

Задания для тренировки

Задание 1. На основе представленных ниже данных оценивается парная линейная регрессия y на x .

y	x
7	2
4	4
5	3
10	6
9	7
12	9
8	3
9	6

- Чему равен коэффициент корреляции Пирсона между зависимой переменной и предсказанным на основе модели (см. условие выше) значением отклика? Ответьте на поставленный вопрос без предварительных расчетов предсказанных значений зависимой переменной. Свое решение обоснуйте
- Рассчитайте коэффициент детерминации для указанной регрессионной модели. Проинтерпретируйте полученное значение
- Проверьте значимость коэффициента детерминации

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.

x	2	5	2	0	1
y	10	3	1	12	5

2. Рассчитайте предсказанное значение зависимой переменной для первого наблюдения.

Задание 3. Ниже представлены результаты оценивания линейной регрессионной модели расходов на здравоохранение в % от ВВП (healthexp), на такие показатели, как

- opp - индекс оппозиции, построенный на основе показателей «Freedom of Assembly and Association» (CIRI database), legislative, executive electoral competitiveness (Database of Political Institutions, the World Bank). Непрерывная переменная от 0 до 100, где 100 соответствует более высокому уровню оппозиции
- ji - Linzer and Staton's measure of latent judicial independence, большие значения соответствуют более высокой степени независимости
- fh - свобода прессы (Freedom of the Press, Freedom House). Шкала перевернута для удобства, так что большему значению соответствует большее значение свободы прессы
- pop_rural - % сельского населения
- mort_u5 - коэффициент смертности детей до 5 лет
- ln_gdp_pc - логарифм ВВП на душу населения

Coefficients:

	coef	std. error	t	Pr> t	[0.025; 0.975]
(Intercept)	5.641894	1.339102			
opp	-0.001480	0.004568			
ji	5.536259	0.704676			
fh	-0.022972	0.007312			
pop_rural	-0.015421	0.007720			
mort_u5	0.015118	0.002309			
ln_gdp_pc	-0.285310	0.133156			

Multiple R-squared: _____

ANOVA

	sum_sq	df	mean_sq	f	PR(>F)
Regression					
Residual	2877.76				
Total	3384.37	637			

1. Постройте 95%-ый доверительный интервал для коэффициента при предикторе *ji*. Так как наблюдений достаточно много, Вы можете использовать при расчете процентных точек нормальную аппроксимацию. Однако в ответе обязательно запишите количество степеней свободы по распределению Стьюдента
2. Проверьте гипотезу о незначимости коэффициента при предикторе *pop_rural* на фиксированном уровне значимости 0.01 против двусторонней альтернативы. Укажите значение необходимой процентной точки. Так как наблюдений достаточно много, Вы можете использовать при расчете процентных точек нормальную аппроксимацию. Однако в ответе обязательно запишите количество степеней свободы по распределению Стьюдента. Приведите промежуточные расчеты, сделайте вывод
3. Рассчитайте коэффициент детерминации

Задание 4.

Ниже представлены оценки регрессионной модели. Зависимая переменная – доля граждан, имеющих наиболее высокий уровень удовлетворенности жизнью. Качество институтов (исходный показатель) измеряется в непрерывной шкале от 0 до 5, где 5 соответствует максимальному значению качества институтов. В модели используется преобразованное значение качества институтов: центрированный показатель (Inst_c). Исследователь сравнивает западноевропейские и латиноамериканские страны. Для групп стран введена дамми-переменная (LA), которая принимает значение 0, если страна – западноевропейская, значение 1 – для латиноамериканской страны.

Life Satisfaction	
Inst_c	0.48*** (5.2)
LA	0.163*** (6.23)
LA × Inst_c	0.04*** (4.24)
Intercept	0.3*** (9.53)

