

## Часть 1. Тест.

**Вопрос 1 ♣** В множественной регрессии с константой матрица  $X$  имеет размер  $56 \times 6$ ,  $TSS = 1100$ ,  $ESS = 100$ . Несмещённая оценка для дисперсии случайных ошибок модели равна

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> A 18.5        | <input type="checkbox"/> D 20          | <input type="checkbox"/> G Нет верного ответа. |
| <input type="checkbox"/> B 30          | <input type="checkbox"/> E 40          |  |
| <input type="checkbox"/> C $\sqrt{30}$ | <input type="checkbox"/> F $\sqrt{40}$ |  |

**Вопрос 2 ♣** Коэффициент  $R^2$  может быть представлен в виде

- |  |  |
|--|--|
| <input type="checkbox"/> A $\sum_{j=2}^k \hat{\beta}_j \frac{\widehat{\text{Cov}}(X_j, Y)}{\widehat{\text{Var}}(Y)}$       | <input type="checkbox"/> D $\sum_{j=2}^k \beta_j \frac{\widehat{\text{Var}}(Y)}{\widehat{\text{Cov}}(X_j, Y)}$       |
| <input type="checkbox"/> B $\sum_{j=2}^k \hat{\beta}_j \frac{\widehat{\text{Cov}}(X_j, Y)}{\widehat{\text{Var}}(\hat{Y})}$ | <input type="checkbox"/> E $\sum_{j=2}^k \hat{\beta}_j \frac{\widehat{\text{Var}}(Y)}{\widehat{\text{Cov}}(X_j, Y)}$ |
| <input type="checkbox"/> C $\sum_{j=2}^k \beta_j \frac{\widehat{\text{Cov}}(X_j, Y)}{\widehat{\text{Var}}(\hat{Y})}$       | <input type="checkbox"/> F $\sum_{j=2}^k \hat{\beta}_j \frac{\widehat{\text{Cov}}(X_j, Y)}{\widehat{\text{Var}}(Y)}$ |
|  | <input type="checkbox"/> G Нет верного ответа.   |

**Вопрос 3 ♣** По данным 570 индивидуумов оценили зависимость почасовой оплаты в долларах,  $EARN$ , от длительности обучения индивидуума в годах,  $S$ , от результата тестирования индивида в баллах,  $ASVABC$ , и пола индивидуума,  $MALE$ , (1 — для мужчин, 0 — для женщин):

$$\ln \widehat{EARN}_i = 0.904 + \underset{(0.124)}{0.01 S_i} + \underset{(0.056)}{0.0157 ASVABC_i} + \underset{(0.1)}{0.27 MALE_i}$$

- |   |  |
|---|--|
| <input type="checkbox"/> A не отличается от оплаты труда женщин | <input type="checkbox"/> D больше на 27 процентов  |
| <input type="checkbox"/> B больше на 0.27 доллара               | <input type="checkbox"/> E больше на 0.27 процента |
| <input type="checkbox"/> C больше на 27 долларов                | <input type="checkbox"/> F Нет верного ответа.     |

**Вопрос 4 ♣** Зависимость спроса на некоторый вид услуг  $Y$  от его цены имеет вид  $\hat{Y}_i = 30 - \underset{(2)}{0.03} \underset{(0.001)}{P_i}$ . Чтобы спрос в среднем снизился на 3% цена должна увеличиться на

- |                                       |                                      |  |
|---------------------------------------|--------------------------------------|--|
| <input type="checkbox"/> A 10%        | <input type="checkbox"/> C 1%        | <input type="checkbox"/> E 10 единиц           |
| <input type="checkbox"/> B 100 единиц | <input type="checkbox"/> D 1 единицу | <input type="checkbox"/> F Нет верного ответа. |

**Вопрос 5 ♣** По данным 500 индивидуумов оценили зависимость веса индивидуума  $Y$ , измеряемого в фунтах (1 фунт  $\approx 0.5$  кг) от его роста, измеряемого в футах (1 фут  $\approx 30$  см):  $\ln \widehat{Y}_i = -4.4 + 2.4 \ln X_i$ . Если рост индивидуума будет измерен в метрах, а вес — в килограммах, то коэффициент перед  $\ln X$  будет равен

- |                                |                                |  |
|--------------------------------|--------------------------------|--|
| <input type="checkbox"/> A 2.4 | <input type="checkbox"/> C 1.2 | <input type="checkbox"/> E Нет верного ответа. |
| <input type="checkbox"/> B 8   | <input type="checkbox"/> D 4.8 |  |

**Вопрос 6 ♣** Если оценивается модель  $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ , а истинной является модель  $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i$ , то МНК-оценка  $\hat{\beta}_2$  оказывается

- ☐ A всегда несмещённой      ☐ C эффективной      ☐ E равной нулю  
☐ B несмещённой при  $\widehat{\text{Cov}}(X, Z) = 0$       ☐ D несмещённой при  $\beta_1 = 0$       ☐ F всегда смещённой  
☐ G Нет верного ответа.

