

Часть 1. Тест.**Вопрос 1 ♣**

Рассмотрим модель $Y = X\beta + \varepsilon$. Условия теоремы Гаусса-Маркова выполнены, причём $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2$, $\hat{Y} = PY$, $P = X(X'X)^{-1}X'$ и I - единичная матрица. Ковариационная матрица случайного вектора \hat{Y} равна

☐ $\sigma_\varepsilon^2(I - P)$

☐ $\sigma_\varepsilon^2(P - I)$

☐ $\sigma_\varepsilon^2 I$

☒ $\sigma_\varepsilon^2 P$

☐ $\sigma_\varepsilon^2(I + P)$

☐ Нет верного ответа.

Вопрос 2 ♣ Для регрессии $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, оцененной по 32 наблюдениям с $R^2 = 0.9$, значение тестовой статистики для проверки гипотезы об адекватности регрессии равно

☐ невозможно вычислить по имеющимся данным☐ 36/27☐ 81☒ 84☐ 243/4☐ Нет верного ответа.**Вопрос 3 ♣**

Если в уравнение регрессии не включена константа, то

☒ R^2 не является показателем качества подгонки регрессии☐ R_{adj}^2 в этой модели всегда неотрицательный☐ К этой модели неприменима теорема Гаусса-Маркова☐ Значимость коэффициентов регрессии нельзя проверять при помощи t-статистики☐ К этой модели неприменим МНК☐ Нет верного ответа.

Вопрос 4 ♣

По 546 наблюдениям за 1987 г. оценили зависимость стоимости частных домов в канаде price (измеряемой в долларах США) от общей площади square (измеряемой в кв. м.) и наличия системы кондиционирования aircond (1 — если есть, 0 — если нет):

переменная	коэффициент	ст. ошибка	t-статистика	P-значение
square	5.017	.488	10.26	0.000
aircon	7613.352	5544.33	1.37	0.170
square*aircon	2.254	.932	2.42	0.016
const	35683.87	2587.205	13.79	0.000

Согласно полученным результатам, при уровне значимости 5%, наличие кондиционера увеличивает стоимость каждого квадратного метра жилья на

- ☒ 2.254 \$ ☐ 0 ☐ 7613.352 \$
☐ 5.017 \$ ☐ 7.271\$ ☐ Нет верного ответа.

Вопрос 5 ♣

Оценена зависимость расходов потребителей на газ и электричество Y в США в 1977-1999 г. в постоянных ценах I квартала 1977 г. от времени ($t = 1$ для 1977, $t = 2$ для 1978 и т.д.) с учётом сезонных факторов ($D_i = 1$, если наблюдение относится к i -ому кварталу и 0 иначе, $i = 1, \dots, 4$):
 $\hat{Y} = 8 + 0.1t - 3D_2 - 2.6D_3 - 2D_4$

Если в качестве базовой категории будет принят не первый квартал, а второй, уравнение регрессии примет вид

- ☐ $\hat{Y} = 8 + 0.1t - 3D_1 - 2.6D_3 - 2D_4$ ☐ $\hat{Y} = 8 + 0.1t + 3D_1 - 2.6D_3 + D_4$
☒ $\hat{Y} = 5 + 0.1t + 3D_1 + 0.4D_3 + D_4$ ☐ $\hat{Y} = 5 + 0.1t - 3D_1 - 2.6D_3 - 2D_4$
☐ $\hat{Y} = 5 + 0.1t - 3D_1 - 0.4D_3 - D_4$ ☐ Нет верного ответа.

Вопрос 6 ♣

Для выбора между линейной и линейной в логарифмах моделями (где EARNINGS — почасовая заработная плата в \$, S — длительность обучения, ASVABC - результаты тестов, характеризующие успеваемость) был проведен тест Дэвидсона, Уайта и МакКиннона и получены следующие результаты:

	Зависимая: <i>EARNINGS</i>	Зависимая: $\ln EARNINGS$
(Intercept)	−28.23 (4.85)	−1.941 (3.2499)
S	2.127 (0.303)	
ASVABC	0.4107 (0.0848)	
lin_add	−18.822 (8.174)	
lnS		0.8809 (0.716)
lnASVABC		0.612 (0.445)
log_add		0.015 (0.094)
R^2	0.2052	0.2140
F	46.05	48.63
Adj. R^2	0.2008	0.2096
Num. obs.	539	540
RSS	91194	149.5
$\hat{\sigma}$	13.056	0.5281

