

Ф.И.О.: \_\_\_\_\_

1. (a)  (b)  (c)  (d)

2. (a)  (b)  (c)  (d)

3.  .

4.  .

5. (a)  (b)  (c)  (d)

6. (a)  (b)  (c)  (d)

7. (a)  (b)  (c)  (d)

8. (a)  (b)  (c)  (d)

9. (a)  (b)  (c)  (d)

10.  .

1. Что такое доверительная вероятность (p-value)? Отметьте справедливые утверждения.

- (a) Низкое значение  $p$  показывает, что значение тестовой статистики экстремально при условии справедливости  $H_0$
- (b)  $(1 - p)$  — это вероятность того, что альтернативная гипотеза верна
- (c)  $p$  оценивает вероятность получить такое значение тестовой статистики, если  $H_0$  верна
- (d) При помощи  $p$  нельзя оценить вероятность того, что  $H_0$  верна

2. Сравните  $G^2$ -тест и тест отношения правдоподобий ( $LRT$ ):

- (a) значения обеих статистик подчиняются  $\chi^2$ -распределению
- (b) значимость обеих статистик можно проверить при помощи  $F$ -критерия
- (c)  $G^2$  - это частный случай  $LRT$  теста
- (d)  $G^2$ -тест сравнивает правдоподобие нулевой модели с правдоподобием предложенной, а  $LRT$ -тест сравнивает правдоподобие полной и уменьшенной моделей

3. Перед вами модель:

```
glm(formula = Y ~ X + K + 1, data = table)
```

Применение функции 'logLik(model1)' дает следующие результаты.

```
> logLik(model1)[1]
```

```
[1] -476.3549
```

Вычислите значение AIC для данной модели, округлите до тысячных

4. Перед вами результаты логистической регрессии на биномиальных данных. Имеются два предиктора PR1 и PR2.

```
> summary(model1)$call
```

```
glm(formula = RES ~ PR1 + PR2, family = "binomial", data = table)
```

```
> summary(model1)$coefficients
```

	Estimate	Std. Error	z value
(Intercept)	20.6811391	4176.289700	0.004952037
PR1high	-0.2006707	1.483921	-0.135230082
PR2slow	-22.8783636	4176.289672	-0.005478155

	Pr(> z )
(Intercept)	0.9960489
PR1high	0.8924300
PR2slow	0.9956291

```
> summary(model1)$null.deviance
```

```
[1] 55.45177
```

```
> summary(model1)$deviance
```

```
[1] 13.38572
```

Рассчитайте AIC для модели.

5. Отметьте верные утверждения, которые характеризуют переобученную модель (overfitted model)
- (a) Модель описывает не только отношения между переменными, но и случайный шум
  - (b) Модель имеет высокую предсказательную силу при использовании с новыми данными
  - (c) Модель хорошо работает на обучающей выборке, но плохо предсказывает на новых данных
  - (d) В модель включено очень много предикторов
6. Перед вами графики, построенные по четырем моделям, описывающим одни и те же данные. Определите, на каком из графиков (рис. 1) приведена самая переобученная модель?

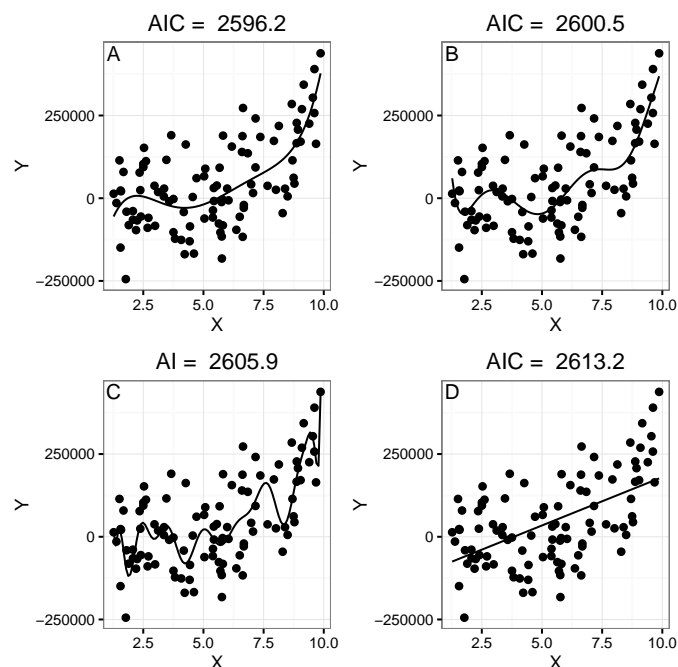


Рис. 1: Найдите на графиках самую переобученную модель.

- (a) A
  - (b) B
  - (c) C
  - (d) D
7. Найдите высказывания, которые характеризуют связывающую функцию (Link function):
- (a) это то же самое что Logit link
  - (b) это функция правдоподобия для предложенной модели
  - (c) присутствует в общих линейных моделях (линейная регрессия, дисперсионный анализ) в виде  $f(z) = z$  (Identity link) и не влияет на расчеты
  - (d) может принимать вид  $f(z) = \log(z)$  для счетных данных

8. Какое из предложенных распределений характеризуют следующие признаки:
- \* Применимо только для целых чисел,
  - \* Пределы варьирования от  $-\text{Inf}$  до  $+\text{Inf}$ ,
  - \* Дискретное,
  - \* Среднее равно дисперсии, с увеличением среднего растет размах варьирования,
  - \* Имеется параметр избыточности дисперсии
- (a) Гамма распределение  
(b) Ни одно из перечисленных  
(c) Пуассоновское распределение  
(d) Нормальное распределение
9. Выберите верные утверждения, относящиеся к избыточной дисперсии:
- (a) больше дисперсии в реальной выборке, чем предсказывает модель  
(b) наличие ненулевой избыточной дисперсии является условием применимости отрицательной биномиальной модели  
(c) отсутствует при моделировании с помощью нормального распределения (хотя модель может быть плохой по другим признакам)  
(d) избыточность дисперсии может быть обусловлена природной гетерогенностью данных, не отраженной в модели
10. Загрузите встроенный датасет `warpbreaks`. В нем приведены подсчеты разрывов шерстяной нити в процессе тканья. Постройте пуассоновскую модель с числом разрывов (`breaks`) в виде зависимой переменной, натяжением нити (`tension`) и типом шерсти (`wool`) в качестве взаимодействующих предикторов. Рассчитайте значение избыточности дисперсии. Результат округлите до сотых.