

Задания для дополнительной тренировки

Задание 1.

Ниже представлены результаты применения метода главных компонент. Исходные индикаторы: X, Y, Z.

	PC1	PC2	PC3
X	0.5884	−0.4993	0.6360
Y	0.6129	−0.2377	−0.7536
Z	0.5274	−0.8332	0.1662
Variance	2.5149	0.4305	0.0545

1. Рассчитайте информативность **первой** главной компоненты?
2. Сформулируйте критерий Кайзера для определения количества извлекаемых главных компонент. Сколько, согласно данному критерию, главных компонент необходимо извлекать в данном случае?
3. Чему равен след (trace) ковариационной матрицы исходных переменных X, Y, Z (то есть, сумма элементов матрицы по главной диагонали)? Проинтерпретируйте это значение.

Задание 2. Известно, что настоящая модель регрессии выглядит следующим образом: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$, и ее МНК-оценки являются BLUE. Вы оцениваете модель $y_i = \theta_0 + \theta_1 x_i + \eta_i$. Чем грозит исключение из модели релевантного регрессора z ? В каком случае при исключении предиктора оценки останутся BLUE?

Задание 3. Изучается влияние факта, занимается ли школьник у репетитора, на его успеваемость. У исследователя возник вопрос, нужно ли включать среднюю успеваемость в его классе в качестве контрольной переменной. Всего в классе 6 учащихся. Рассмотрите случаи как включения средней успеваемости в качестве контрольной переменной, так и невключения. Какие могут быть последствия в случае первого и второго решения? Свой ответ объясните.

Задание 4. Ознакомьтесь с постановочной частью исследования [Lü, Landry, 2014](#).

1. Проинтерпретируйте оценки коэффициентов Table 2, Model 1.
2. Проинтерпретируйте график Figure 4, Panel 2. Что можно сказать о значимости предельного эффекта?

Задание 5. Ниже представлены результаты оценивания линейной регрессионной модели процента детей, получающих образование, (*eduger*), на социально-экономические показатели. *polityIV* — индекс демократии Polity IV (от -10 до 10); *gini_8090* — средний индекс имущественного неравенства в 1980 и 1990 г. (от 0 до 100); *ssw* — консолидированные бюджетные расходы на социальное обеспечение (% ВВП); *prot80* — процент протестантов от численности населения в 1980 г.

	coef	std. error	t	Pr> t	[0.025; 0.975]
(Intercept)	59.0982	9.2578			
<i>polityIV</i>	1.0462	0.4522			
<i>gini_8090</i>	0.2556	0.1952			
<i>ssw</i>	1.4918	0.3066			
<i>prot80</i>	0.1016	0.0576			

ANOVA

	sum_sq	df	mean_sq	f	PR(>F)
Regression					
Residual	5665.644				
Total	14178.832	57			

- Проверьте гипотезу о незначимости коэффициента при предикторе *prot80* против двусторонней альтернативы на фиксированном уровне значимости 0.05. Отметьте необходимое значение критической точки из представленного ниже списка
 - квантиль распределения Стьюдента, 0.95, df = 56: 1.673
 - квантиль распределения Стьюдента, 0.95, df = 57: 1.6721
 - квантиль распределения Стьюдента, 0.95, df = 58: 1.6715
 - квантиль распределения Стьюдента 0.975, df = 56: 2.0032
 - квантиль распределения Стьюдента 0.975, df = 57: 2.0025
 - квантиль распределения Стьюдента 0.975, df = 58: 2.0017
- Проинтерпретируйте оценку коэффициента при предикторе *ssw*
- Рассчитайте коэффициент детерминации и проинтерпретируйте полученное значение
- Проверьте гипотезу, что регрессия на константу не хуже модели с предикторами, на фиксированном уровне значимости 0.01. Запишите значение статистики и ее промежуточные расчеты, а также выберите необходимую критическую точку – квантиль. Сделайте вывод.
 - квантиль хи-квадрат распределения уровня 0.99, df= 57: **84.733**
 - квантиль хи-квадрат распределения уровня 0.01, df = 57: **35.131**
 - квантиль распределения Фишера уровня 0.99, df1 = 5, df2 = 53: **3.384**
 - квантиль распределения Фишера уровня 0.99, df1 = 4, df2 = 53: **3.695**
 - квантиль распределения Фишера уровня 0.01, df1 = 5, df2 = 53: **0.108**

(f) квантиль распределения Фишера уровня 0.01, $df1 = 4$, $df2 = 53$: **0.073**

- Задание 6.**
1. Модель, представленная в предыдущем задании, была проверена на гетероскедастичность посредством теста Уайта. Запишите количество параметров в дополнительной (вспомогательной) модели, которую необходимо построить на промежуточном этапе реализации теста Уайта
 2. Порассуждайте, какие могут быть источники гетероскедастичности в данном случае? Приведите не менее двух, свой ответ поясните