
```

```
diag(H)
```

```
## [1] 0.5 0.5 1.0
```

```
S_hat_white <- diag(as.vector(e_hat^2))
S_hat_HC3 <- diag(as.vector(e_hat^2)/(1-diag(H))^2)
S_hat_HC3 # look at the problem
```

```
##      [,1] [,2] [,3]
## [1,]    4    0    0
```

```
## [2,]    0    4    0
## [3,]    0    0  Inf
# vcov White
solve(crossprod(X)) %*% t(X)  %*% S_hat_white %*% X %*% solve(crossprod(X))
##      [,1] [,2]
## [1,]  0.5 -0.5
## [2,] -0.5  0.5
# vcov HC3
solve(crossprod(X)) %*% t(X)  %*% S_hat_HC3 %*% X %*% solve(crossprod(X))
##      [,1] [,2]
## [1,]  NaN  NaN
## [2,]  NaN  NaN
```

8.32.

8.33. при гомоскедастичности $\hat{\mu} = \bar{y}$, при гетероскедастичности

$$\hat{\mu} = \frac{\sum \tilde{x}_i \tilde{y}_i}{\sum \tilde{x}_i^2} = \frac{\sum i^2 \cdot y_i}{\sum i^2}$$

9.1.

9.2.

9.3.

9.4.

9.5.

9.6.

9.7. чтобы избежать переполнения при подсчете произведения всех y_i

9.8.

9.9.

10.1. $u_i^2 = \varepsilon_i^2 = 1$, $\mathbb{E}(\hat{\beta} \mid x_1 = 0,\, x_2 = 1) = 0.2\beta$, $\mathbb{E}(\hat{\beta} \mid x_1 = 0,\, x_2 = 2) = 0.8\beta$. Интуитивно объясняем: рисуем прямую по двум точкам. Мы знаем абсциссы точек с точностью ± 1. Если точки близки, то это может сильно менять оценку наклона, если точки далеки, то случайность слабо влияет на наклон.

10.2.

10.3.

10.4.

10.5.

10.6.

10.7.

10.8.

11.1.

11.2.

1. Процесс $AR(2)$, т.к. две первые частные корреляции значимо отличаются от нуля, а гипотезы о том, что каждая последующая равна нулю не отвергаются.

2. Можно использовать одну из двух статистик

$$\text{Ljung-Box} = n(n+2)\sum_{k=1}^{3} \frac{\hat{\rho}_k^2}{n-k} = 0.4288623$$

$$\text{Box-Pierce} = n\sum_{k=1}^{3} \hat{\rho}_k^2 = 0.4076341$$

Критическое значение хи-квадрат распределения с 3-мя степенями свободы для $\alpha = 0.05$ равно $\chi^2_{3,crit} = 7.8147279$. Вывод: гипотеза H_0 об отсутствии корреляции ошибок модели не отвергается.

11.3.

1. H_0: ряд содержит единичный корень, $\beta = 0$; H_0: ряд не содержит единичного корня, $\beta < 0$
2. $ADF = -0.4/0.1 = -4$, $ADF_{crit} = -2.89$, H_0 отвергается
3. Ряд стационарен
4. При верной H_0 ряд не стационарен, и t-статистика имеет не t-распределение, а распределение Дики-Фуллера.

11.4.
11.5.
11.6.
11.7.
11.8.
11.9.
11.10.

1. $\mathbb{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma^2/(1-\rho^2)$
2. $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \rho^h \cdot \sigma^2/(1-\rho^2)$
3. $\mathrm{Corr}(\varepsilon_t, \varepsilon_{t+h}) = \rho^h$

11.11.
11.12. все линейные комбинации стационарны
11.13. Они будут примерно одинаковы. Оценка наклона определяется автоковариационной функцией.
11.14.
11.15. $x_t = (1-\mathrm{L})^t y_t$
11.16. $F_n = \mathrm{L}(1+\mathrm{L})F_n$, значит $F_n = \mathrm{L}^k(1+\mathrm{L})^k F_n$ или $F_{n+k} = (1+\mathrm{L})^k F_n$
11.17. а - неверно, б - верно.
11.18.
11.19.
11.20.
11.21.
11.22.
11.23. 1, 2, 2
11.24.
11.25.
11.26.
11.27.
11.28.
11.29. 1. Поскольку имеют место соотношения $\varepsilon_1 = \rho\varepsilon_0 + u_1$ и $Y_1 = \mu + \varepsilon_1$, то из условия задачи получаем, что $\varepsilon_1 \sim N(0, \sigma^2/(1-\rho^2))$ и $Y_1 \sim N(\mu, \sigma^2/(1-\rho^2))$. Поэтому

$$f_{Y_1}(y_1) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\rho^2)}} \exp\left(-\frac{(y_1-\mu)^2}{2\sigma^2/(1-\rho^2)}\right).$$

