

**Часть 1. Тест.****Вопрос 1 ♣**

В множественной регрессии с двумя регрессорами выборочные корреляции между зависимой переменной и регрессорами составили:  $\widehat{\text{Corr}}(Y, X_1) = 0.7$ ,  $\widehat{\text{Corr}}(Y, X_2) = 0.2$ . Тогда  $R^2$  будет равен

- |                                 |   |  |
|---------------------------------|---|--|
| <input type="checkbox"/> A 0.49 | <input type="checkbox"/> D 0.81                         | <input type="checkbox"/> F 0.25                |
| <input type="checkbox"/> B 0.9  | <input type="checkbox"/> E Не хватает данных для ответа | <input type="checkbox"/> G Нет верного ответа. |
| <input type="checkbox"/> C 0.04 |   |  |

**Вопрос 2 ♣**

Исследовательница Алевтина вновь изучает зависимость размера порции в мишленовском ресторане от его звёздности,  $star$  (от 1 до 3), и уровня цен,  $price$ . Она оценила модель вида  $size_i = \beta_1 + \beta_2 star2_i + \beta_3 star3_i + \beta_4 price_i + u_i$ , где  $star1, star2, star3$  - дамми-переменные, равные 1 для ресторанов с соответствующим числом звезд, и 0 иначе. У Алевтины есть  $n$  наблюдений. С помощью какой статистики она будет проверять гипотезу об отсутствии влияния звёздности на размер порции?

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> A $t_{n-3}$   | <input type="checkbox"/> D $F_{3,n-3}$ | <input type="checkbox"/> F $F_{2,n-3}$         |
| <input type="checkbox"/> B $F_{3,n-4}$ | <input type="checkbox"/> E $F_{2,n-4}$ | <input type="checkbox"/> G Нет верного ответа. |
| <input type="checkbox"/> C $t_{n-4}$   |  |  |

**Вопрос 3 ♣**

Исследователь Валериан оценил модель зависимости дохода человека,  $income$ , от его возраста,  $age$ , и получил следующую зависимость  $\widehat{income}_i = 150 + 900age_i - 10age_i^2$ . Все коэффициенты значимы на 5% уровне значимости. В каком возрасте доход достигает максимума?

- |   |  |
|---|--|
| <input type="checkbox"/> A 90 лет   | <input type="checkbox"/> E 45 лет  |
| <input type="checkbox"/> B Недостаточно данных                                      | <input type="checkbox"/> F Доход возрастает с ростом возраста, у функции нет максимума |
| <input type="checkbox"/> C 30 лет   | <input type="checkbox"/> G Нет верного ответа.   |
| <input type="checkbox"/> D Доход убывает с ростом возраста, у функции нет максимума |  |

**Вопрос 4 ♣**

Регрессия с тремя регрессорами оценена по 100 наблюдениям. Какая из этих гипотез НЕ может быть проверена при помощи статистики, имеющей  $F_{2,96}$  распределение?

- |   |  |  |
|---|--|--|
| <input type="checkbox"/> A $H_0 : \beta_2 = 2\beta_3$       | <input type="checkbox"/> D $H_0 : \beta_1 + \beta_2 = 10; \beta_3 = 1$ | <input type="checkbox"/> G Нет верного ответа. |
| <input type="checkbox"/> B $H_0 : \beta_2 = \beta_3 = 0$    | <input type="checkbox"/> E $H_0 : \beta_1 = \beta_2 = \beta_3$         |  |
| <input type="checkbox"/> C $H_0 : \beta_1 = 1; \beta_3 = 5$ | <input type="checkbox"/> F $H_0 : \beta_2 = 3\beta_3 = 5$              |  |

**Вопрос 5 ♣**

Исследовательница Алевтина изучает зависимость размера порции в мишленовском ресторане от его звёздности,  $star$  (от 1 до 3), и уровня цен,  $price$ . Она оценила модель вида  $size_i = \beta_1 + \beta_2 star2_i + \beta_3 star3_i + \beta_4 price_i + u_i$ , где  $star1, star2, star3$  - дамми-переменные, равные 1 для ресторанов с соответствующим числом звезд, и 0 иначе. Алевтина считает, что размер порции уменьшается вдвое с каждой дополнительной звездой. Какую гипотезу ей нужно проверить?

