



Практическая задача

Charity Direct Mail – задача из области клиентской аналитики

Цель анализа

Организация поддержки ветеранов собирает регулярные пожертвования. Используя ответы, из предыдущей кампании, нужно построить модель чтобы предсказать будущие ответы.

Особенности данных

- Собраны данные предыдущих компаний
- Выборка сбалансирована по бинарному отклику
- **Реальный положительный отклик был 5%**
- **Стоимость контакта 0.68 USD**

Таблицы победителей

Rank	Total Profit	Overall Avg. Profit	Rank	Total Profit	Overall Avg. Profit	
1.	\$14,712	\$0.153	11.	\$ 10,720	\$ 0.111	
2.	14,662	0.152	12.	10,706	0.111	\$10,560
3.	13,954	0.145	13.	10,112	0.105	\$ 0.110
4.	13,825	0.143	14.	10,049	0.104	Total profit
5.	13,794	0.143	15.	9,741	0.101	Avg. profit
6.	13,598	0.141	16.	9,464	0.098	для
7.	13,040	0.135	17.	5,683	0.059	модели
8.	12,298	0.128	18.	5,484	0.057	«звать
9.	11,423	0.119	19.	1,925	0.020	всех»
10.	11,276	0.117	20.	1,706	0.018	

Описание данных

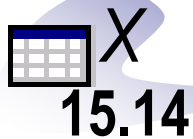
Name	Type	Label
TargetB	Numeric	Target Gift Flag
ID	Character	Control Number
TargetD	Currency	Target Gift Amount
GiftCnt36	Numeric	Gift Count 36 Months
GiftCntAll	Numeric	Gift Count All Months
GiftCntCard36	Numeric	Gift Count Card 36 Months
GiftCntCardAll	Numeric	Gift Count Card All Months
GiftAvgLast	Currency	Gift Amount Last
GiftAvg36	Currency	Gift Amount Average 36 Months
GiftAvgAll	Currency	Gift Amount Average All Months
GiftAvgCard36	Currency	Gift Amount Average Card 36 onths
GiftTimeLast	Numeric	Time Since Last Gift
GiftTimeFirst	Numeric	Time Since First Gift

Name	Type	Label
PromCnt12	Numeric	Promotion Count 12 Months
PromCnt36	Numeric	Promotion Count 36 Months
PromCntAll	Numeric	Promotion Count All Months
PromCntCard12	Numeric	Promotion Count Card 12 Months
PromCntCard36	Numeric	Promotion Count Card 36 Months
PromCntCardAll	Numeric	Promotion Count Card All Months
StatusCat96NK	Character	Status Category 96NK
StatusCatStarAll	Numeric	Status Category Star All Months
DemCluster	Character	Demographic Cluster
DemAge	Numeric	Age
DemGender	Character	Gender
DemHomeOwner	Character	Home Owner
DemMedHomeValue	Currency	Median Home Value Region
DemPctVeterans	Numeric	Percent Veterans Region
DemMedIncome	Currency	Median Income Region

Подходы к решению

Риск = Expected Profit =

$$P_TargetB * (P_TargetD - 0.68) - 0.68 * (1 - P_TargetB)$$


15.14

Прогнозируем один отклик,
второй фиксируем

$E(B|X)$
 $E(D|X)$

Отдельные модели и
расчет ожидаемого риска

$E(B, D|X)$

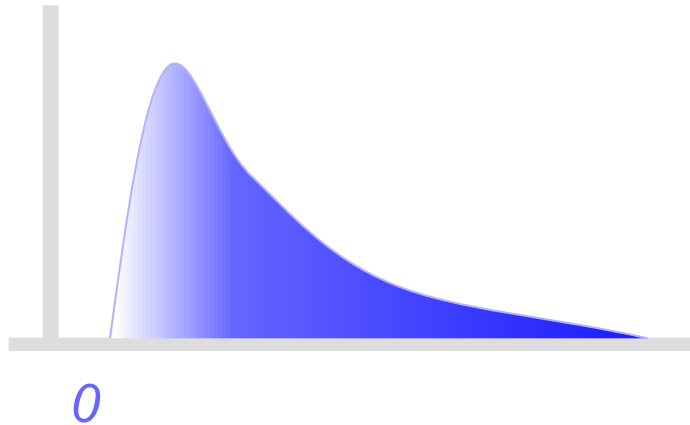
Совместное моделирование
откликов (например, нейросетью, но
сложно настраивать)