Далее, найдем $f_{Y_2|Y_1}(y_2|y_1)$. Учитывая, что $Y_2 = \rho Y_1 + (1-\rho)\mu + u_2$, получаем $Y_2|\{Y_1 = y_1\} \sim N(\rho y_1 + (1-\rho)\mu, \sigma^2)$. Значит,

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_2 - \rho y_1 - (1-\rho)\mu)^2}{2\sigma^2}\right).$$

Действуя аналогично, получаем, что для всех $t \geqslant 2$ справедлива формула

$$f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \rho y_{t-1} - (1-\rho)\mu)^2}{2\sigma^2}\right).$$

Таким образом, находим функцию правдоподобия

$$\mathrm{L}(\mu,\rho,\sigma^2)=f_{Y_T,\ldots,Y_1}(y_T,\ldots,y_1)=f_{Y_1}(y_1)\prod_{t=2}^{T}f_{Y_t|Y_{t-1}}(y_t|y_{t-1}),$$

где $f_{Y_1}(y_1)$ и $f_{Y_t|Y_{t-1}}(y_t|y_{t-1})$ получены выше.

2. Для нахождения неизвестных параметров модели запишем логарифмическую условную функцию правдоподобия:

$$l(\mu,\rho,\sigma^2|Y_1=y_1)=\sum_{t=2}^{T}\log f_{Y_t|Y_{t-1}}(y_t|y_{t-1})=$$

$$=-\frac{T-1}{2}\log(2\pi)-\frac{T-1}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t-\rho y_{t-1}-(1-\rho)\mu)^2.$$

Найдем производные функции $l(\mu,\rho,\sigma^2|Y_1=y_1)$ по неизвестным параметрам:

$$\frac{\partial l}{\partial\mu}=-\frac{1}{2\sigma^2}\sum_{t=2}^{T}2(y_t-\rho y_{t-1}-(1-\rho)\mu)\cdot(\rho-1),$$

$$\frac{\partial l}{\partial\rho}=-\frac{1}{2\sigma^2}\sum_{t=2}^{T}2(y_t-\rho y_{t-1}-(1-\rho)\mu)\cdot(\mu-y_{t-1}),$$

$$\frac{\partial l}{\partial\sigma^2}=-\frac{T-1}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{t=2}^{T}(y_t-\rho y_{t-1}-(1-\rho)\mu)^2.$$

Оценки неизвестных параметров модели могут быть получены как решение следующей системы уравнений:

$$\begin{cases}\dfrac{\partial l}{\partial\mu}=0,\\ \dfrac{\partial l}{\partial\rho}=0,\\ \dfrac{\partial l}{\partial\sigma^2}=0.\end{cases}$$

Из первого уравнения системы получаем, что

$$\sum_{t=2}^{T}y_t-\hat\rho\sum_{t=2}^{T}y_{t-1}=(T-1)(1-\hat\rho)\hat\mu,$$

откуда

$$\hat\mu=\frac{\sum_{t=2}^{T}y_t-\hat\rho\sum_{t=2}^{T}y_{t-1}}{(T-1)(1-\hat\rho)}=\frac{3-\hat\rho\cdot 3}{4\cdot(1-\hat\rho)}=\frac{3}{4}.$$

Далее, если второе уравнение системы переписать в виде

$$\sum_{t=2}^{T}(y_t-\hat\mu-\hat\rho(y_{t-1}-\hat\mu))(y_{t-1}-\hat\mu)=0,$$

то легко видеть, что

$$\hat\rho=\frac{\sum_{t=2}^{T}(y_t-\hat\mu)(y_{t-1}-\hat\mu)}{\sum_{t=2}^{T}(y_{t-1}-\hat\mu)^2}.$$

Следовательно, $\hat\rho=-1/11=-0.0909$.

Наконец, из третьего уравнения системы

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - \hat{\rho}y_{t-1} - (1-\hat{\rho})\hat{\mu})^2.$$

Значит, $\hat{\sigma}^2 = 165/242 = 0.6818$. Ответы: $\hat{\mu} = 3/4 = 0.75$, $\hat{\rho} = -1/11 = -0.0909$, $\hat{\sigma}^2 = 165/242 = 0.6818$.

11.30. Рассмотрим модель без константы. Тогда ковариационная матрица коэффициентов пропорциональна матрице

$$\begin{pmatrix} 1 & -\hat{\rho}_1 \\ -\hat{\rho}_1 & 1 \end{pmatrix}$$

11.31.

11.32.

11.33.

11.34.

11.35.

11.36.

11.37.

11.38. Процесс стационарен только при $y_1 = 4 + \frac{2}{\sqrt{3}}\varepsilon_1$. Фразу нужно понимать как «у стохастического разностного уравнения $y_t = 2 + 0.5y_{t-1} + \varepsilon_t$ есть стационарное решение».

11.39.

1. $\mathbb{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_1) = \sigma^2$, $\mathrm{Var}(\varepsilon_t) = 2\sigma^2$ при $t \geqslant 2$. Гетероскедастичная.

2. $\mathrm{Cov}(e_t, e_{t+1}) = \sigma^2$. Автокоррелированная.

3. $\hat{\beta}$ — несмещенная, неэффективная