t-statistics are given in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

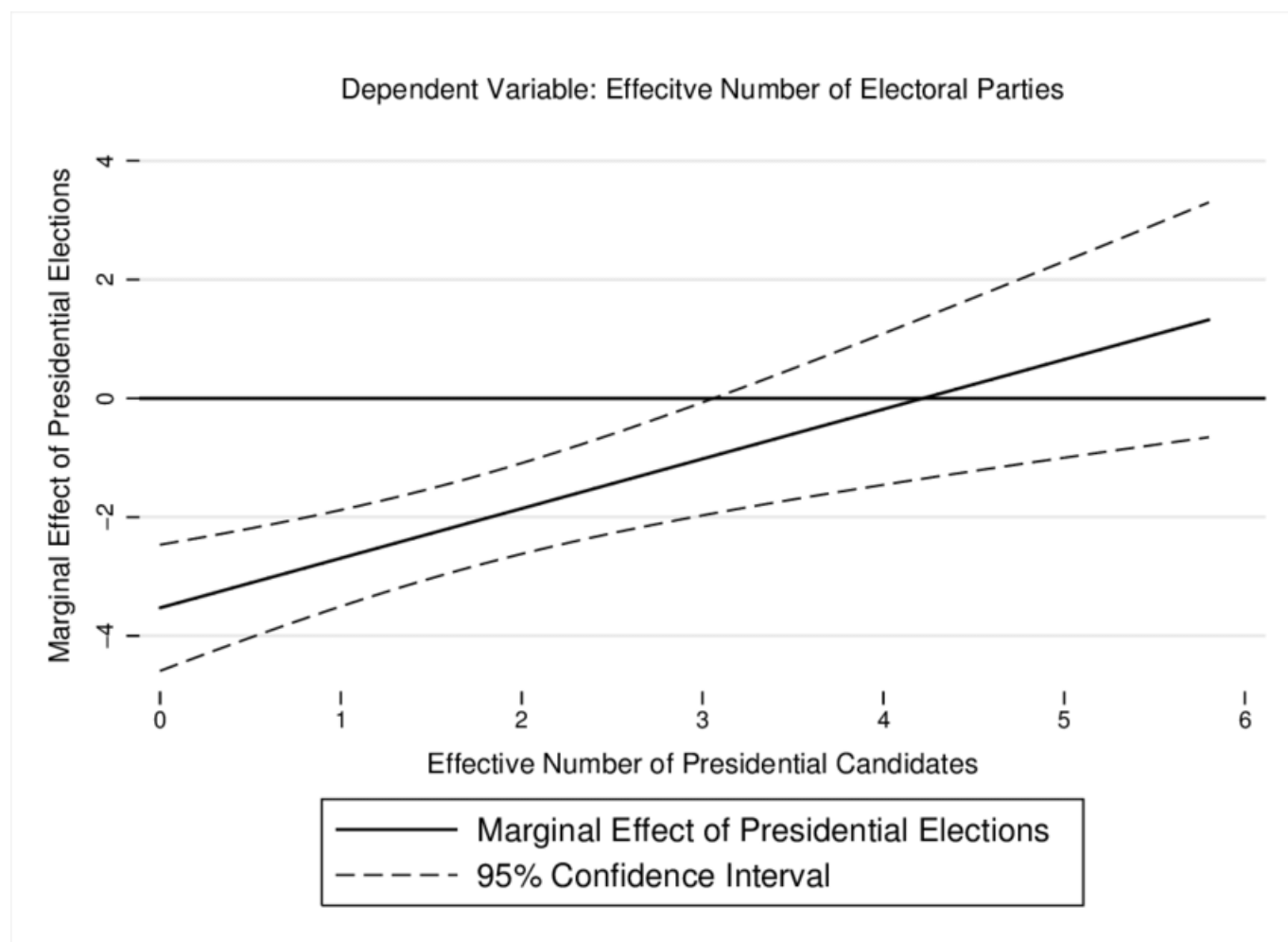
1. Проинтерпретируйте оценку коэффициента при переменной LA.
2. Проинтерпретируйте оценку коэффициента при переменной взаимодействия.
3. Рассчитайте значение предельного эффекта качества политических институтов в случае, если рассматривается латиноамериканская страна.

Задание 5. Ниже представлена спецификация модели, отражающая взаимосвязь трат на еду и месячного дохода. Переменные измерены в единицах – «фунтиках».

$$food_exp_i = 20 + 2.5income_i - 0.25income_i^2 + \epsilon_i$$

1. Рассчитайте предельный эффект месячного дохода на траты на еду, если месячный доход составляет 10 «фунтиков».
2. Проинтерпретируйте оценку коэффициента при предикторе $income^2$
3. Вычислите «переломную» точку (значение показателя $income$), после которой характер взаимосвязи $income$ и зависимой переменной меняет знак. В ответе представьте необходимые расчеты.

Задание 6. На основе оценивания регрессионной модели с переменной взаимодействия был построен следующий график зависимости предельного эффекта ключевого предиктора от значений модератора:



При каких значениях модератора предельный эффект является значимым (уровень доверия 0.95)? Если таковых значений на графике нет, напишите об этом, кратко пояснив свой ответ.

Задание 7. Объясните нижеприведенное утверждение. А также укажите, на что в таком случае должен опираться исследователь?

«The analyst cannot infer whether X has a meaningful conditional effect on Y from the magnitude and significance of the coefficient on the interaction term.»

Задание 8. Прочитайте нижеприведенный отрывок из статьи и продолжите его, аргументировав позицию авторов:

Much of the concern about multicollinearity arises when the analyst observes that the coefficients from a linear-additive model change when an interaction term is introduced. In the linear-additive world, the sensitivity of results to the inclusion of an additional variable is often taken as a sign of multicollinearity. However, this need not be the case with interaction models.

Задание 9. Ознакомьтесь с нижеприведенной выдачей (корреляционная матрица и оценки коэффициентов линейной регрессии):

	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

	coef	std.error	t	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:

	x	z	w
	2.154286	12.03485	7.954813

1. Можно ли говорить о том, что модель страдает от мультиколлинеарности? Приведите не менее двух свидетельств из выдачи в качестве обоснования. Можно приводить как прямые, так и косвенные свидетельства мультиколлинеарности.
2. Проинтерпретируйте значение VIF для переменной w. Что оно показывает?