

Имя, фамилия и номер группы:

.....

### Тест

**Вопрос 1.** (1 балл) Чудо-швабры производятся на разных заводах по одной из двух технологий,  $A$  или  $B$ . Исследователь оценил две модели зависимости выпуска,  $Y$ , от количества сырья,  $X$ , и технологии:

$$\hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 A_i + \hat{\alpha}_2 X_i + \hat{\alpha}_3 A_i X_i;$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 B_i + \hat{\beta}_2 X_i + \hat{\beta}_3 B_i X_i.$$

Переменная  $A_i$  равна единице для заводов с технологией  $A$  и нулю иначе, а переменная  $B_i$  равна единице для заводов с технологией  $B$  и нулю иначе.

Оценки коэффициентов связаны соотношением

☐  $\hat{\alpha}_1 = \hat{\beta}_0$

☐  $\hat{\alpha}_0 = \hat{\beta}_0 + \hat{\beta}_1$

☐  $\hat{\alpha}_2 = \hat{\beta}_2$

☐  $\hat{\alpha}_0 + \hat{\alpha}_1 = \hat{\beta}_0$

☐  $\hat{\alpha}_0 = \hat{\beta}_0$

☐ нет верного ответа

**Вопрос 2.** (1 балл) Исследователь Теофан оценил с помощью МНК модель  $Y = \beta_0 I + \beta_1 Z + \beta_2 W + u$ , где  $I$  – столбец из единиц. Для матрицы факторов,  $X = (IZW)$ , известно, что

$$(X'X)^{-1} = \begin{pmatrix} 0.04 & 0.012 & -0.008 \\ 0.012 & 0.03 & -0.007 \\ -0.008 & -0.007 & 0.02 \end{pmatrix}$$

Предпосылки теоремы Гаусса-Маркова выполнены. Отношение дисперсии оценки  $\hat{\beta}_0$  к дисперсии оценки  $\hat{\beta}_2$  равно

☐  $3/2$

☐  $1/2$

☐  $10/3$

☐  $-5/1$

☐  $2$

☐ нет верного ответа

**Вопрос 3.** (2 балла) Для регрессионной модели со свободным членом известно, что

$$X'X = \begin{pmatrix} 20 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & 5 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 40 \\ 10 \\ 13 \end{pmatrix}, \quad \sum_{i=1}^n Y_i^2 = 140.$$

Коэффициент  $R^2$  в этой модели равен

☐  $0.5$

☐  $13/14$

☐  $9/35$

☐  $0.6$

☐ недостаточно информации

☐ нет верного ответа

Вопрос 4. (1 балл) Выборочная корреляция между регрессорами  $X$  и  $Z$  равна 0.5. В регрессии  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 Z_i$  показатель  $VIF$  для регрессора  $X$  равен

- ☐ A 1/2                      ☐ C 3/4                      ☐ E 4/3  
☐ B 1/4                      ☐ D 2                      ☐ F нет верного ответа

Вопрос 5. (1 балл) Какое условие НЕ требуется в теореме Гаусса-Маркова?

- ☐ A модель  $Y = X\beta + \varepsilon$  правильно специфицирована                      ☐ D случайные ошибки  $\varepsilon_i$  не коррелированы  
☐ B случайные ошибки  $\varepsilon_i$  имеют одинаковые дисперсии                      ☐ E случайные ошибки  $\varepsilon_i$  нормально распределены  
☐ C матрица регрессоров  $X$  имеет полный ранг                      ☐ F нет верного ответа

Вопрос 6. (2 балла) Исследовательница Клеопатра оценила модель  $\ln Y_i = \beta_0 + \beta_1 \ln X_i + \beta_2 \ln Z_i + \beta_3 \ln W_i + u_i$ . Клеопатра хочет протестировать гипотезу  $H_0: \beta_3 + 2\beta_1 = 1$ . Для этой цели можно оценить вспомогательную регрессию