Фиксированный непрерывный

ОТКЛИК

solicit

ignore



*primary
outcome*

15.14

0

*secondary
outcome*

-0.68

0

Выбрать больший

Expected Profit Solicit = $15.14 \hat{p}_1 - 0.68 \hat{p}_0$

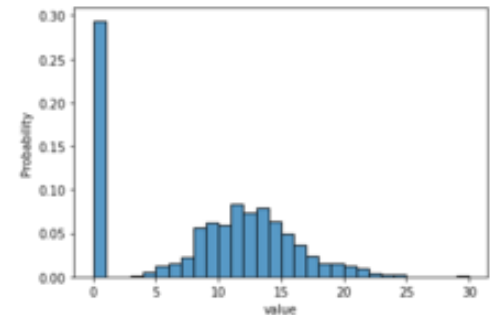
Expected Profit Ignore = 0

$p_1 \geq 0.68 / 15.82 \Rightarrow$ Solicit

$p_1 < 0.68 / 15.82 \Rightarrow$ Ignore

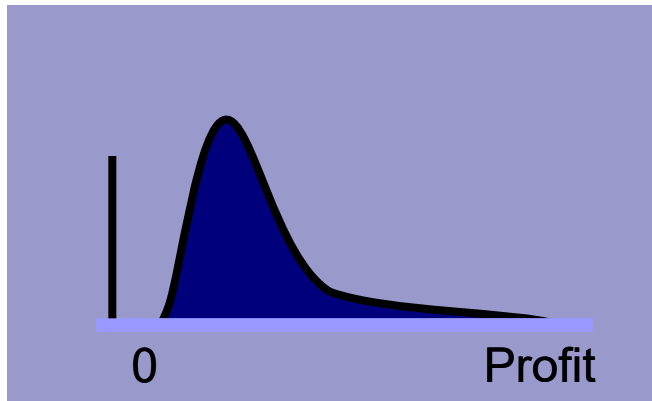
Прогнозируем только непрерывный отклик

- Считаем дискретный прогноз постоянным и равным априорной вероятности:
 - **Expected Profit = $P_TargetB * (P_TargetD - 0.68) - 0.68 * (1 - P_TargetB) \sim P_TargetD$**
- Zero-inflated модели - наличие пика распределения:
 - обычно в 0, но может быть и нет
 - пиков может быть несколько, например в 1 и 0
 - Много 0 еще не значит Zero – inflated – может быть просто маленькое мат. ожидание с чрезмерной дисперсией
 - должно быть две моды – в 0 и не в 0
 - Моделируются как порождающая модель смеси распределений: $p(x) = \sum_{j=1}^k w_j \varphi(x, \theta_j)$, $\sum_{j=1}^k w_j = 1$, $w_j \geq 0$

