

# Метрика

Задача 1.

Предполагая нормальность остатков, оцените модель  $y_i = \beta x_i + \varepsilon_i$  по трём наблюдениям:  $x_1 = 0, x_2 = 0, x_3 = 1, y_1 = 0, y_2 = 1, y_3 = 4$ . Найдите  $\hat{\beta}, RSS, TSS, ESS, R^2$ .

Задача 2.

По историческим данным начала 20-го века была оценена зависимость длины тормозного пути от скорости машины. По 50 наблюдениям была оценена регрессия:

$$\widehat{dist}_i = -17.57 + 3.9speed_i$$

Стандартные ошибки двух оцененных коэффициентов равны 6.7 и 0.41 соответственно. Ковариация оценённых коэффициентов равна  $-2.65$  и  $\hat{\sigma} = 15.38$ .

1. Проверьте гипотезы о значимости каждого коэффициента на 10%-ом уровне значимости.
2. Постройте 90%-ый предиктивный интервал для длины тормозного пути машины ехавшей со скоростью 50 км/ч.

Подсказка: Нормальное  $N(0; 1)$  распределение с вероятностью 90% лежит в пределах от -1.65 до 1.65.

Задача 3.

Винни-Пух построил логит-модель для предсказания качества мёда:  $y_i = 0$  означает неправильный мёд, а  $y_i = 1$  — правильный. Правильность мёда зависит от разных характеристик дупла, в частности от правильности пчёл ( $x_i$ ) и их количества ( $z_i$ ). Оценка модели выглядит так:

$$P[y_i = 1] = \Lambda(-0.2 + 2x_i + 0.02z_i)$$

Здесь  $\Lambda()$  — логистическая функция,  $\Lambda(t) = e^t / (1 + e^t)$ .

1. Спрогнозируйте вероятность правильного мёда для дупла с сотней правильных пчёл.
2. Во сколько отличается вероятность правильного мёда для дупла с правильными и аналогичного дупла с неправильными пчёлами?
3. Стандартные ошибки коэффициентов равны 0.1, 0.2 и 0.3 соответственно. Проверьте значимость каждого коэффициента и дайте Винни-Пуху содержательные рекомендации касательно правильности мёда.

	Model 1
(Intercept)	8.45*** (0.00)
log(carat)	1.68*** (0.00)
R <sup>2</sup>	0.93
Adj. R <sup>2</sup>	0.93
Num. obs.	53940
RMSE	0.26
*** $p < 0.001$ , ** $p < 0.01$ , * $p < 0.05$	

Таблица 1: Statistical models

Задача 4.

Исследовательница Мишель оценила зависимость логарифма цены бриллиантов от логарифма массы. Результаты приведены в таблице 1.

```
model <- lm(data = diamonds, log(price) ~ log(carat))
texreg(model)
```

А затем провела тест Уайта (таблица 2):

```
bptest(model,
  varformula = ~log(carat) + I(log(carat) ^ 2),
  data = diamonds) %>% pander
```

Таблица 2: studentized Breusch-Pagan test: model

Test statistic	df	P value
560.4	2	2.002e-122 * * *

1. Аккуратно сформулируйте  $H_0$  и  $H_a$  данного теста
2. К каким выводам пришла Мишель по результатам теста
3. Что результаты теста говорят о проверке гипотез о значимости коэффициентов?

	Pooling	FE	RE
(Intercept)	10.36*** (1.20)		9.07*** (1.12)
wage	-0.32*** (0.05)	-0.14*** (0.03)	-0.16*** (0.03)
capital	2.10*** (0.04)	0.80*** (0.06)	1.13*** (0.06)
R <sup>2</sup>	0.69	0.16	0.27
Adj. R <sup>2</sup>	0.69	0.03	0.27
Num. obs.	1031	1031	1031

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

Таблица 3: Statistical models

Задача 5. Исследователь Василиса хочет сравнить три модели (FE/RE/Pooled регрессия) в исследовании безработицы. Результаты тестов и оценок моделей приводятся в таблицах 3-6:

```
data("EmplUK", package = "plm")
employment <- pdata.frame(EmplUK, index = c("firm", "year"))
pooling <- plm(data = employment, emp ~ wage + capital, model = "pooling")
fe <- plm(data = employment, emp ~ wage + capital, model = "within")
re <- plm(data = employment, emp ~ wage + capital, model = "random")
```

```
texreg(list(pooling, fe, re), custom.model.names = c("Pooling", "FE", "RE"))
```

```
phptest(fe, re) %>% pander
```

Таблица 4: Hausman Test:  $\text{emp} \sim \text{wage} + \text{capital}$

Test statistic	df	P value	Alternative hypothesis
181.7	2	3.538e-40 * * *	one model is inconsistent

```
plmtest(re, type = "bp") %>% pander
```

Таблица 5: Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels:  $\text{emp} \sim \text{wage} + \text{capital}$

Test statistic	df	P value	Alternative hypothesis
3335	1	0 * * *	significant effects

```
pFtest(fe, pooling) %>% pander
```

Таблица 6: F test for individual effects:  $\text{emp} \sim \text{wage} + \text{capital}$

Test statistic	df1	df2	P value	Alternative hypothesis
115.6	139	889	0 * * *	significant effects

1. Выберите оптимальную модель объяснив свой выбор
2. Приведите пример ситуации, в которой FE модель невозможно оценить