

## ATE

$y_i$  - результат измерения

$x_i$  - проме переменные

$d_i = 1$  treatment group

$d_i = 0$  control group

$y_i(1)$  - значение  $y_i$ , если  $d_i = 1$

$y_i(0)$  - значение  $y_i$ , если  $d_i = 0$

Проблема:

$$d_i = 1 \Rightarrow y_i = y_i(1)$$

$$d_i = 0 \Rightarrow y_i = y_i(0)$$

Всем только одно

состоит

$$y_i = y_i(0) + d_i(y_i(1) - y_i(0))$$

$$y_i = d_i \cdot y_i(1) + (1 - d_i) \cdot y_i(0)$$

$$\frac{\text{treatment effect}}{\text{effect}} = \text{uplift} = U_i = y_i(1) - y_i(0)$$

Он же сбросывает, то он компенсирует

$$\text{ATE} = \mathbb{E}(y_i(1) - y_i(0))$$

average treatment effect

$$\text{CATE} = \mathbb{E}(y_i(1) - y_i(0) | x_i)$$

↑  
conditional

↑  
возраст

Если оба вида 2 независимые выборки, то тогда

оценка несущая:

$$\frac{1}{n} \sum_{i=1}^n [y_i(1) - y_i(0)] \quad \text{Несущая}$$

Мы можем посчитать:

$$\bar{y}^T - \bar{y}^C = \frac{1}{n} \sum_{i=1}^n [y_i(1) \cdot d_i - y_i(0) \cdot (1-d_i)]$$

$$\begin{aligned} & \mathbb{E}(y_i(1) | d_i=1) - \mathbb{E}(y_i(0) | d_i=0) = \mathbb{E}(y_i(1) | d_i=1) - \mathbb{E}(y_i(0) | d_i=0) = \\ & = \mathbb{E}(y_i(1) | d_i=1) - \mathbb{E}(y_i(0) | d_i=1) + \mathbb{E}(y_i(0) | d_i=1) - \mathbb{E}(y_i(0) | d_i=0) = \\ & = \mathbb{E}(y_i(1) - y_i(0) | d_i=1) + \mathbb{E}(y_i(0) | d_i=1) - \mathbb{E}(y_i(0) | d_i=0) \end{aligned}$$

ATE

average treatment effect on the treated

ср. эфф. бол. на тех,

кто уже получал

	$d_i = 0$	$d_i = 1$
$y_i(1)$	red blob	green blob
$y_i(0)$	green blob	red blob

Selection bias SB

секвестрация

самоотбор

Лекарство

$y_i(1)$  - здоровье тех, кто

получил

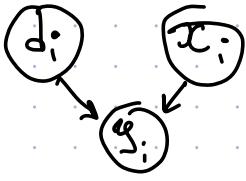
$y_i(0)$  - не получив

$SB < 0$  т.к. люди

выбирают лекарство

с меньшим  $y_i(0)$

Если бинарная переменная:



$$\mathbb{E}(y_i(0) | d_i=1) = \mathbb{E}(y_i(0))$$

$$\mathbb{E}(y_i(0) | d_i=0) = \mathbb{E}(y_i(0)) \Rightarrow SB = 0$$

$$\mathbb{E}(y_i | d_i=1) - \mathbb{E}(y_i | d_i=0) = ATET + SB =$$

$$= ATET + 0 = \mathbb{E}(y_i(1) - y_i(0) | d_i=1) = \mathbb{E}(y_i(1) - y_i(0)) =$$

$$= ATE$$

$$\hat{ATE} = \bar{y}^T - \bar{y}^C \xrightarrow{P} ATE$$

составленная  
обертка

Torga:

$$\hat{y}_i = \hat{w}_0 + \hat{w}_1 d_i$$

Наиболее

$$\hat{y}_i = \hat{w}_0 + \hat{w}_1 d_i + \hat{w}_2 x_i$$

перп.

$$\hat{y}_i = \hat{w}_0 + \hat{w}_1 d_i + \hat{w}_2 x_i$$

CV PED

$$\hat{z}_i = y_i - \hat{w}_2 x_i \quad \hat{\lambda}_i = \hat{z}_0 + \hat{z}_1 d_i$$

Как писать есть ли проблема в дизайне AB-теста?