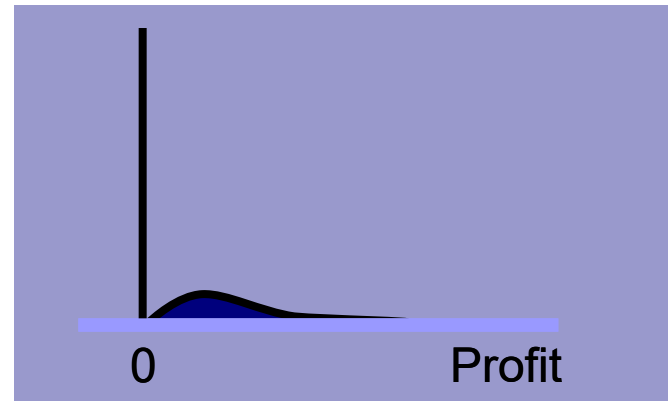


Прогнозирование риска

Primary Decision



Secondary Decision

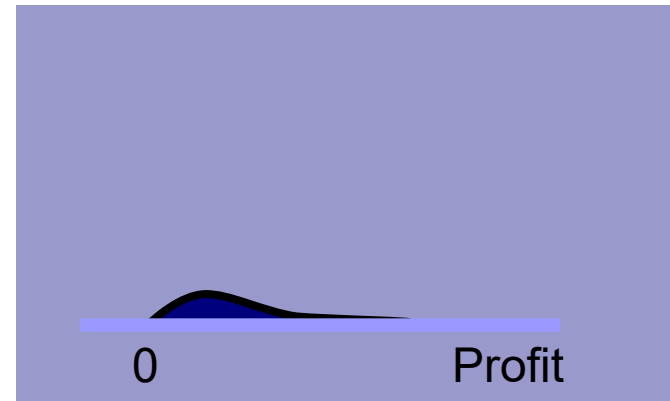
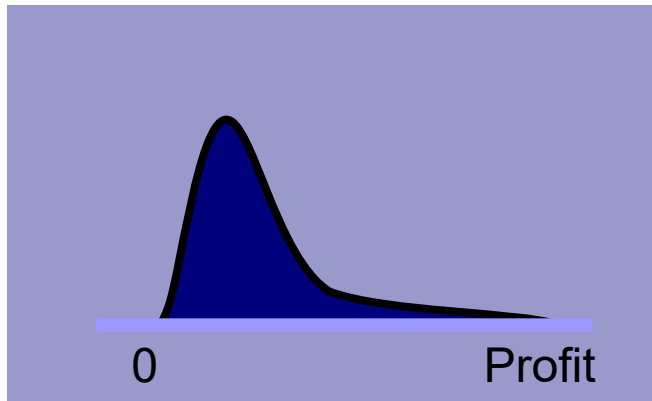


Оценка прогноза по исходам

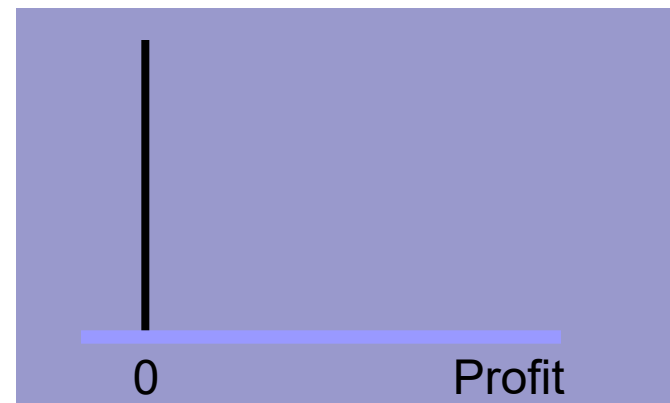
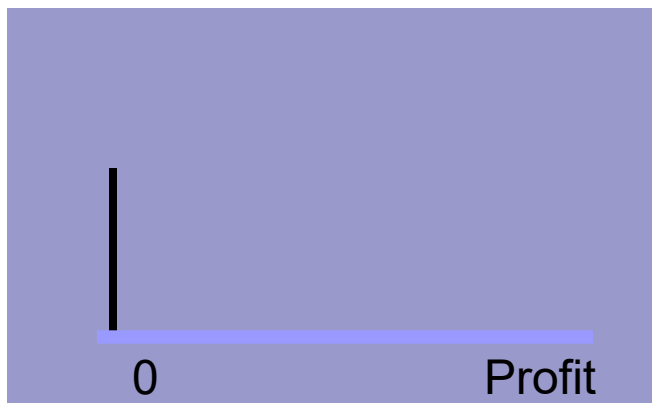
Primary Decision