Где $\text{lin_add} = \ln(\hat{Y}) - \ln Y$, $\text{log_add} = \hat{Y} - \exp(\ln Y)$ и в скобках указаны стандартные ошибки. На уровне значимости 5% можно сделать вывод, что

- | | | | |
|---------------------------------------|------------------------------------|----------------------------|--|
| <input type="checkbox"/> A | Лучше полулогарифмическая модель | <input type="checkbox"/> E | Между линейной и линейной в логарифмах моделями нет статистической разницы |
| <input checked="" type="checkbox"/> B | Лучше линейная в логарифмах модель | | |
| <input type="checkbox"/> C | Лучше линейная модель | | |
| <input type="checkbox"/> D | Невозможно выбрать лучшую модель | <input type="checkbox"/> F | Нет верного ответа. |

Вопрос 7 ♣

По данным для 27 фирм была оценена зависимость выпуска Y от труда L и капитала K с помощью моделей:

$$\ln Y_i = b_1 + b_2 \ln L_i + b_3 \ln K_i + \varepsilon_i \quad (1)$$

$$\ln Y_i = b_1 + b_2(\ln L_i + \ln K_i) + \varepsilon_i \quad (2)$$

Суммы квадратов остатков в этих моделях известны, $RSS_1 = 4$ и $RSS_2 = 6$. F -статистика для проверки гипотезы о равенстве эластичностей по труду и по капиталу равна

☐ A 8

☐ C 4

☐ E 2

☐ B 6

☒ 12

☐ F Нет верного ответа.

Вопрос 8 ♣

По одним и тем же наблюдениям оценили две регрессии: $\hat{Y} = 1 + 3X_1$ и $\hat{Y} = 2 + 5X_2$.

Известно, что $\widehat{\text{Cov}}(X_1, X_2) = 0$. Оценки МНК коэффициентов регрессии $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$:

☐ A $\hat{\beta}_1, \hat{\beta}_2$ найти невозможно, $\hat{\beta}_0 = 3$

☐ D оценки всех коэффициентов невозможно найти по имеющимся данным

☒ $\hat{\beta}_0$ найти невозможно, $\hat{\beta}_1 = 3, \hat{\beta}_2 = 5$

☐ E $\hat{\beta}_0 = 3, \hat{\beta}_1 = 3, \hat{\beta}_2 = 5$

☐ C $\hat{\beta}_0 = 1.5, \hat{\beta}_1 = 3, \hat{\beta}_2 = 5$

☐ F Нет верного ответа.

Вопрос 9 ♣

По данным для 27 фирм исследована зависимость прибыли Y от числа работников X вида $Y = \beta_0 + \beta_1 X + \varepsilon$ и получено $\hat{\beta}_0 = 7, \hat{\beta}_1 = 3, \hat{\sigma}^2 = 25$ и матрица $(X'X)^{-1} = \begin{pmatrix} 0.36 & -0.03 \\ -0.03 & 0.09 \end{pmatrix}$.

95% доверительный интервал для β_0 :

☐ A [1.12; 12.88]

☐ C [5.04; 8.96]

☒ [0.82; 13.18]

☐ B невозможно вычислить по имеющимся данным

☐ D [-0.09; 6.09]

☐ F Нет верного ответа.

Вопрос 10 ♣

Исследователь Анатолий оценил параметры нескольких моделей:

Модель	Уравнение
1	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
2	$Y = \beta_1 + \beta_4 X_4 + \beta_5 X_5 + u$
3	$Y/X_2 = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
4	$\ln Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$

С помощью R_{adj}^2 можно выбрать лучшую из моделей

☐ A 2 и 4

☐ C 1 и 3

☐ E 1 и 4

☐ B 2 и 3

☒ 1 и 2

☐ F Нет верного ответа.

Имя, фамилия и номер группы:

.....

Вопрос 1 : ☐ A ☐ B ☐ C ☐ D ☒ E ☐ F

Вопрос 2 : ☐ A ☒ B ☐ C ☐ D ☐ E ☐ F

Вопрос 3 : ☒ A ☐ B ☐ C ☐ D ☐ E ☐ F

Вопрос 4 : ☒ A ☐ B ☐ C ☐ D ☐ E ☐ F

Вопрос 5 : ☐ A ☒ B ☐ C ☐ D ☐ E ☐ F

Вопрос 6 : ☐ A ☒ B ☐ C ☐ D ☐ E ☐ F

Вопрос 7 : ☐ A ☐ B ☐ C ☒ D ☐ E ☐ F

Вопрос 8 : ☐ A ☒ B ☐ C ☐ D ☐ E ☐ F

Вопрос 9 : ☐ A ☐ B ☐ C ☐ D ☒ E ☐ F

Вопрос 10 : ☐ A ☐ B ☐ C ☒ D ☐ E ☐ F

Часть 2. Задачи.