- AA-тест

A	A	B
---	---	---

$$H_0: \mu_{A_1} = \mu_{A_2}$$

$$H_A: \mu_{A_1} \neq \mu_{A_2}$$

гипотеза  $\Rightarrow$   
OTB.

некоторый  
дизайн  
AB

Что может помешать тебе Ты?

→ Учебные тексты не соизглагают с подсказкой

# Coca Cola u TEHËK

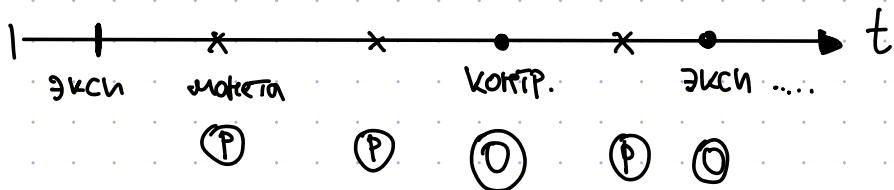
→ Сетевой эффект — пользователь влияет на  
Yandex.Taxi. Новоделные друзья новых.

Mama & BacL  
treatment control  
SB, capping

Pazitiv yeterli na Takcu  
 $\Rightarrow$  egzit no geneboi yene  
bere CT

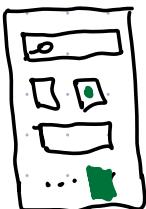
## Монти Бороться по-разному

switchback: pangasius melaleu pemulu  
7kcha



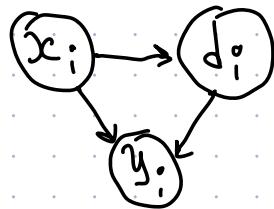
Наго же начопоться же проблемы со временем

→ On početné metody; NPS



Бабоги забыт от  
места опоса!

## Данные на основе



Важность - возраст влияет на решение  
самостоятельно

### Хорошая характеристика:

| Если  $y(0), y(1)$  не зависят от  $d$  |  $\exists$   
 $\Rightarrow$  все старое работает т.к.  $SB = 0$

Да, возраст влияет, то

$x=20$  лет в рамках него  $d$  не влияет на  $y(0), y(1)$

Чел не может заплатить в будущем, потому что  
у него только  $y(1)$  и выкинуться

Лекарства:

$$y(0) \downarrow \Rightarrow d_i = 1 \quad SB < 0$$

Что делать если заб. есть?

Эксперимента: этого техник

causal inference

Пример:  $\hat{ATE}$  - можно оценка, если  $SB \neq 0$

$$LATE = E(y_i(1) - y_i(0) | d_i(1) - d_i(0) > 0)$$

Эфф. средн. тех, кто решил участвовать....

но это не то....

$$\mathbb{E}(y_i(1) - y_i(0))$$

приятность ≠ корреляция

а как оценить зависимость от SB

## МЭТРИКИ

перфоманс, CV, AUC, ... — параметрические меры  
сделать АБ-тест

МЭТРИКИ — некараметрический мерой  
( $SB = 0$ , не реш. проблема)

$y_i$	$d_i$	$x_i$
6	1	21
7	0	40
8	1	70
10	1	45
1	0	15
3	0	79

$$ATE = \mathbb{E}(y_i(1) - y_i(0))$$

Числ:

считать разницу  $y_i(1) - y_i(0)$

у большинства но  $x$  кандидатов

$$(6-1) + (8-3) + (10-7) = \hat{ATE}_{\text{MATCH}}$$

KNN  $k=1$

в.з. как подсчитать  $k$

$$\hat{y}_i(0) = \begin{cases} y_i, & d_i = 0 \\ NN_k^1(x_i), & d_i = 1 \end{cases}$$

$$\hat{y}_i(1) = \begin{cases} y_i, & d_i \geq 2 \\ NN_k^0(x_i), & d_i = 0 \end{cases}$$

А как ноктю, что отмече укачио?