Secondary Decision

Primary Outcome



Secondary Outcome

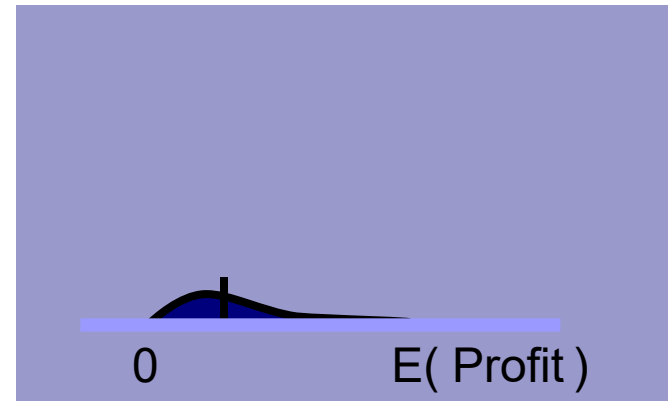
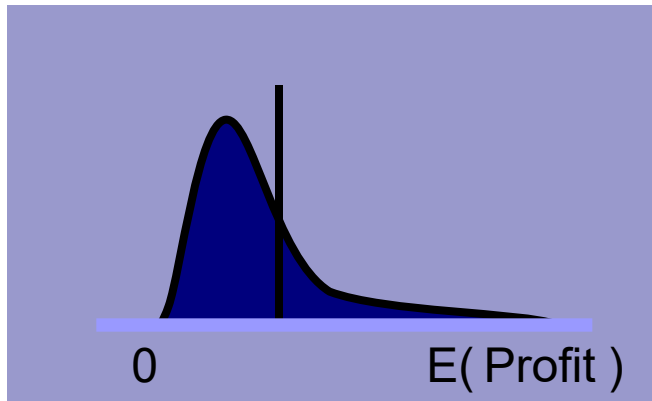


Ожидаемый выигрыш

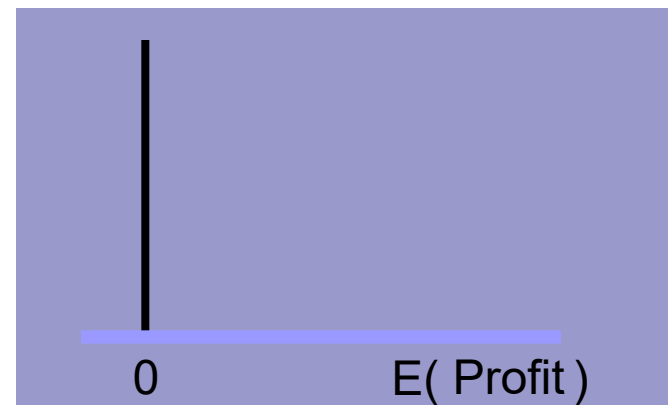
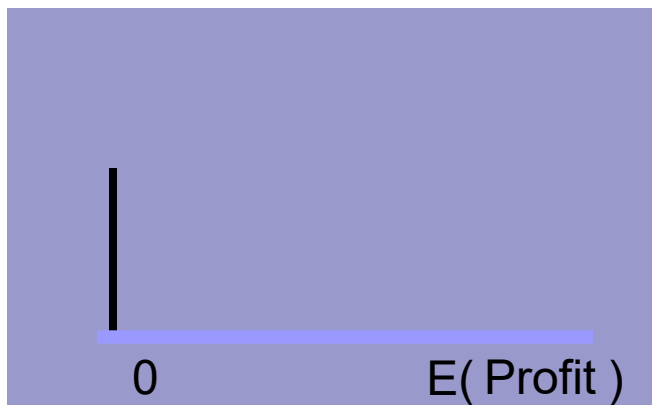
Primary Decision

