

## Часть 1. Тест.

**Вопрос 1 ♣** Предпосылка об отсутствии систематической ошибки в модели означает, что для всех наблюдений

☐ A  $\text{Var}(\varepsilon_i) \neq 0$

☐ C Нет верного ответа.

☐ B  $\text{Var}(\varepsilon_i) = 0$

**Вопрос 2 ♣** После применения МНК к модели  $y_i = \beta x_i + \varepsilon_i$  сумма  $ESS + RSS$

☐ A обязательно равна  $TSS$

☐ B может быть не равна  $TSS$

**Вопрос 3 ♣** При наличии ошибок измерения зависимой переменной МНК-оценки коэффициентов модели

☐ A состоятельны

☐ B несостоятельны

**Вопрос 4 ♣** Индексы вздутия дисперсии (VIF) в случае отсутствия мультиколлинеарности лежат в интервале

☐ A  $[0; 1]$

☐ B  $[1; +\infty)$

**Вопрос 5 ♣** Незначимость всех коэффициентов регрессии

☐ A может быть не связана с мультиколлинеарностью

☐ B обязательно свидетельствует о наличии мультиколлинеарности

**Вопрос 6 ♣** Если выполнены все предпосылки теоремы Гаусса-Маркова, но остатки модели не подчиняются нормальному закону распределения, то МНК-оценки коэффициентов регрессии являются

☐ A несмещёнными

☐ B смещёнными

**Вопрос 7 ♣** Если в модель добавили незначимый фактор, то коэффициент детерминации  $R^2$

☐ A не изменится

☐ B упадёт

☐ C вырастет

**Вопрос 8 ♣** Нулевая гипотеза в тесте Дарбина-Уотсона состоит в

☐ A отсутствии автокорреляции

☐ B наличии автокорреляции

**Вопрос 9 ♣** При диагностике автокорреляции третьего порядка тест Бройша-Годфри

☐ A применим

☐ B неприменим

**Вопрос 10 ♣** Стандартные ошибки в форме Уайта в случае гетероскедастичности помогают устранить несостоятельность оценок коэффициентов

☐ A верно

☐ B неверно

## Часть 2. Задачи.

1. На основании опроса была оценена следующая модель:

$$\ln(wage_i) = \beta_1 + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 married_i + \beta_5 educ_i + \beta_6 black_i + \varepsilon_i$$

где:

- $wage_i$  — величина заработной платы в долларах
- $exper_i$  — опыт работы в годах
- $educ_i$  — количество лет обучения
- $married_i$  — наличие супруга/супруги (1 — есть, 0 — нет)
- $black_i$  — принадлежность к негроидной расе (1 — да, 0 — нет)

Показатель	Значение
$R^2$	<b>B7</b>
Скорректированный $R^2$	0.219
Стандартная ошибка регрессии	<b>B6</b>
Количество наблюдений	<b>B2</b>

Результаты дисперсионного анализа:

	df	SS	MS	F	P-значение
Регрессия	<b>B1</b>	5.993	1.199	<b>B5</b>	0.000
Остаток	134	18.240	0.136		
Итого	<b>B3</b>	<b>B4</b>			

Коэффициент	Оценка	$se(\hat{\beta})$	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Константа	4.529	0.331	13.688	0.000	3.874	5.183
$exper$	0.090	0.037	2.419	0.017	0.016	0.164
$exper^2$	-0.003	0.002	-1.790	0.076	-0.006	0.000
$married$	0.240	0.079	3.045	0.003	<b>B8</b>	<b>B9</b>
$educ$	0.078	0.017	<b>B10</b>	0.000	0.045	0.111
$black$	0.073	0.171	0.424	0.672	-0.266	0.411

Найдите пропущенные числа **B1–B10**.

Ответ округляйте до 3-х знаков после запятой. Кратко поясняйте, например, формулой, как были получены результаты.

2. На основании данных по ценам на квартиры в Москве были построена модель

$$\ln(price_i) = \beta_1 + \beta_2 totsp_i + \beta_3 metrdist_i + \beta_4 dist_i + \beta_5 floor_i + \varepsilon_i,$$

где:

- $\ln(price_i)$  — логарифм цены квартиры в тысячах долларов
- $totsp_i$  — общая площадь квартиры в кв. м.
- $metrdist_i$  — расстояние до метро в минутах
- $dist_i$  — расстояние до центра города в км
- $floor_i$  — дамми-переменная (1 — если квартира не на первом и последнем этажах, 0 — иначе)

Модели были оценены на пяти разных выборках, результаты представлены в таблице:

Коэффициент	Выборка А	Выборка В	Выборка С	Выборка D	Выборка Е
Константа	3.980***	3.926***	3.929***	3.719***	4.224***
$totsp$	0.0155***	0.0148***	0.0163***	0.0179***	0.0139***
$metrdist$	-0.00858***	-0.0169***	-0.00566**	-0.0108***	-0.0077
$dist$	-0.0267***	-0.0186***	-0.0253***	-0.0150***	-0.0350***
$floor$	0.0419**	0.0633*	0.0224	0.0225	0.0228
Наблюдений	460	145	315	150	150
$R^2$	0.693	0.684	0.723	0.328	0.520
$RSS$	15.120	4.503	9.408	2.163	8.545

\* — значимость на 10%, \*\* — значимость на 5%, \*\*\* — значимость на 1%.