Ну метк D.U., регион S.e.

$0 \in [\hat{ATE}_L^{\text{MATCH}}, \hat{ATE}_R^{\text{MATCH}}] \Rightarrow$  легкачко

Теория  $\hat{ATE}$   $\Rightarrow$  бүтсірәп  
нет

Хорошая новость: Это работает, оценка соч.

Кобзевка

$\hat{ATE}^{\text{MATCH}} \xrightarrow{P} ATE$

Эта сходимость Тен не доказал, но  
согласие предикторов в KNN

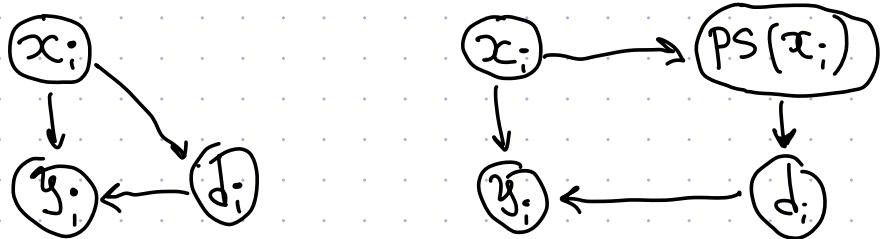
$$PS(x_i) = P(d_i = 1 | x_i)$$

propensity score с вероятностью обьекта залечеть в  
treatment, если испытать эксп-  
тиса самотек

Если эксп. рандомизированы, тогда

$$PS(x_i) = \frac{1}{2}$$

$\hat{PS}(x_i)$  — избр/бунт h.t.h.



$PS(x_i)$  – единственный фактор для  $d_i$

KNN только на  $PS(x_i)$

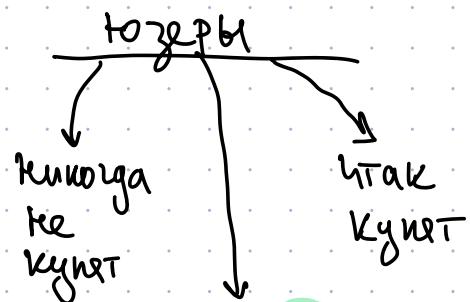
методика + PS

Есть теорема

| Быстро! Тут уже не работает  
| никто не знает откуда взять  $se \dots$

## UPLIFT - моделирование

		действие тозера	
		позвоким	не позвоким
возд.	позвоким	купил	не купил
	не позв.	купил	не купил



наго Такие  
майти и  
не звонить

купят, если  
только и не  
позвонить

$y_i(1)$

$y_i(0)$

$$T_i := y_i(1) - y_i(0) \quad \text{uplift}$$

Эффект воздействия

$\tilde{U}_i = 0$  звонок не побудил человека

$\tilde{U}_i = 1$  звонок заставил купить

$\tilde{U}_i = -1$  звонок заставил человека пересудить

Это все категориальные переменные

$$\text{CATE: } \tilde{U}(x_i) = \mathbb{E}(y_i(1)|x_i) - \mathbb{E}(y_i(0)|x_i)$$

$$d_i \in \{0, 1\}$$

$$y_i = d_i y_i(1) + (1-d_i) y_i(0)$$

$$PS(x_i) = \mathbb{P}(d_i=1|x_i)$$

## ① Two-model approach

Идея: Разделение выборки на две (одни козг./не одни)

$$X_0 = \{(x_i, y_i) | d_i = 0\} \quad q_0(x)$$

$$X_1 = \{(x_i, y_i) | d_i = 1\} \quad q_1(x)$$

$$\hat{U}(x_i) = \text{uplift}(x_i) = q_1(x_i) - q_0(x_i)$$

Преимущества:

- модель простая
- $a(x) = \hat{U}(x)$

Недостатки:

- модель уч. моделя  $y_i(0)$  и не уч.  $y_i(1)$ , то кое-что не засчитано
- разрывность

## ② Class Transformation approach

Идея:  $y_i$  — бинарная переменная

$$y_i^* = y_i \cdot d_i + (1 - y_i) \cdot (1 - d_i) = \begin{cases} y_i(1), & d_i = 1 \\ 1 - y_i(0), & d_i = 0 \end{cases}$$