**Вопрос 7 ♣** Если в регрессии отсутствует свободный член, то в общем случае

- ☐ A сумма квадратов остатков равна нулю      ☐ D  $R^2$  является мерой качества подгонки регрессии  
☐ B  $\sum_i \hat{Y}_i = \sum_i Y_i$       ☐ E сумма остатков равна нулю  
☐ C  $TSS \neq ESS + RSS$       ☐ F Нет верного ответа.

**Вопрос 8 ♣** На третьем шаге теста МакАлера исследователь получил следующие результаты:

$$\widehat{\ln Y}_i = 1.54 - \underset{(0.20)}{1.2} X_i + \underset{(0.3)}{3.1} \hat{v}_{\exp(\widehat{\ln Y}), i}$$

$$\hat{Y}_i = \underset{(0.02)}{1.24} + \underset{(0.03)}{1.1} X_i + \underset{(0.2)}{2.4} \hat{v}_{\ln \hat{Y}, i}$$

На основании этих результатов исследователь

- ☐ A должен предпочесть полулогарифмическую модель  
☐ B должен отвергнуть гипотезу об адекватности исходной модели  
☐ C должен предпочесть линейную модель  
☐ D должен сделать вывод об отсутствии пропущенных переменных  
☐ E не может выбрать ни логарифмическую, ни линейную модель  
☐ F должен сделать вывод о наличии пропущенных переменных  
☐ G Нет верного ответа.

**Вопрос 9 ♣** Известны 95%-ые доверительные интервалы для коэффициентов регрессии:  $\beta_1 \in [-4; 10]$ ,  $\beta_2 \in [2; 10]$ . На уровне значимости 5%

- ☐ A  $\hat{\beta}_1$  значим,  $\hat{\beta}_2$  не значим      ☐ C  $\hat{\beta}_1$  не значим,  $\hat{\beta}_2$  значим      ☐ E значимость проверить невозможно  
☐ B  $\hat{\beta}_1$  значим,  $\hat{\beta}_2$  значим      ☐ D  $\hat{\beta}_1$  не значим,  $\hat{\beta}_2$  не значим      ☐ F Нет верного ответа.

Вопрос 10 ♣ Запись  $3.20E - 16$  означает

☐ A Ошибка с кодом 16

☐ D  $3.2 \cdot (e - 16)$

☐ G Нет верного ответа.

☐ B  $3.2^{e-16}$

☐ E  $3.2 \cdot e - 16$

☐ C  $3.2 \cdot 10^{-16}$

☐ F  $3.2 \cdot e^{-16}$

## Часть 2. Задачи.

1. На основании опроса была оценена следующая модель:

$$\ln(wage_i) = \beta_1 + \beta_2 exper_i + \beta_3 age_i + \beta_4 sex_i + \varepsilon_i$$

где:

- $wage_i$  — величина заработной платы в долларах
- $exper_i$  — опыт работы в годах
- $age_i$  — возраст в годах
- $sex_i$  — пол (1 — мужской, 0 — женский)

Показатель	Значение
$R^2$	0.903
Скорректированный $R^2$	<b>B7</b>
Стандартная ошибка регрессии	<b>B6</b>
Количество наблюдений	<b>B2</b>

Результаты дисперсионного анализа:

	df	SS	MS	F	P-значение
Регрессия	3	2638.3	879.4	<b>B5</b>	0.000
Остаток	<b>B1</b>	282.1	16.6		
Итого	<b>B3</b>	<b>B4</b>			

Коэффициент	Оценка	$se(\hat{\beta})$	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Константа	-6.96	12.38	-0.56	0.58	-33.08	19.16
$exper$	2.65	0.32	8.38	0.000	1.98	3.32
$age$	<b>B8</b>	<b>B9</b>	8.06	0.000	4.13	7.06
$sex$	-10.73	1.79	<b>B10</b>	0.000	-14.49	-6.95

а) Найдите пропущенные числа **B1–B10**.

б) Как изменятся результаты оценки регрессии, если переменную  $sex_i$  переопределить так, чтобы 0 соответствовал мужчинам, 1 — женщинам?

Ответ округляйте до 2-х знаков после запятой. Кратко поясняйте, например, формулой, как были получены результаты.

2. Юный эконометрист Вениамин очень любит опаздывать на первую пару. Он считает, что время (в минутах), на которое он опаздывает,  $time_t$ , зависит от количества снега (в миллиметрах), выпавшего за последние сутки,  $snow_t$ . После месяца упорного сбора данных, он смог оценить следующую модель:

$$\widehat{time}_t = 12 + 0.2snow_t$$

Оценка ковариационной матрицы коэффициентов,  $\widehat{\text{Var}}(\hat{\beta}) = \begin{pmatrix} 5 & 0.03 \\ 0.03 & 0.01 \end{pmatrix}$

Оценка дисперсии ошибок равна  $\hat{\sigma}^2 = 1.45$ .

За последние сутки выпало 15 миллиметров снега.

