

| Тест | 1 | 2 | 3 | Итого |
|------|---|---|---|-------|
| | | | | |

← для проверяющего!

Фамилия, имя, номер группы:

.....

Ответы на тест:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|
| | | | | | | | | | |

Тест

Вопрос 1. Использование МНК к регрессии с бинарной зависимой переменной приведет к возникновению:

- ☐ А Гетероскедастичности остатков
- ☐ Б Незначимости всей регрессии
- ☐ В Мультиколлинеарности в модели
- ☐ Г Остатки модели будут иметь нормальное распределение
- ☐ Д нет верного ответа

Вопрос 2. В качестве функции правдоподобия для оценки ММП парной регрессионной модели выступает функция:

- ☐ А $L(\beta_0; \beta_1) = \prod_{i=1}^n (p_i^{y_i} + (1 - p_i)^{1-y_i})$
- ☐ Б $L(\beta_0; \beta_1) = \prod_{i=1}^n (p_i^{y_i} \cdot (1 - p_i)^{1-y_i})$
- ☐ В $L(\beta_0; \beta_1) = \prod_{i=1}^n (p_i^{y_i} - (1 - p_i)^{1-y_i})$
- ☐ Г $L(\beta_0; \beta_1) = \prod_{i=1}^n (1 - p_i)^{1-y_i}$
- ☐ Д нет верного ответа

Вопрос 3. Была оценена логистическая регрессия зависимости вероятности просрочки (1 — есть просрочка, 0 — нет) по кредиту в зависимости от возраста заемщика (Age):

$$\mathbb{P}(Y = 1) = F(Z), Z = -2,101 - 0,025 \cdot \text{Age} + u$$

Абсолютная разница в вероятности просрочки для заемщика 36 лет и заемщика 55 лет, округленная до сотых, составляет:

- ☐ А нельзя найти по имеющимся данным
- ☐ Б 0, отсутствует
- ☐ В 0.5
- ☐ Г 0.02
- ☐ Д 0.05
- ☐ Е нет верного ответа

Вопрос 4. Истинная зависимость имеет вид $Y_i = \beta_0 + \beta_1 \cdot Z_i + v_i$. При этом Z_i измеряется с ошибкой: $Z_i^{obs} = Z_i + w_i$. Известно, что $\beta_1 = -0,4$, $\sigma_w^2 = 6$, $\sigma_z^2 = 3$, $\text{Cov}(w_i, v_i) = 0$. Исследователь оценивает регрессию $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot Z_i^{obs}$. Предел по вероятности оценки $\hat{\beta}_1$ будет отличаться от истинного значения параметра на

- ☐ A $-0.1(3)$
☐ C $0.2(6)$
☐ E Оценка не будет асимптотически смещена
- ☐ B $0.1(3)$
☐ D $-0.2(6)$
☐ F нет верного ответа

Вопрос 5. Валидность инструмента Z_i в модели $Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$ обозначает:

- ☐ A Инструмент Z_i коррелирует с X_i
☐ C Инструмент Z_i не коррелирует с X_i
☐ E В модели есть эндогенность
- ☐ B Инструмент Z_i не коррелирует с ошибкой
 ☐ D Инструмент Z_i коррелирует с ошибкой
 ☐ F нет верного ответа

Вопрос 6. В линейной модели $Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$ регрессор X_i является эндогенным. Состоятельные оценки коэффициентов можно получить с помощью

- ☐ A МНК
 ☐ C Взвешенного МНК
 ☐ E переменных
- ☐ B Обобщенного МНК
 ☐ D Метода инструментальных
 ☐ E нет верного ответа

Вопрос 7. Известно, что $Y_i = \beta_0 + \epsilon_i$, при этом $\text{Var}(\epsilon_i) = i^2$. Какая из этих оценок β_0 будет эффективной?