Зависимость

$$\hat{T}(x_i) = 2 \cdot P(y_i^* = 1 | x_i) - 1 \quad \left[ \begin{array}{l} \text{Такая формула} \\ \text{получается, если} \\ \text{исходная выборка} \\ \text{сделана симметричной} \end{array} \right]$$

$$\hat{E}(y_i^* | x_i)$$

$$\forall x_i, PS(x_i) = 1/2$$

А что делать, если выборка несбалансирована?

$$y_i^* = y_i(1) \cdot \frac{d_i}{\hat{PS}(x_i)} - y_i(0) \cdot \frac{1 - d_i}{1 - \hat{PS}(x_i)}$$

$$PS(x_i) = P(d_i = 1 | x_i)$$

$\hat{PS}(x_i)$  — близкое значение

$$\begin{aligned} \hat{E}(y_i^* | x_i) &= P(d_i = 1 | x_i) \cdot \hat{E}(y_i^* | x_i, d_i = 1) + \\ &\quad + P(d_i = 0 | x_i) \cdot \hat{E}(y_i^* | x_i, d_i = 0) = \end{aligned}$$

$$= PS(x_i) \cdot \frac{\hat{E}(y_i(1) | x_i)}{PS(x_i)} + (1 - PS(x_i)) \cdot \frac{\hat{E}(y_i(0) | x_i)}{1 - PS(x_i)}$$

$$= \mathbb{E}(y_i(1) - y_i(0) | x_i) = \text{CATE} = T(x_i)$$

$$\mathbb{E}(y_i^*(x_i)) = T(x_i)$$



оценка этого мат. ож. — это полная оценка  
gpl uplift

MSE — мин

$$a(x) = \mathbb{E}(y|x)$$

$$a(x_i) \approx y_i^* \quad \text{MSE}$$

Есть другие преобразованные таргеты

В т.ч. gpl пересечки

Преимущества:

- модель простая, легче  
обучить
- $a(x) = V$

Недостатки:

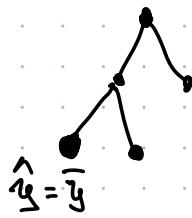
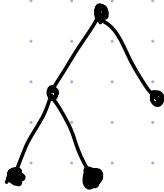
- $y_i^*$  — синтетич.
- корреляция с真实. ф.
- нельзя назвать uplift

### ③ Modeling uplift directly

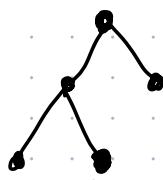
Модель: Предлагаемые подходы неоднозначны  
но-to-to левое и правое близкое из  $y^*$  uplift,

а мы хотим моделировать наивысшую

uplift-деревья



пересеч



$\hat{y} = \text{самый}$   
 $\text{част. класс}$

классиф.

$$\hat{T} = \frac{\sum g_i d_i}{\sum d_i} - \frac{\sum g_i (1-d_i)}{\sum (1-d_i)} = p - q$$

$$\bar{y}^T - \bar{y}^C \quad \begin{matrix} \text{противоположн.} \\ \text{множ} \end{matrix}$$

Критерий разбиения на вершины:

a)  $\Delta = |\hat{T}_{left} - \hat{T}_{right}|$  разница между метриками

b)  $\Delta = D_{\underset{\text{split}}{\text{after}}}(p^T, p^C) - D_{\underset{\text{split}}{\text{before}}}(p^T, p^C)$