- |  |  |
|--|--|
| <input type="checkbox"/> A $H_0 : \beta_3 = 4, \beta_2 = 2, \beta_1 = 1$ | <input type="checkbox"/> E $H_0 : \beta_1 < \beta_2 < \beta_3$   |
| <input type="checkbox"/> B $H_0 : \beta_1 = 2\beta_2 = 4\beta_3$         | <input type="checkbox"/> F $H_0 : \beta_1 = 2\beta_2 = 3\beta_3$ |
| <input type="checkbox"/> C $H_0 : 3\beta_1 = 2\beta_2 = \beta_3$         | <input type="checkbox"/> G Нет верного ответа.                   |
| <input type="checkbox"/> D $H_0 : 4\beta_1 = 2\beta_2 = \beta_3$         |  |

**Вопрос 6 ♣**

В регрессии с четырьмя регрессорами, оцененной по 21 наблюдению, оказалось, что  $TSS = 250$ ,  $R^2_{adj} = 0.75$ . Тогда  $R^2$  равен

- |                                |                                 |  |
|--------------------------------|---------------------------------|--|
| <input type="checkbox"/> A 0.7 | <input type="checkbox"/> D 0.8  | <input type="checkbox"/> G Нет верного ответа. |
| <input type="checkbox"/> B 0.6 | <input type="checkbox"/> E 0.96 |  |
| <input type="checkbox"/> C 0.2 | <input type="checkbox"/> F 0.3  |  |

**Вопрос 7 ♣**

Истинной является модель  $Y_i = \beta_1 + \beta_2 X_i + u_i$ . Глафира оценивает две регрессии:  $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$  и  $\hat{Y}_i = \hat{\gamma}_1 + \hat{\gamma}_2 X_i + \hat{\gamma}_3 Z_i$  с помощью МНК. Известна выборочная корреляция  $\widehat{\text{Corr}}(X_i, Z_i) = -0.2$ . Тогда оценка  $\hat{\gamma}_2$  является

- |   |  |
|---|--|
| <input type="checkbox"/> A смещённой на $-0.2\hat{\gamma}_3$ относительно $\beta_2$ ,<br>но эффективной оценкой | <input type="checkbox"/> E несостоятельной, но эффективной оценкой для $\beta_2$ |
| <input type="checkbox"/> B состоятельной, но смещённой на $-0.2$ относительно $\beta_2$ оценкой                 | <input type="checkbox"/> F несостоятельной и неэффективной оценкой для $\beta_2$ |
| <input type="checkbox"/> C состоятельной, но неэффективной оценкой для $\beta_2$                                | <input type="checkbox"/> G Нет верного ответа.                                   |
| <input type="checkbox"/> D состоятельной, но смещённой на $-0.2\hat{\gamma}_3$                                  |  |

**Вопрос 8 ♣**

Исследователь выполнил второй шаг в РЕ-тесте МакКиннона. В регрессии  $\ln Y_i$  на исходные регрессоры и  $Z_i = \hat{Y}_i - \exp(\widehat{\ln Y_i})$  коэффициент при  $Z_i$  оказался значимым. А в регрессии  $Y_i$  на исходные регрессоры и  $W_i = \ln \hat{Y}_i - \widehat{\ln Y_i}$  коэффициент при  $W_i$  оказался незначимым. Из результатов следует сделать вывод, что

- |  |   |
|--|---|
| <input type="checkbox"/> А в исходной модели пропущен регрессор $Z_i$                          | <input type="checkbox"/> Д следует предпочесть логарифмическую модель     |
| <input type="checkbox"/> В тесты противоречат друг другу, ни одна из моделей не предпочитается | <input type="checkbox"/> Е следует предпочесть полулогарифмическую модель |
| <input type="checkbox"/> С в исходной модели пропущен регрессор $W_i$                          | <input type="checkbox"/> Ф следует предпочесть линейную модель            |
|  | <input type="checkbox"/> Г Нет верного ответа.                            |

**Вопрос 9 ♣**

Если для регрессора используется преобразование Бокса-Кокса с параметром  $\theta = -1$ , а для зависимой переменной — с параметром  $\lambda = 0$ , то регрессионное уравнение представимо в виде

- |  |  |
|--|--|
| <input type="checkbox"/> А $\ln Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + u_i$ | <input type="checkbox"/> Е $Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + u_i$ |
| <input type="checkbox"/> В $Y_i = \beta_1 + \beta_2 X_i + u_i$               | <input type="checkbox"/> Ф $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$       |
| <input type="checkbox"/> С $\ln Y_i = \beta_1 + \beta_2 (X_i - 1) + u_i$     | <input type="checkbox"/> Г Нет верного ответа.                           |
| <input type="checkbox"/> Д $\ln Y_i = \beta_1 + \beta_2 X_i^2 + u_i$         |  |

**Вопрос 10 ♣**

Исследователь Феофан оценил регрессию  $Y$  на  $Z$  и получил, что  $\hat{Y}_i = 10 + 2Z_i$ . После этого, он оценил регрессию  $Y$  на  $Z$  и новую переменную  $X$ . Известна выборочная ковариация,  $\widehat{\text{Cov}}(Y, X) = 0$ . Выберите верное утверждение:

- |  |  |
|--|--|
| <input type="checkbox"/> А $R_{adj}^2$ вырос и стал равен 1                  | нулю   |
| <input type="checkbox"/> В $R^2$ не снизился по сравнению с исходной моделью | <input type="checkbox"/> Е Коэффициент при $Z$ стал незначимым         |
| <input type="checkbox"/> С Коэффициент при $Z$ остался значимым              | <input type="checkbox"/> Ф Коэффициент при переменной $Z$ не изменился |
| <input type="checkbox"/> Д Коэффициент при переменной $X$ равен              | <input type="checkbox"/> Г Нет верного ответа.                         |

## Часть 2. Задачи.

1. Пусть  $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$  и  $i = 1, \dots, 5$  — классическая регрессионная модель. Также имеются следующие данные:  $\sum_{i=1}^5 y_i^2 = 55$ ,  $\sum_{i=1}^5 x_i^2 = 3$ ,  $\sum_{i=1}^5 x_i y_i = 12$ ,  $\sum_{i=1}^5 y_i = 15$ ,  $\sum_{i=1}^5 x_i = 3$ .

а) Найдите  $\hat{\beta}_1, \hat{\beta}_2$

б) Найдите  $TSS, ESS, RSS, R^2, \hat{\sigma}^2$

2. Исследовательница Глафира изучает зависимость спроса на молоко от цены молока и дохода семьи. В её распоряжении есть следующие переменные:

- *price* — цена молока в рублях за литр
- *income* — ежемесячный доход семьи в тысячах рублей
- *milk* — расходы семьи на молоко за последние семь дней в рублях

В данных указано, проживает ли семья в сельской или городской местности. Поэтому Глафира оценила три регрессии: (All) — по всем данным, (Urban) — по городским семьям, (Rural) — по сельским семьям.

	(All)	(Urban)	(Rural)
(Intercept)	2.198 (4.648)	13.570* (5.849)	-5.089 (6.726)
income	0.203*** (0.057)	0.085 (0.082)	0.229** (0.075)
price	-0.252 (0.181)	-0.316 (0.219)	-0.061 (0.268)
R-squared	0.1	0.1	0.2
adj. R-squared	0.1	0.0	0.1
sigma	5.6	4.7	5.8
F	6.4	1.2	4.8
P-value	0.0	0.3	0.0
RSS	3032.1	1007.6	1625.1
n observations	100	48	52

а) Проверьте значимость в целом регрессии (All) на 5%-ом уровне значимости.

б) На 5%-ом уровне значимости проверьте гипотезу, что зависимость спроса на молоко является единой для городской и сельской местности.

3. У Эконометрессы Глафиры было четыре наблюдения и она решила оценить модель парной регрессии  $y_i = \beta_1 + \beta_2 x_i + u_i$ . Её подруга эконометресса Анжелла решила, что четыре наблюдения — мало, и поэтому учла каждое наблюдение 10 раз, так что в результате у неё вышло 40 наблюдений.

- а) Во сколько раз будут отличаться оценки  $\hat{\beta}_1$  у Глафиры и Анжелы? Оценки  $\hat{\beta}_2$ ?  
б) Во сколько раз у будут отличаться  $RSS$ ?  $TSS$ ?  $ESS$ ?  $R^2$ ?

4. У эконометрессы Агнессы есть дамми-переменная  $male_i$ , равная 1 для мужчин, и дамми-переменная  $female_i$ , равная 1 для женщин. Зависимая переменная  $y_i$  — доход индивида.

$$A : \hat{y}_i = \hat{\beta} male_i$$

$$B : \hat{y}_i = \hat{\gamma} female_i$$

$$C : \hat{y}_i = \hat{\alpha}_1 male_i + \hat{\alpha}_2 female_i$$

$$D : \hat{y}_i = \hat{\delta}_1 + \hat{\delta}_2 male_i$$

- а) Проинтерпретируйте оценки коэффициентов во всех регрессиях.  
б) Как связаны между собой оценки коэффициентов в регрессиях С и D?

5. По 2040 наблюдениям оценена модель зависимости стоимости квартиры в Москве (в 1000\$) от общего метража и метража жилой площади.

	Estimate	Std. Error	t value	Pr(> t )
Константа	-88.81	4.37	-20.34	0.00
Общая площадь	1.70	0.10	17.78	0.00
Жилая площадь	1.99	0.18	10.89	0.00

Сумма квадратов остатков равна  $RSS = 2.2216891 \times 10^6$ . Оценка ковариационной матрицы  $\widehat{Var}(\hat{\beta})$  имеет вид

	(Intercept)	totsp	livesp
(Intercept)	19.0726	0.0315	-0.4498
totsp	0.0315	0.0091	-0.0151
livesp	-0.4498	-0.0151	0.0335

- а) Постройте 95%-ый доверительный интервал для ожидаемой стоимости квартиры с жилой площадью 30 м<sup>2</sup> и общей площадью 60 м<sup>2</sup>.
- б) Постройте 95%-ый прогнозный интервал для фактической стоимости квартиры с жилой площадью 30 м<sup>2</sup> и общей площадью 60 м<sup>2</sup>.