- ☐ A $\frac{\sum_{i=1}^n \frac{Y_i}{i^2}}{\sum_{i=1}^n \frac{1}{i^2}}$
☐ D $\overline{\left(\frac{1}{i^2}\right)}$
- ☐ B $\frac{\sum_{i=1}^n \frac{Y_i}{i}}{\sum_{i=1}^n \frac{1}{i}}$
☐ E \bar{Y}
- ☐ C $\overline{\left(\frac{1}{i}\right)}$
☐ F нет верного ответа

Вопрос 8. Какой из этих тестов на гетероскедастичность не требует выбора переменной, по которой подозревается гетероскедастичность:

- ☐ A Тест Уайта
 ☐ C Тест Глейзера
 ☐ E Тест Хаусмана
- ☐ B Тест Голдфелда-Куандта
 ☐ D Тест Дарбина-Уотсона
 ☐ F нет верного ответа

Вопрос 9. При использовании МНК оценок параметров регрессионного уравнения и робастных ошибок в форме Уайта,

- ☐ A Оценки $\hat{\beta}$ будут состоятельными и неэффективными, доверительные интервалы, полученные по $\widehat{\text{Var}}_{HCE}(\hat{\beta})$ можно использовать
- ☐ B Оценки $\hat{\beta}$ будут состоятельными и эффективными, доверительные интервалы, полученные по $\widehat{\text{Var}}_{HCE}(\hat{\beta})$ можно использовать
- ☐ C Оценки $\hat{\beta}$ будут несостоятельными и неэффективными, доверительные интервалы, полученные по $\widehat{\text{Var}}_{HCE}(\hat{\beta})$ нельзя использовать
- ☐ D Оценки $\hat{\beta}$ будут несостоятельными и неэффективными, доверительные интервалы, полученные по $\widehat{\text{Var}}_{HCE}(\hat{\beta})$ можно использовать

- ☐ E Оценки $\hat{\beta}$ будут состоятельными и эффективными, доверительные интервалы, полученные по $\widehat{Var}_{HCE}(\hat{\beta})$ нельзя использовать
- ☐ F нет верного ответа

Вопрос 10. Для модели $Y = X\beta + \epsilon$ с $\text{Var}(\epsilon) = \Omega \neq \sigma_\epsilon^2 I$ где I – единичная матрица, эффективные оценки параметров β можно получить с помощью критерия:

- ☐ A $\min(Y - X\beta)' \Omega (Y - X\beta)$ ☐ C $\min(Y - X\beta)' (Y - X\beta)$ ☐ E Минимизация невозможна
- ☐ B $\min(Y - X\beta)' \Omega^{-1} (Y - X\beta)$ ☐ D $\min(Y - X\beta)^2$ ☐ F нет верного ответа

Вопрос 11. Если функция плотности удовлетворяет условиям регулярности, то оценки метода максимального правдоподобия являются

- ☐ A несмещенными ☐ C неотрицательными ☐ E равномерно распределенными
- ☐ B несостоятельными ☐ D инвариантными ☐ F нет верного ответа

Вопрос 12. Тест Саргана для проверки валидности инструментов можно использовать, если число эндогенных переменных среди объясняющих

- ☐ A Больше числа экзогенных переменных ☐ C Меньше числа инструментов ☐ E Меньше 3
- ☐ B Больше числа инструментов ☐ D Не превышает 10 ☐ F нет верного ответа

Вопрос 13. Если обобщенный метод моментов будет применен в случае, когда число используемых моментов совпадает с числом оцениваемых параметров, то минимизируемая функция в точке оптимума

- ☐ A Равна нулю ☐ C Меньше нуля ☐ E Равна числу моментных тождеств
- ☐ B Больше нуля ☐ D Может быть как больше нуля, так и меньше нуля ☐ F нет верного ответа

Вопрос 14. При проверке гипотезы $H_0: g(\beta) = 0$ для параметров модели $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \epsilon_i$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ с помощью теста множителей Лагранжа, необходимо знать оценки параметров

- ☐ A Регрессии на константу ☐ C Только модели без ограничений ☐ E Как модели с ограничениями, так и модели без ограничений
- ☐ B Регрессии на все факторы, кроме константы ☐ D Только модели с ограничениями ☐ F нет верного ответа

Вопрос 15. Для проверки значимости коэффициента регрессии $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \epsilon_i$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, оцененной с помощью ММП по n наблюдениям, исследователь использует LR статистику. Она имеет распределение

☐ A точно $\mathcal{N}(0, 1)$ ☐ C t_{n-4} ☐ E асимптотически χ^2_{n-4} ☐ B асимптотически $\mathcal{N}(0, 1)$ ☐ D асимптотически χ^2_1 ☐ F нет верного ответа

Фамилия, имя, номер группы:

.....

