

Демо-версия экзаменационной работы

Задание 1. Выберите верный(-е) вариант(-ы) ответа. Если верных утверждений нет, то напишите в ответе слово «НЕТ»:

- К-фолдовая кросс-валидация
 1. решает проблему переобучения модели
 2. обеспечивает статистическую независимость фолдов
 3. предполагает, что в каждой из k итераций проверки один фолд используется как тестовая выборка, а остальные (k-1) фолдов объединяются в обучающую выборку
 4. предполагает, что большее количество фолдов дает менее смещенную оценку обобщающей способности модели
- Сравниваются две линейные вложенные модели при помощи F-теста. Известно, что p-value составило 0.43. На основании этого можно сказать, что нужно выбрать
 1. модель менее экономную
 2. модель более экономную
 3. нет разницы между этими моделями, можно выбрать любую
 4. в этом случае нельзя использовать F-тест, более корректно будет обратиться к тесту отношения правдоподобия
- Стьюдинизированные остатки
 1. принимают значения от -1 до 1
 2. используются для диагностики нетипичных наблюдений по предиктору
 3. используются для диагностики нетипичных наблюдений по отклику
 4. более устойчивы к выбросам, чем обычные стандартизованные остатки
- Параллельность регрессий в логистической регрессионной модели с порядковым откликом предполагает, что
 1. предсказанные вероятности одинаковы для всех категорий
 2. модель, оцененная в этом допущении, более экономна по сравнению с соответствующей мультиномиальной моделью
 3. эффект предиктора одинаков при сравнении $Y \leq k$ и $Y > k$ для любого k
 4. графически линии логитов для разных пороговых значений совпадают
- В условиях гетероскедастичности
 1. оценки коэффициентов регрессионной модели, оцененной при помощи МНК, становятся смещенными
 2. оценки коэффициенты регрессионной модели, оцененной при помощи МНК, остаются BLUE
 3. стандартные ошибки оценок коэффициентов регрессионной модели, оцененной при помощи МНК, всегда занижены
 4. оценки коэффициенты регрессионной модели, оцененной при помощи МНК, остаются состоятельными

Задание 2. Ниже представлены результаты анализа разложения вариации по линейной регрессионной модели, построенной по выборке 161 наблюдений.

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	236.833	236.833		
x2	1	77.850	77.850		
x3	1	159.474	159.474		
x4	1	15.318	15.318		
x5	1	3.141	3.141		
Residuals	...	144.703	...		

Используя информацию из данной таблицы,

1. рассчитайте значение коэффициента детерминации
2. рассчитайте значение необходимой статистики критерия для проверки гипотезы о незначимости коэффициента детерминации
3. сделайте вывод на основе значения p-value из соответствующего F-теста, если известно, что $p\text{-value} \approx 0$

Задание 3. Ниже представлена спецификация модели, отражающая взаимосвязь трат на еду и месячного дохода. Переменные измерены в единицах – «фунтиках»

$$\widehat{food_exp}_i = 20 + 0.35income_i - 0.004income_i^2$$

1. Изобразите схематично взаимосвязь трат на еду и месячного дохода, опираясь на оценки представленной модели
2. Рассчитайте «пороговую» точку: начиная с какого дохода рассматриваемая взаимосвязь меняет знак?
3. Проинтерпретируйте оценку коэффициента при $income_i^2$

Задание 4. Известно, что p-value, рассчитанное по плацебо-тесту, составляет 0.15, объясните суть плацебо-теста, запишите последовательные шаги его реализации и проинтерпретируйте полученное значение p-value

Задание 5. Модель со случайными эффектами была оценена посредством метода FGLS. В рамках формулы для получения оценок коэффициентов в этой модели используется ковариационная матрица случайных ошибок. Обозначим дисперсию случайного эффекта как σ_α^2 , а дисперсию ошибок ϵ_{it} как σ_ϵ^2 . Запишите ковариационную матрицу ошибок для одной заданной пространственной единицы, прокомментируйте, на какие допущения модели со случайными эффектами Вы опирались

Задание 6. Ниже представлена таблица, представляющая результаты исследования S. Galiani, E. Schargrodsky «Effects of Land Titling on Child Health». Ознакомьтесь также с аннотацией данной статьи:

This paper examines the impact of urban land titling (оформление прав собственности на землю) on child health. We hypothesize that land titling may translate into positive effects on health through its impact on housing investments and household structure. To address selection concerns (для решения проблемы смещения отбора), we take advantage of a natural experiment (естественный эксперимент) of land occupation in a suburban area of Buenos Aires, Argentina, that ensures that the allocation of property rights is exogenous to the characteristics of the squatters (распределение прав собственности экзогенно по отношению к характеристикам поселенцев).

Weight-for-height Z-score

Age group	Property right offer = 1	Property right offer = 0	Difference
All (0–11 years old)	0.279 (0.065) [239]	0.065 (0.087) [132]	0.214 (0.109)
0–4 years old	0.110 (0.100) [114]	-0.028 (0.119) [60]	0.139 (0.156)
5–11 years old	0.434 (0.083) [125]	0.143 (0.126) [72]	0.291 (0.151)

Note: Weight-for-height Z-scores are available for boys through 138 months of age and less than 145 cm of height, and girls through 120 months of age and less than 137 cm of height. The sample is restricted to children born after 1991. Standard errors in parentheses. Number of observations in brackets.

Запишите в общем виде спецификацию регрессионной модели, оценивающей различия в характере связи прав собственности и здоровья детей между возрастными группами от 0 до 4 лет и от 5 до 11 лет

Задание 7. Пояснение к выдаче мультиномиальной логистической модели представлено ниже:
prioInt: 2 — есть серьезный конфликт (случилось более 1000 смертей от конфликтов в год), 1 — слабый конфликт конфликт (случилось больше 25 и меньше 1000 смертей), 0 — отсутствие конфликта.
prioIntLag: переменная prioInt с лагом в один год.
mount: процент горной местности.

PrioInt:		
	1	2
mount	0.003 (0.004)	0.016*** (0.006)
prioIntLag	2.039*** (0.153)	3.303*** (0.208)

Note: *p<0.1; **p<0.05; ***p<0.01

Проинтерпретируйте оценки коэффициентов при mount в терминах отношения шансов (рисков)