Secondary Decision

Primary Outcome



Secondary Outcome

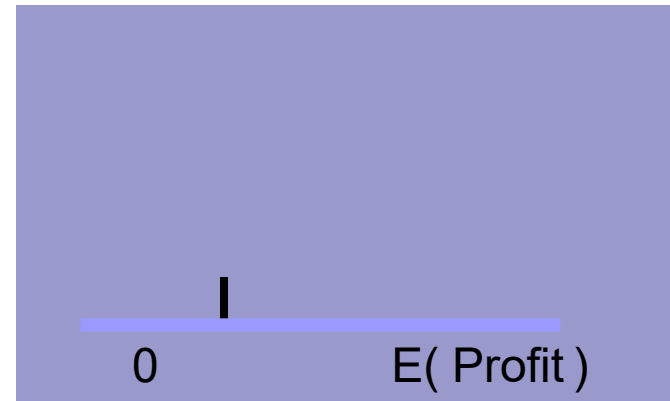
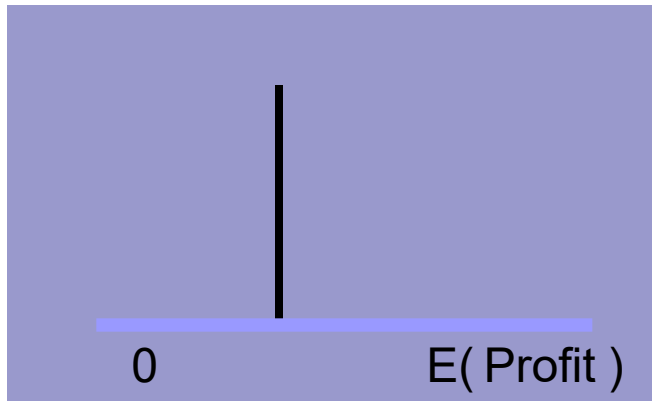


Упрощенная матрица выигрыша

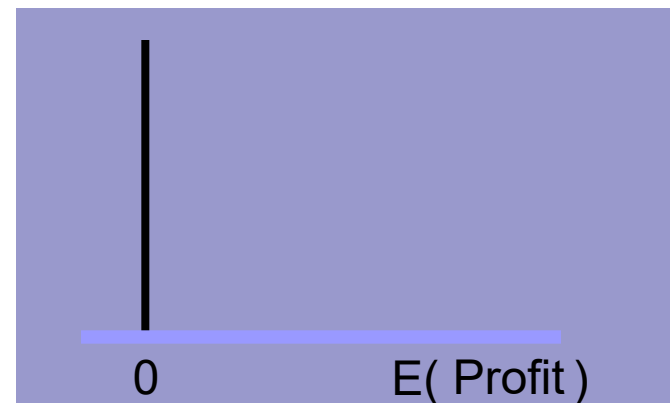
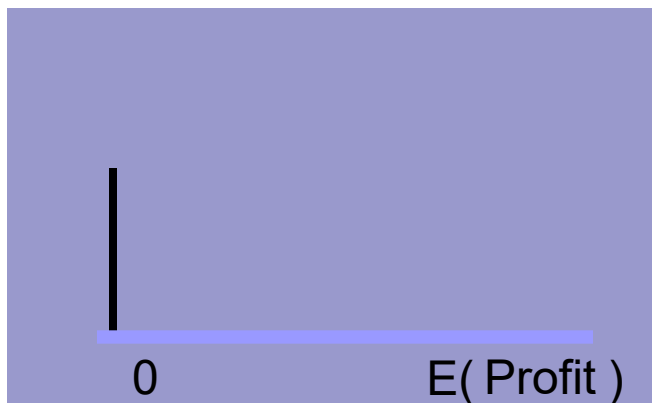
Primary Decision 