- Для всей выборки (выборка А) проинтерпретируйте коэффициент при переменной  $dist_i$ .
- Определите на 5%-ом уровне значимости, можно ли использовать одну модель для квартир, находящихся в пешей доступности от метро (выборка С), и квартир, находящихся в транспортной доступности (выборка В).
- Исследователь предположил, что дисперсия ошибок модели возрастает с увеличением площади квартиры. Проверьте, есть ли в модели гетероскедастичность на 10% уровне значимости на основании соответствующего теста. В выборку D включены 150 квартир с наименьшей общей площадью, в выборку Е — 150 квартир с наибольшей общей площадью.

При проверке гипотез: выпишите  $H_0$ ,  $H_a$ , найдите значение тестовой статистики, укажите её распределение, найдите критическое значение, сделайте выводы

3. По ежемесячным данным, 146 наблюдений, была оценена зависимость

$$\widehat{credit}_t = 362.21 - 7.50r\_credit_t - 13.09ipc_t, R^2 = 0.44$$

где:

- $credit_t$  — объём потребительских кредитов, выданных домашним хозяйствам РФ
- $r\_credit_t$  — ставка процента по кредитам
- $ipc_t$  — индекс потребительских цен

Известно, что  $\sum_{t=2}^{146} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 = 266491$ ,  $\sum_{t=1}^{146} \hat{\varepsilon}_t^2 = 438952$ ,  $\sum_{t=2}^{146} |\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}| = 3617$ ,  $\sum_{t=1}^{146} |\hat{\varepsilon}_t| = 6382$ .

Кроме того, была оценена вспомогательная модель для остатков исходной модели:

$$\hat{\varepsilon}_t = 15.67 - 0.75r\_credit_t + 0.02ipc_t + 0.39\hat{\varepsilon}_{t-1} + 0.21\hat{\varepsilon}_{t-2} + 0.24\hat{\varepsilon}_{t-3}, R^2 = 0.56$$

- а) На 1%-ом уровне значимости проверьте гипотезу об адекватности исходной регрессии
- б) Проведите тест Дарбина-Уотсона на 5%-ом уровне значимости
- в) Проведите тест Бройша-Годфри на 5%-ом уровне значимости

При проверке гипотез: выпишите  $H_0$ ,  $H_a$ , найдите значение тестовой статистики, укажите её распределение, найдите критическое значение, сделайте выводы

4. Домохозяйка Глаша очень любит читать романы Л.Н. Толстого и смотреть сериалы. Её сын Петя учится на третьем курсе ВШЭ. Последние 30 дней он записывал, сколько Глаша прочитала страниц «Анны Карениной»,  $pages_t$ , и посмотрела серий «Доктора Хауса»,  $series_t$ . На основании этих наблюдений при помощи МНК Петя оценил следующую модель:

$$\widehat{pages}_t = 100 - 3series_t$$

Оценка ковариационной матрицы коэффициентов,  $\widehat{\text{Var}}(\hat{\beta}) = \begin{pmatrix} 11 & 0.5 \\ 0.5 & 1 \end{pmatrix}$

Оценка дисперсии ошибок равна  $\hat{\sigma}^2 = 323$ .

Завтра Глаша собирается посмотреть 10 серий «Доктора Хауса».

- а) Постройте точечный прогноз количества прочитанных Глашей страниц романа
- б) Постройте 95%-ый доверительный интервал для  $E(pages_t | series_t = 10)$ , ожидаемого количества прочитанных страниц
- в) Постройте 95%-ый предиктивный интервал для фактического количества прочитанных страниц

### Часть 3. Теоретические вопросы

5. Опишите МНК для парной регрессии: выпишите целевую функцию, систему нормальных уравнений, оценки коэффициентов, оценки дисперсий коэффициентов.
6. Сформулируйте теорему Гаусса-Маркова для детерминированных регрессоров.
7. Опишите тест Уайта: сформулируйте нулевую и альтернативную гипотезы, способ получения тестовой статистики, её распределение при верной нулевой гипотезе, вид критической области.