- а) Постройте точечный прогноз времени опоздания Вениамина
  - б) Постройте 95%-ый доверительный интервал для  $E(time_t | snow_t = 15)$  для ожидаемого опоздания Вениамина
  - в) Постройте 95%-ый предиктивный интервал (доверительный интервал) для фактического опоздания Вениамина
3. По данным 27 фирм, упорядоченных по капиталу,  $K_1 < K_2 < \dots < K_n$ , была оценена зависимость выпуска  $Y_i$  от труда  $L_i$  и капитала  $K_i$  с помощью моделей

$$(1) \ln Y_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + u_i$$

$$(2) \ln Y_i = \beta_1 + \beta_2 \ln \frac{L_i}{K_i} + u_i$$

	Модель (1)	Модель (2)
константа	1.1706 (0.326)	1.0692 (0.1317)
$\ln L$	0.6029 (0.125)	
$\ln K$	0.375 (0.085)	
$\ln \frac{L}{K}$		0.6369 (0.0754)
R-squared	0.943	0.74
F	200.24	71.351
P-value	0.0	0.0
RSS	0.851	
N	27	27

- а) На уровне значимости  $\alpha = 0.05$  проверьте для модели (1) гипотезы  $H_0: \beta_2 = \beta_3 = 0$  и  $H_0: \beta_3 = 0.5$ .
- б) Объясните, почему модель (2) является ограниченной версией модели (1). Явно выпишите ограничения. На уровне значимости  $\alpha = 0.05$  проверьте гипотезу об этих ограничениях.

- в) Фирмы разделили на маленькие,  $i \leq 14$ , и большие,  $i \geq 15$ . Для этих двух групп оценили отдельные регрессии. Можно ли считать, что производственные функции для больших и маленьких фирм не различаются? Ответ подтвердите подходящим тестом.

	Модель для $i \leq 14$	Модель для $i \geq 15$
константа	0.6998 (0.649)	1.4082 (0.678)
$\ln L$	0.9000 (0.133)	0.0081 (0.226)
$\ln K$	0.2100 (0.056)	0.805 (0.179)
R-squared	0.896	0.908
F	47.84	49.81
P-value	0.0	0.0
RSS	0.119	0.362
N	14	13

4. С помощью теста Бокса-Кокса оценили зависимость веса индивида (в килограммах) от его роста (в сантиметрах):

Log likelihood = -2659.5656

Number of obs	=	540
LR chi2(2)	=	230.68
Prob > chi2	=	0.000

	W	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
/lambda		1.055498	1.892654	0.56	0.577	-2.654035 4.76503
/theta		-.0263371	.1471576	-0.18	0.858	-.3147607 .2620865

Estimates of scale-variant parameters

	Coef.
Notrans	
_cons	2.936809
Trans	
H	.0237224
/sigma	.1660251

Test H0:	Restricted log likelihood	chi2	Prob > chi2
theta=lambda = -1	-2680.8693	42.61	0.000
theta=lambda = 0	-2659.7618	0.39	0.531
theta=lambda = 1	-2685.5201	51.91	0.000

Какую спецификацию модели (линейную, линейную в логарифмах, полулогарифмическую) следует предпочесть и почему?

5. По данным для 23 демократических стран оценили зависимость индекса Джини<sup>1</sup> от ВВП на душу населения с учетом ППС (паритета покупательной способности). Затем провели тест Рамсея.

```
. reg gini gdp if democ==1
```

Source	SS	df	MS	Number of obs = 23		
Model	506.853501	1	506.853501	F( 1, 21) = 13.05		
Residual	815.572523	21	38.8367868	Prob > F = 0.0016		
				R-squared = 0.3833		
				Adj R-squared = 0.3539		
Total	1322.42602	22	60.1102738	Root MSE = 6.2319		

  

gini	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
gdp	-.0006307	.0001746	-3.61	0.002	-.0009937	-.0002676
_cons	44.30983	3.572733	12.40	0.000	36.87993	51.73974

```
. ovtest
```

```
Ramsey RESET test using powers of the fitted values of gini
Ho: model has no omitted variables
F(3, 18) = 5.16
Prob > F = 0.0095
```

- Сформулируйте нулевую и альтернативную гипотезу теста Рамсея
- Опишите пошагово, как проводится тест Рамсея
- Прокомментируйте результаты теста Рамсея

<sup>1</sup>Индекс Джини — это мера неравенства доходов. Чем выше индекс Джини, тем выше неравенство.



Немного хинтов:

1.

```
2. y_hat <- 12 + 0.2 * 15  
se_y_hat <- 5 + 225 * 0.01 + 2 * 15 * 0.03  
se_forecast_error <- se_y_hat + 1.45  
t_crit <- qt(0.975, df = 30 - 2)
```

```
3. t_obs <- (0.385 - 0.5) / 0.085  
t_crit <- qt(0.975, df = 27 - 3)  
F_obs <- (0.943 - 0.74) / (1 - 0.943) * (27 - 3)  
F_crit <- qf(0.95, df1 = 1, df2 = 27 - 3)
```