Задачи

- В линейной модели $Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$ регрессор X коррелирован с ошибкой: $\text{corr}(X_i, \epsilon_i) \neq 0$.
 - Объясните, каким образом для данной модели можно получить состоятельные оценки с помощью двухшагового МНК. Какие модели необходимо оценить на каждом шаге?
 - Покажите, что в случае с одним регрессором оценка коэффициента β_1 , полученная с помощью двухшагового МНК и одним инструментом, эквивалентна IV-оценке.
- Илон Маск оценивает два параметра по выборке из 1000 наблюдений методом максимального правдоподобия. Известна логарифмическая функция правдоподобия:

$$\ell(\gamma, \beta) = -6\gamma^2 - 4\beta^2 + \gamma\beta - \beta + 3$$
 - Найдите оценки параметров γ и β методом максимального правдоподобия
 - Найдите оценку информационной матрицы Фишера.
 - Постройте 95%-й интервал для параметра β
 - С помощью LM-теста проверьте гипотезу $\gamma = 1$ на уровне значимости 5
 - С помощью теста Вальда проверьте гипотезу $\gamma = \beta$ на уровне значимости 5

Критические значения хи-квадрат распределения для 5%-го уровня значимости равны 3.8, 6.0, 7.8, 9.5, 11.1 (для 1-5 степеней свободы).

- По данным открытого скрининга здоровья людей (2730 съёмов) на наличие диабета второго типа ($Y = 0$ если диабета нет и 1, если диабет обнаружен) была построена логит-модель вероятности наличия диабета в зависимости от параметров:

пол (gen, 0 — М, 1 — Ж), возраст (полных лет, от 18-ти), частота сердечного ритма, пульс (ударов в минуту, pulse)

Результаты оценки модели представлены ниже:

| | $\hat{\beta}$ | $se(\hat{\beta})$ |
|-------|---------------|-------------------|
| age | 0.07 | 0.005 |
| pulse | 0.1 | 0.265 |
| gen | -0.3 | 0.103 |
| const | -8.7 | 0.611 |

Ниже дана таблица классификации при пороге отсечения 0,25:

| | $\hat{Y}_i = 0$ | $\hat{Y}_i = 1$ |
|-----------|-----------------|-----------------|
| $Y_i = 0$ | 1560 | 600 |
| $Y_i = 1$ | 220 | 350 |

- а) Какие проблемы возникнут при рассмотрении линейной регрессии вместо логистической в этой модели?
 - б) Рассчитайте значения чувствительности и специфичности данной модели
 - в) Посчитайте вероятность наличия диабета для мужчины возрастом 60 лет, пульсом 80.
 - г) Оцените предельный эффект увеличения возраста для женщины 43 лет и со значением пульса 80. Кратко, одной-двумя фразами, прокомментируйте смысл полученных цифр.
4. Известно, что количество решённых задач по экзамену по Очень Сложному Предмету зависит от количества выпитого накануне кофе и любви к котикам (обе переменные непрерывные, даже любовь к котикам! И вообще, любовь к котикам может быть даже бесконечной, ведь нет предела совершенству!). То есть, истинная зависимость имеет вид:

$$Np_i = \beta_0 + \beta_1 Coffee_i + \beta_2 Cats_i + \epsilon_i$$

Исследователь Вениамин разделил выборку на три части по степени любви к кофе и оценил отдельные регрессии для каждой из трёх подвыборок. Известно, что в выборку кофеманов вошло 33 человека, в выборку кофеманов 23, а неопределившихся оказалось 44. После этого Вениамин провёл тест Голдфелда-Куандта и хотел было вписать его результаты в текст исследования, но тут, к его ужасу, прибежала кошка и съела часть цифр! Помогите Вениамину восстановить их!

- а) Зная, что значение тестовой статистики равно 6 и RSS в группе кофеманов выше, скажите, какие степени свободы имела тестовая статистика
- б) Найдите RSS_{coffee_mans} , зная, что $RSS_{coffee_mans} = 30$
- в) Предложите способ получить эффективные оценки $\hat{\beta}$, если известно, что $\sigma_{coffee_mans}^2 = 1$, $\sigma_{neutrals}^2 = 2$, $\sigma_{coffee_mans}^2 = 10$.