1. По 29 наблюдениям оценили функцию спроса на яблоки

$$\widehat{\ln Q} = 14 - 5 \ln P_{apple} + 2.4 \ln P_{orange} + 2 \ln P_{banana},$$

где Q — спрос на яблоки, а P_{apple} — цена яблок, P_{orange} — цена апельсинов и P_{banana} — цена бананов.

Известна оценка ковариационной матрицы коэффициентов регрессии:

$$\widehat{\text{Var}}(\hat{\beta}) = \begin{pmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.1 & 2 & 0.5 & 0.7 \\ 0.2 & 0.5 & 3 & 0.6 \\ 0.3 & 0.7 & 0.6 & 4 \end{pmatrix}$$

На уровне значимости 5% проверьте гипотезу о том, что $\beta_{apple} + 2\beta_{orange} = 0$.

2. По ежемесячным данным за 2 года была оценена зависимость \hat{Y} - количества покупаемых автомобилей Honda, от цены P и дохода покупателя I с помощью трёх моделей:

$$\hat{Y} = -7.04 + 1.01 P + 0.52 I, R^2 = 0.88$$

(25.53) (1.147) (0.09)

$$\hat{Y} = -63.8 + 4.5 P, R^2 = 0.34$$

(53.3) (2.2)

$$\hat{Y} = 15.2 + 0.57 I, R^2 = 0.87$$

(4.2) (0.07)

Известно, что $\widehat{\text{Corr}}(P, I) = 0.52$, в скобках указаны стандартные отклонения коэффициентов.

Какую модель Вы предпочтёте и почему?

3. Для регрессии в отклонениях $y = \beta_1 x + \beta_2 z + \varepsilon$, оцененной по 100 наблюдениям, известны следующих суммы:

$$\sum y_i^2 = \frac{493}{3}, \sum x_i^2 = 30, \sum z_i^2 = 3, \sum x_i y_i = 30, \sum z_i y_i = 20, \sum x_i z_i = 0$$

Найдите оценки МНК коэффициентов β_1, β_2 и коэффициент детерминации R^2 .

4. Приведены результаты оценки зависимости логарифма арендной платы жилья в России, \lnPRICE , от общей площади, $GENSQUARE$ (кв. м.), наличия газа, GAS , (1 – есть, 0 – нет), наличия телефона, $PHONE$ (1 – есть, 0 – нет).

Модели 1 и 2 оценены для городов с численностью населения более миллиона человек, модель 3 – для сёл, модель 4 – для обеих выборок. В скобках в таблице приведены стандартные ошибки.

- а) Проверьте значимость коэффициентов в модели 2 при уровне значимости 5% и дайте интерпретацию полученным результатам.
- б) Проверьте гипотезу о равенстве 0.02 коэффициента при переменной $GENSQUARE$ при уровне значимости 0.01 в модели 2.
- в) Можно ли утверждать, что наличие газа и телефона в крупном городе не влияет на стоимость аренды? Ответ обоснуйте формулировкой и проверкой подходящей гипотезы.
- г) Опираясь на результаты оценки моделей 2, 3 и 4, можно ли утверждать, что зависимость стоимости жилья от рассмотренных выше факторов одинакова для крупного города и села? Ответ обоснуйте подходящим тестом.

	Модель 1	Модель 2	Модель 3	Модель 4
(Intercept)	7.168 (0.1334)	6.884 (0.2398)	5.413 (0.0372)	5.422 (0.0385)
$GENSQUARE$	0.012 (0.002687)	0.012 (0.0027)	0.0038 (0.000506)	0.00305 (0.00052)
GAS		0.119 (0.166)	0.217 (0.035)	0.301 (0.0355)
$PHONE$		0.197 (0.123)	0.334 (0.0372)	0.4937 (0.0385)
R^2	0.1330	0.1541	0.0661	0.0976
F	19.95	7.78	73.14	116.55
Adj. R^2	0.1264	0.1343	0.0652	0.0968
Num. obs.	132	132	3104	3236
RSS	23.05	22.494	2801.506	3138.69
$\hat{\sigma}$	0.42113	0.41921	0.95064	0.98546