ЛЕКЦИЯ 12. Сравнение моделей: AIC, BIC и как выбрать лучшую

Введение

К этому моменту вы построили уже не одну регрессионную модель.

Вы научились:

- выбирать переменные,
- применять регуляризацию (Ridge, Lasso),
- следить за значимостью и простотой модели,
- анализировать R^2, остатки, коэффициенты.

Но когда вы создали **несколько моделей**, возникает вопрос:

“А какая из них **лучшая**?”

- Та, у которой выше R^2?
- Та, у которой меньше переменных?
- Та, у которой коэффициенты “логичнее”?

Выбор не всегда очевиден.

Иногда одна модель объясняет больше, но сложнее.

Другая — проще, но чуть менее точна.

Чтобы сравнивать модели **объективно**, используются специальные метрики:

AIC и **BIC**.

Что такое AIC и BIC?

Это **критерии качества модели**, которые:

- учитывают **точность модели**,
- **штрафуют** за лишние переменные,

- помогают сравнивать модели **с разным числом факторов**.

AIC (Akaike Information Criterion)

- Чем **ниже AIC**, тем **лучше модель**
- Баланс между точностью и простотой
- Наказывает за “перегруженность”

BIC (Bayesian Information Criterion)

- Похож на AIC, но **штрафует сильнее**
- Больше подходит, когда у вас **много наблюдений**
- Тоже: **меньше = лучше**

Принцип работы

Формулы (упрощённо, без логарифмов):

```
AIC = -2 * логарифм правдоподобия + 2 * k
BIC = -2 * логарифм правдоподобия + k * log(n)
```

Где:

k — количество параметров модели (включая свободный член),

n — количество наблюдений

> Главное — **не считать вручную**, а сравнивать результаты, которые дают системы анализа (например, Python или R).

Как сравнивать

Модель	AIC	BIC	Вывод
Модель 1	128.3	134.2	Точнее, но сложнее
Модель 2	130.1	131.5	Проще, менее точна
✅ Победитель	128.3	134.2	AIC → модель 1

✍️ Пример из курса

Гипотеза: успеваемость зависит от сна, стресса, и мотивации.

Выстроили 2 модели:

- Модель A: сон + стресс
- Модель B: сон + стресс + мотивация

Модель	R^2	AIC	Вывод
A	0.68	145.2	Простая, но не самая точная
B	0.75	140.6	Чуть сложнее, но заметно лучше

✅ **Вывод:** модель B — предпочтительнее (ниже AIC)

Почему AIC лучше, чем просто R^2

Показатель	Что делает	Проблема
R^2	Считает, сколько объясняется	Всегда растёт при добавлении переменных
AIC	Балансирует точность и сложность	Может быть выше у “перегруженной” модели
BIC	Делает то же, но строже	Предпочитает простоту

Где получить AIC и BIC?

Среда	Поддержка
Python (sklearn, statsmodels)	✅ Да
R	✅ Да
Excel, Google Sheets	❌ Нет напрямую, но можно обсчитать вручную (сложно)

В рамках курса можно:

- сравнить R^2 и количество переменных,
- визуально обсудить: “Что проще?”, “Что переобучено?”,
- объяснить идею AIC и BIC как **логики выбора, а не только цифры**.

ИИ-поддержка

Инструмент	Что делает
ChatGPT	Объясняет разницу между AIC и BIC
Excel Copilot	Помогает визуализировать сравнение моделей
Notion AI	Формулирует вывод о том, почему выбрана конкретная модель

🚫 Запрещено:

- Опираться только на R^2 для выбора модели
- Игнорировать сложность модели (10 переменных ≠ лучше)
- Использовать AIC/BIC, не объяснив, что они значат
- Сравнивать модели с разным y (должна быть одна и та же цель)

Вывод

AIC и BIC — это не “магические числа”.

Это инструменты, которые учат **выбирать осознанно**:

не просто “что лучше объясняет”, а “что делает это экономно, ясно и честно”.

Это уже мышление зрелого аналитика:

“Не просто построить модель, а выбрать **лучшую** из возможных.”