$$KL(p||q) = \sum_k p_k \cdot \log \frac{p_k}{q_k}$$

$k \in \{\text{left}, \text{right}\}$

$$p_k, q_k - \text{уз } \hat{T}$$

$$E(p||q) = \sum_k (p_k - q_k)^2$$

$$\chi^2(p||q) = \sum_k \frac{(p_k - q_k)^2}{q_k}$$

Есть другие критерии

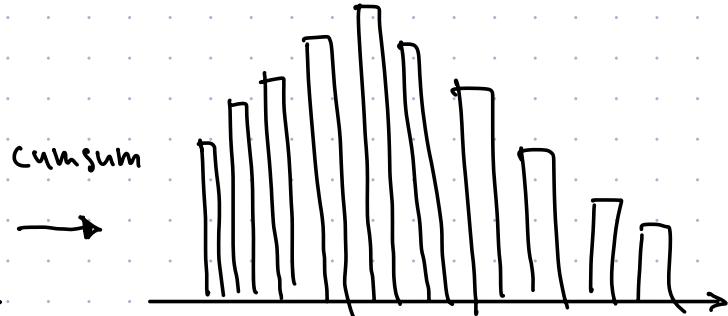
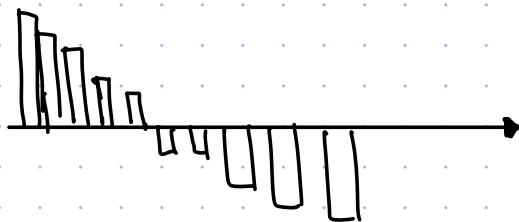
могуко построить?  
дерево .  $\Rightarrow$  можно построить  
лес / бустинг

### метрики качества

У нас нет истинных данных для обеих групп  
 $\Rightarrow$  кросс-валидация не работает !!

$\Rightarrow$  ~~Коэффициент~~  
оценка  
похожести

a) Графики

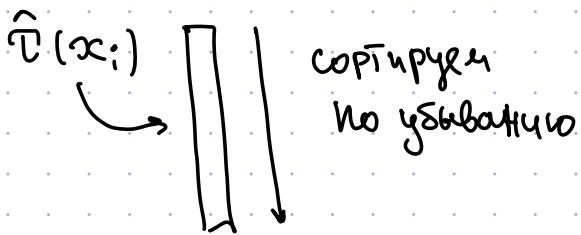


У хорошей модели

высокие значения в

1-й и 2-й квартилях

$\Rightarrow$  AUUC Area under uplift curve

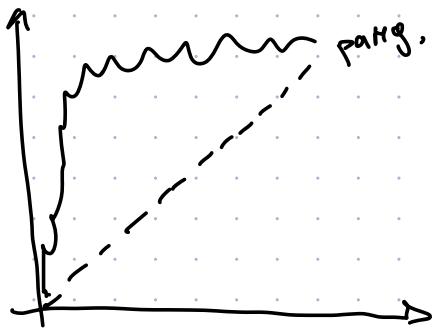
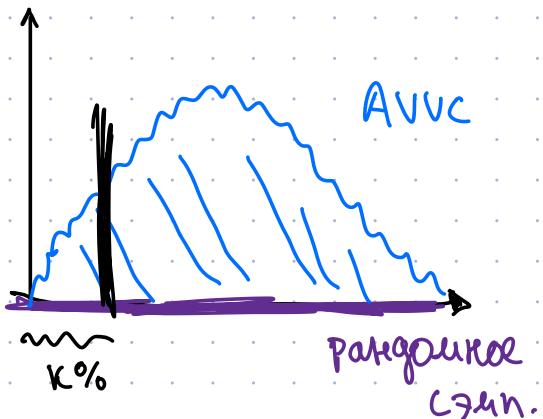


uplift  
curve

$$y_t^T = \sum_{i=1}^t y_i^T$$

$$f(t) = \left( \frac{y_t^T - y_t^c}{N_t^T} \right) \circ (N_t^T + N_t^c)$$

uplift      • kva-f0



VPLIFT@K - K%

Сколько & запасотал г.е. при работе  
на K% клиентов.

Qini - curve

$$f(t) = \frac{g(t) \cdot (N_t^T + N_t^C)}{N_t^T}$$

$$g(t) \approx y_t^T - \frac{y_t^C \cdot N_t^T}{N_t^C}$$

8) метрики оценивания на  $y^*$

$$\text{MSE}(y_i^*, \hat{\tau}_i) = \frac{1}{n} \sum_{i=1}^n (y_i^* - \hat{\tau}_i)^2$$

III. аппроксимация для  $\text{MSE}(\tau_i, \hat{\tau}_i)$