

Демонстрационная версия
проверочной работы №2

Имя и фамилия студента:

Задание 1. Ознакомьтесь с нижеприведенной выдачей по результатам оценивания линейной регрессионной модели:

```
> cor(df)
      y      x      z      w
y  1.0000  0.8035 -0.203 -0.216
x  0.8035  1.0000 -0.717 -0.692
z  -0.203 -0.717  1.0000  0.9514
w  -0.216 -0.692  0.9514  1.0000

> model.1 <- lm(y ~ x + z + w, data=df)
> summary(model.1)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.7762      3.5310  -3.052  0.03796 *
x              1.5434      0.2138   7.218  0.00195 **
z              0.1507      0.3014   0.500  0.64971
w              0.8212      0.4382   1.874  0.09497 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9486, Adjusted R-squared:  0.8990

> vif(model.1)
      x      z      w
2.154286 12.03485 7.954813
```

1. Можно ли говорить о том, что model.1 страдает от мультиколлинеарности? Приведите не менее двух свидетельств из выдачи в качестве обоснования. Можно приводить как прямые, так и косвенные свидетельства мультиколлинеарности. (2 балла)

2. Проинтерпретируйте значение VIF для переменной w. Что оно показывает? (1 балл)

Задание 2. 1. По указанным ниже значениям предиктора и отклика (x и y соответственно) найдите вектор оценок коэффициентов в регрессионной модели y на x . Используйте для этого общую формулу оценки коэффициентов в векторно-матричном виде (релевантную как для парной, так и для множественной регрессии). Запишите промежуточные расчеты. В качестве ответа запишите сам вектор и полную спецификацию модели, подставив эти оценки коэффициентов. (3 балла)

x	4	0	2	2	1
y	2	1	1	5	2

2. Рассчитайте предсказанное значение зависимой переменной для первого наблюдения. (1 балл)

Задание 3. Ниже представлены результаты анализа разложения вариации по линейной регрессионной модели, построенной по выборке 161 наблюдений.

Analysis of Variance Table

Response: cpi

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dem	1	236.833	236.833		
fp	1	77.850	77.850		
loggdppc	1	159.474	159.474		
stab	1	15.318	15.318		
britcol	1	3.141	3.141		
Residuals	...	144.703	...		

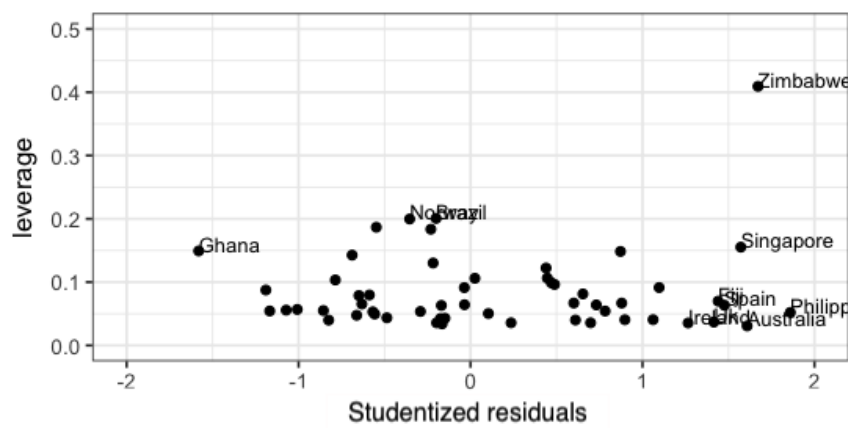
Используя информацию из данной таблицы,

1. рассчитайте значение коэффициента детерминации (1 балл)
2. рассчитайте значение необходимой статистики критерия для проверки гипотезы о незначимости коэффициента детерминации (1 балл)
3. сделайте вывод о (не)значимости коэффициента детерминации на основе значения p-value, если известно, что p-value: $< 2.2e-16$ (представлена запись из R) (1 балл)

Задание 4. Отметьте ВСЕ верные утверждения, если таковые имеются. Если таковых нет, напишите слово «НЕТ»: (1 балл)

- 1) В результате сильной мультиколлинеарности оценки в регрессионной модели перестают быть BLUE (best linear unbiased estimators).
- 2) Матрица проекции (Hat-matrix) является симметричной.
- 3) Одним из следствий сильной мультиколлинеарности является неустойчивость оценок коэффициентов.
- 4) Элементы матрицы проекции (hat-matrix) позволяют выявить наблюдения с нетипичными значениями по зависимой переменной.

Задание 5. По результатам регрессионного анализа был построен график для определения нетипичных наблюдений. На основе этого графика (см. ниже) укажите, есть ли нетипичные наблюдения, для удобства можете указать их на графике. Охарактеризуйте эти нетипичные наблюдения, если они имеются. (2 балла)



Задание 6. После оценивания линейной регрессионной модели y на x_1, x_2 по 100 наблюдениям в критерии Уайта, используемом для тестирования гетероскедастичности, было получено значение R^2 из вспомогательной регрессии равное 0.2. Сформулируйте нулевую и альтернативную гипотезы, рассчитайте статистику критерия и сделайте вывод. Ниже представлен список возможных процентных точек, отметьте необходимую(-ые) при уровне значимости 0.05 (2 балла):

1. квантиль распределения хи-квадрат уровня 0.95, $df = 2$: 5.99
2. квантиль распределения хи-квадрат уровня 0.975, $df = 2$: 7.37
3. квантиль распределения хи-квадрат уровня 0.95; $df = 4$: 9.49
4. квантиль распределения хи-квадрат уровня 0.975, $df = 4$: 11.14
5. квантиль распределения хи-квадрат уровня 0.95; $df = 5$: 11.07
6. квантиль распределения хи-квадрат уровня 0.975, $df = 5$: 12.83