- ☐ A  $\ln(Y_i/W_i) = \gamma_0 + \gamma_1 \ln(X_i/W_i^2) + \gamma_2 \ln Z_i + u_i$                       ☐ D  $\ln(Y_i/W_i) = \gamma_0 + \gamma_1 \ln(X_i \cdot W_i^2) + \gamma_2 \ln Z_i + u_i$   
☐ B  $\ln(Y_i \cdot W_i) = \gamma_0 + \gamma_1 \ln(X_i/W_i^2) + \gamma_2 \ln Z_i + u_i$                       ☐ E  $\ln(Y_i \cdot W_i) = \gamma_0 + \gamma_1 \ln(X_i \cdot W_i^2) + \gamma_2 \ln Z_i + u_i$   
☐ C  $\ln(Y_i/W_i^2) = \gamma_0 + \gamma_1 \ln(X_i/W_i) + \gamma_2 \ln Z_i + u_i$                       ☐ F нет верного ответа

Вопрос 7. (2 балла) Арамис построил регрессию по 66 наблюдениям:

$$\hat{Y}_i = \underset{(0.4)}{4} + \underset{(5)}{6} X_i + \underset{(2)}{4.4} Z_i - \underset{(2)}{3} Q_i - \underset{(3)}{9} R_i + \underset{(10)}{16} S_i.$$

В скобках указаны стандартные ошибки. Показатель  $R_{adj}^2$  может увеличиться при удалении из модели группы факторов

- ☐ A  $X, Q, S$                       ☐ C  $X, S$                       ☐ E  $Q, S$   
☐ B  $X, Q$                       ☐ D  $S$                       ☐ F нет верного ответа

Вопрос 8. (2 балла) Для регрессии  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 Z_i + \hat{\beta}_3 W_i$ , оценённой по 24 наблюдениям,  $R^2 = 0.9$ . При проверке гипотезы о неадекватности модели  $F$ -статистика равна

- ☐ A 45                      ☐ C 200.27                      ☐ E 5/9  
☐ B 189/2                      ☐ D 60                      ☐ F нет верного ответа

Вопрос 9. (1 балл) Исследовательница Надежда оценила регрессию в отклонениях,  $\hat{y}_i = x_i + 2z_i$  с помощью МНК. Известно, что  $\bar{Y} = 5$ ,  $\bar{X} = 6$ ,  $\bar{Z} = -2$ . В регрессии нецентрированных переменных,  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 Z_i$ , оценка коэффициента  $\hat{\beta}_0$  равна

☐ A 5☐ C 4☐ E 3☐ B 2☐ D 1☐ F нет верного ответа

Вопрос 10. (1 балл) Портос построил регрессию по 66 наблюдениям,  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 W_i + \hat{\beta}_3 Z_i$ ,  $RSS = 140$ . Затем Портос оценил вспомогательную регрессию,  $\hat{Y}_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_i + \hat{\gamma}_2 W_i + \hat{\gamma}_3 Z_i + \hat{\delta}_2 \hat{Y}_i^2 + \hat{\delta}_3 \hat{Y}_i^3$ ,  $RSS = 120$ .

При проверке гипотезы о правильной спецификации модели в тесте Рамсея  $F$ -статистика равна

☐ A 5☐ C 30/7☐ E 6☐ B 11/3☐ D 10/3☐ F нет верного ответа

### Задачи

1. (5 баллов) Рассмотрим алгоритм LASSO с параметром регуляризации  $\lambda$  для модели  $Y = X\beta + \varepsilon$ , где все переменные центрированы.
- а) Выпишите целевую функцию алгоритма.
  - б) Что произойдет с оценками  $\hat{\beta}_{LASSO}$  при  $\lambda \rightarrow \infty$ ?
  - в) Что произойдет с оценками  $\hat{\beta}_{LASSO}$  при  $\lambda \rightarrow 0$ ?

2. (5 баллов) По 200 фирмам была оценена зависимость выпуска  $Y$  от труда  $L$  и капитала  $K$  с помощью двух моделей:

Модель Кобба-Дугласа:  $\ln Y_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + \varepsilon_i$

Транслоговая модель:  $\ln Y_i = \gamma_0 + \gamma_1 \ln L_i + \gamma_2 \ln K_i + \gamma_3(0.5 \ln^2 L_i) + \gamma_4(0.5 \ln^2 K_i) + \gamma_5 \ln K_i \ln L_i + \varepsilon_i$

Оценки коэффициентов обеих моделей (в скобках приведены стандартные ошибки):

Переменная	Модель Кобба-Дугласа	Транслоговая модель
константа	1.1706 (0.326)	0.9441 (2.911)
$\ln L$	0.6029 (0.125)	3.613 (1.548)
$\ln K$	0.375 (0.085)	-1.893 (1.016)
$0.5 \ln^2 L$		-0.964 (0.707)
$0.5 \ln^2 K$		0.0852 (0.2922)
$\ln L \ln K$		0.3123 (0.4389)
$R^2$	0.9	0.954

В модели Кобба-Дугласа  $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.0096$ .