4. Более эффективной будет $\hat{\beta}_{gls} = (X'V^{-1}X)^{-1}X'V^{-1}y$, где

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Матрица V известна с точностью до константы σ^2, но в формуле для $\hat{\beta}_{gls}$ неизвестная σ^2 сократится.

Другой способ построить эффективную оценку — применить МНК к преобразованным наблюдениям, т.е. $\hat{\beta}_{gls} = \frac{\sum x_i' y_i'}{\sum x_i'^2}$, где $y_1' = y_1$, $x_1' = x_1$, $y_t' = y_t - y_{t-1}$, $x_t' = x_t - x_{t-1}$ при $t \geqslant 2$.

12.1.
12.2.
12.3. $f(x_1, x_2) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2)$
12.4.
12.5.
13.1.
13.2.
13.3.
14.1.
14.2.
14.3.
14.4.
14.5.
14.6.
14.7.
14.8.
14.9.
14.10.
14.11.
14.12.
14.13.
14.14. Например, $A = (1, 2, 3)$, $B = (1, 0, 1)'$
14.15. $\mathrm{tr}(I) = n$, $\mathrm{tr}(\pi) = 1$, $\mathrm{tr}(P) = k$
14.16.
14.17.
14.18. $n \times m$, $m \times n$, I
14.19.
14.20.
14.21.
15.1.

15.2.

15.3.

15.4.

15.5.

15.6.

15.7.

15.8.

15.9. По определению ковариационной матрицы:

$$\mathrm{Var}(\xi) = \begin{pmatrix} \mathrm{Var}(\xi_1) & \mathrm{Cov}(\xi_1,\xi_2) & \mathrm{Cov}(\xi_1,\xi_3) \\ \mathrm{Cov}(\xi_2,\xi_1) & \mathrm{Var}(\xi_2) & \mathrm{Cov}(\xi_2,\xi_3) \\ \mathrm{Cov}(\xi_3,\xi_1) & \mathrm{Cov}(\xi_3,\xi_2) & \mathrm{Var}(\xi_3) \end{pmatrix} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 4 \end{pmatrix}$$

$$\mathrm{Var}(\xi_1+\xi_2+\xi_3) = \mathrm{Var}\left(\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}\right) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \mathrm{Var}\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 9 \quad (1.41)$$

15.10.

$$\mathbb{E}(z_1) = \mathbb{E}\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathbb{E}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad (1.42)$$

$$\mathrm{Var}(z_1) = \mathrm{Var}\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathrm{Var}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

15.11. $\mathbb{E}(z_2) = \mathbb{E}\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathbb{E}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

Поскольку $z_2 = z_1 + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, где z_1 — случайный вектор из предыдущей задачи, то $\mathrm{Var}(z_2) = \mathrm{Var}(z_1)$. Сдвиг случайного вектора на вектор-константу не меняет его ковариационную матрицу.

15.12. *В данном примере проиллюстрирована процедура центрирования случайного вектора — процедура вычитания из случайного вектора его математического ожидания.*

$$\mathbb{E}(z_3) = \mathbb{E}\left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} - \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix}\right) = \mathbb{E}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} - \mathbb{E}\begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix} - \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Заметим, что вектор z_3 отличается от вектора z_1 (из задачи 15) сдвигом на вектор-константу $\begin{pmatrix}\mathbb{E}\xi_1\\ \mathbb{E}\xi_2\end{pmatrix}$, поэтому $\mathrm{Var}(z_3) = \mathrm{Var}(z_1)$.

15.13.

15.14.

15.15.

15.16.

15.17. Каждый из вариантов возможен

15.18.

16.1.

16.2.

16.3.

16.4.

16.5.

16.6.

16.7.

16.8.

16.9.

16.10.

16.11. по χ^2-распределению

16.12. $u \sim N(0, I)$

17.1.

17.2.

17.3.

17.4.

17.5.

17.6.

17.7.

17.8.

17.9.

17.10.

17.11.

17.12.