

Решение задач

Задание 1. Ниже в таблице представлены значения переменных: X , Z , Y .

X	-2	2	1	-1	0
Z	2	0	0	-1	-1
Y	2	6	3	2	10

1. Без использования Python получите оценки коэффициентов в регрессии Y на X с помощью общей формулы получения оценок коэффициентов, подходящей как для парной, так и для множественной регрессии. Представьте промежуточные расчеты, выпишите полученный вектор оценок коэффициентов и запишите спецификацию модели, подставив эти оценки в уравнение

Для начала запишем матрицу X :

$$X = \begin{pmatrix} 1 & -2 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$$

$$(X^T X)^{-1} = \frac{1}{50} \times \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 2 & 1 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} 2 \\ 6 \\ 3 \\ 2 \\ 10 \end{pmatrix} = \begin{pmatrix} 23 \\ 9 \end{pmatrix}$$

$$(X^T X)^{-1} X^T Y = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix} \times \begin{pmatrix} 23 \\ 9 \end{pmatrix} = \begin{pmatrix} 4.6 \\ 0.9 \end{pmatrix}$$

2. Без использования Python получите оценки коэффициентов в регрессии Y на X и Z с помощью общей формулы получения оценок коэффициентов, подходящей как для парной, так и для множественной регрессии. Представьте промежуточные расчеты, выпишите полученный вектор оценок коэффициентов и запишите спецификацию модели, подставив эти оценки в уравнение

$$X = \begin{pmatrix} 1 & -2 & 2 \\ 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & -1 \\ 1 & 0 & -1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 10 & -3 \\ 0 & -3 & 6 \end{pmatrix}$$

$$\det(X^T X) = 5 \times 10 \times 6 - (-3) \times (-3) \times 5 = 255$$

$$(X^T X)^{-1} = \frac{1}{255} \times \begin{pmatrix} 51 & 0 & 0 \\ 0 & 30 & 15 \\ 0 & 15 & 50 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 2 & 1 & -1 & 0 \\ 2 & 0 & 0 & -1 & -1 \end{pmatrix} \times \begin{pmatrix} 2 \\ 6 \\ 3 \\ 2 \\ 10 \end{pmatrix} = \begin{pmatrix} 23 \\ 9 \\ -8 \end{pmatrix}$$

$$(X^T X)^{-1} X^T Y = \frac{1}{255} \times \begin{pmatrix} 51 & 0 & 0 \\ 0 & 30 & 15 \\ 0 & 15 & 50 \end{pmatrix} \times \begin{pmatrix} 23 \\ 9 \\ -8 \end{pmatrix} = \begin{pmatrix} 4.6 \\ 0.588 \\ -1.039 \end{pmatrix}$$

Задание 2. Ниже представлены результаты анализа разложения вариации по линейной парной регрессионной модели, построенной по выборке из 15 наблюдений.

Analysis of Variance Table

Response: y					
	df	sum_sq	mean_sq	f	PR(>F)
x	0.6526	...
Residual	...	56.116	...		

Восстановим пропуски в таблице:

- df для x = k = 1
- df для Residual = n-k-1 = 15-2 = 13
- mean_sq для Residual = $\frac{56.116}{13} \approx 4.317$
- $f = \frac{mean_sq(x)}{mean_sq(Residual)} = \frac{mean_sq(x)}{4.317} = 0.6526$

Следовательно, mean_sq для x = $4.317 \times 0.6526 \approx 2.817$

- Рассчитаем p-value, помнив о том, что нас интересует односторонняя альтернатива. В Python можно рассчитать следующим образом:

```
from scipy.stats import f
f.sf(0.6526, 1, 13)
```

В итоге получили 0.434, что говорит о том, что R^2 неотличим от 0.

$$R^2 = \frac{2.817}{56.116 + 2.817} \approx 0.048$$

Analysis of Variance Table

Response: y					
	df	sum_sq	mean_sq	f	PR(>F)
x	1	2.817	2.817	0.6526	0.434
Residual	13	56.116	4.316		

Задание 3.

На данных по 44 городам построена модель, объясняющая динамику уровня преступности за последние 10 лет. `change_in_crime_rate` — прирост преступности в %, `change_in_pop` — прирост численности населения, %; `kids` — процент детей; `free_lunch` — процент бесплатных школьных обедов; `income_change` — прирост доходов домохозяйств.

Восстановим пропуски в таблице:

```
Coefficients:
      coef    std. err   t      Pr>|t|  [0.025  0.975]
Intercept  -22.3548    12.3097 -1.816   0.0771  -47.253;  2.544
change_in_pop  0.3188     0.2052  1.533   0.1333   -0.096;  0.734
kids         1.1128     0.2869  3.879   0.0004    0.532;  1.693
free_lunch   -0.3681     0.0973 -3.783   0.0005   -0.565; -0.171
income_change -0.1944     0.3681 -0.528   0.6004   -0.939;  0.551
```

```
      df    sum_sq  mean_sq  f      PR(>F)
change_in_pop  1     803.2    803.2  6.248   .000
kids           1    1380.1   1380.1
free_lunch     1    3186.6   3186.6
income_change  1      60.6    60.6
Residual      39   8476.0    217.3
```

$$R^2 = \frac{803.2 + 1380.1 + 3186.6 + 60.6}{803.2 + 1380.1 + 3186.6 + 60.6 + 8476} = 0.39$$

То есть, данная модель лучше, чем модель на константу.