На уровне значимости  $\alpha = 0.05$  проверьте следующие гипотезы:

- В модели Кобба-Дугласа эластичность выпуска по капиталу равна единице.
- В модели Кобба-Дугласа эластичности выпуска по труду и капиталу одинаковы.
- В транслоговой модели  $\gamma_4 = 0$ .
- В транслоговой модели  $\gamma_3 = \gamma_4 = \gamma_5 = 0$ .



3. (4 балла) Исследователь оценил зависимость продолжительности жизни  $Y$  от концентрации промышленных выбросов в атмосфере  $X$  и ежегодных частных расходов на медицинскую помощь  $Z$ .

Для 300 жителей индустриальных центров,  $\hat{Y}_i = 65.91 - 0.03 X_i - 0.036 Z_i$ ,  $RSS = 400$ .  
(10.43) (0.0001) (0.019)

Для 200 сельских жителей,  $\hat{Y}_i = 58.4 - 0.017 X_i - 0.024 Z_i$ ,  $RSS = 200$ .  
(15.3) (0.006) (0.007)

А также по общей выборке,  $\hat{Y}_i = 63.2 - 0.02 X_i - 0.031 Z_i$ ,  $RSS = 800$ .  
(12.4) (0.005) (0.001)

В скобках приведены стандартные ошибки.

Можно ли считать, что зависимость едина для городских и сельских жителей? Ответ обоснуйте подходящим тестом, аккуратно выписав тестируемую гипотезу.

4. (5 баллов) Исследователь Д'Артаньян стандартизировал (центрировал и нормировал) все имеющиеся регрессоры и поместил их в столбцы матрицы  $\tilde{X}$ . Выборочная корреляционная матрица регрессоров равна:

$$\begin{pmatrix} 1 & 0.95 & 0 \\ 0.95 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- а) Найдите параметр обусловленности (condition number) матрицы  $\tilde{X}^T \tilde{X}$ .
- б) Вычислите одну или две главные компоненты, объясняющие не менее 70% суммарной дисперсии стандартизированных регрессоров. Выпишите найденные компоненты как линейные комбинации столбцов матрицы  $\tilde{X}$ .

5. (6 баллов) Для 400 голландских магазинов модной одежды с помощью трёх моделей оценили зависимость продаж в расчете на квадратный метр в гульденах,  $Sales$ , от:

- общей площади магазина,  $Size$ , в  $m^2$ ;
- количества сотрудников, работающих целый день,  $Nfull$ ;
- количества временных рабочих,  $Ntemp$ ;
- дамми-переменной  $Owner$ , равной единице, если собственник один, и нулю иначе.

$$\widehat{Sales}_i = 6083 - 15.25Size_i + 1452.8Nfull_i + 420.15Ntemp_i - 1464.1Owner_i$$

(718)      (1.59)                      (171)                      (423)                      (361)

$$\ln \widehat{Sales}_i = 8.59 - 0.0024Size_i + 0.183Nfull_i + 0.102Ntemp_i - 0.209Owner_i$$

(0.11)      (0.00024)                      (0.026)                      (0.066)                      (0.056)

$$\ln \widehat{Sales}_i = 10.08 - 0.31 \ln Size_i + 0.22 \ln Nfull_i + 0.066 \ln Ntemp_i - 0.19 \ln Owner_i$$

(0.21)      (0.043)                      (0.061)                      (0.118)                      (0.059)

В скобках приведены стандартные ошибки.

- а) Дайте интерпретацию коэффициента при переменной  $Size$  в каждой из трёх моделей;
- б) Подробно опишите, как выбрать наилучшую из этих моделей.