Secondary Decision 

Primary Outcome



Secondary Outcome

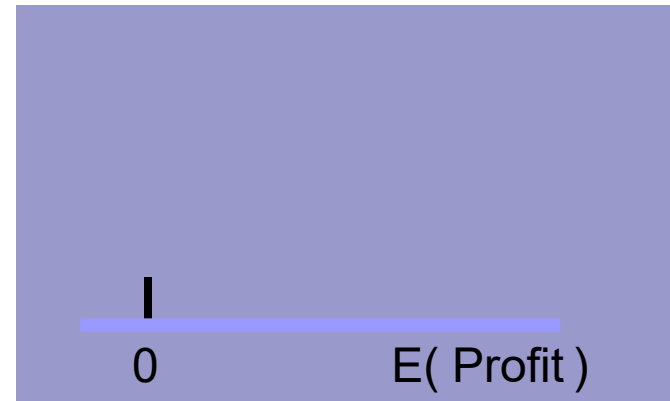
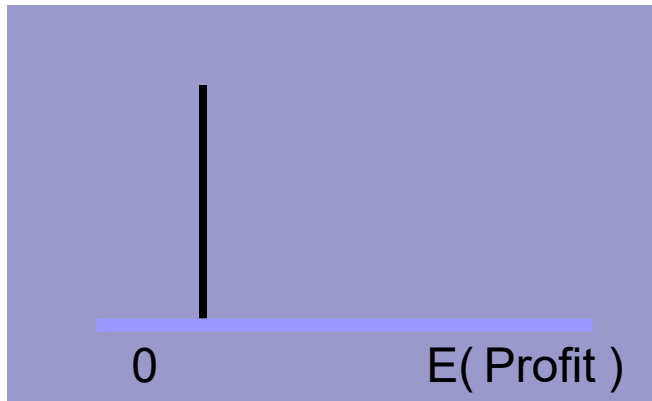


Упрощенная матрица выигрыша

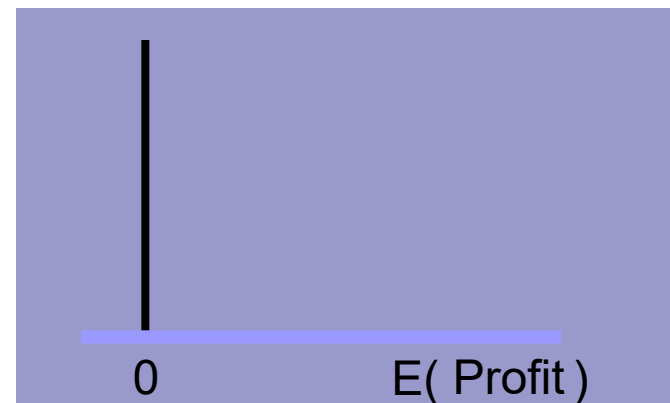
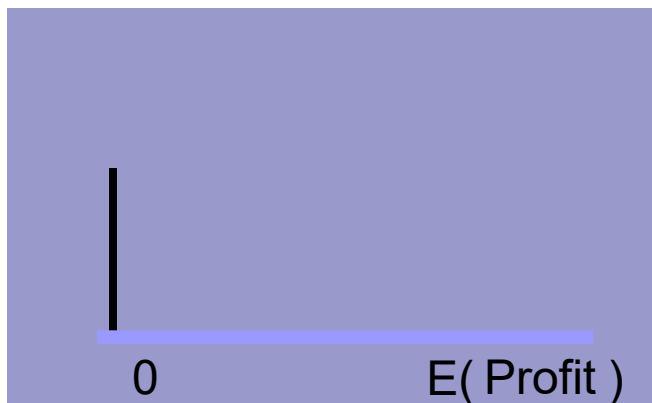
Primary Decision 

Secondary Decision 

Primary Outcome



Secondary Outcome

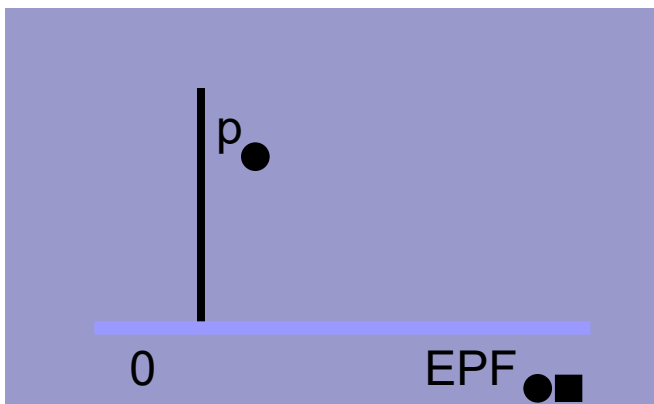


Оценка

Ожидаемый
Выигрыш

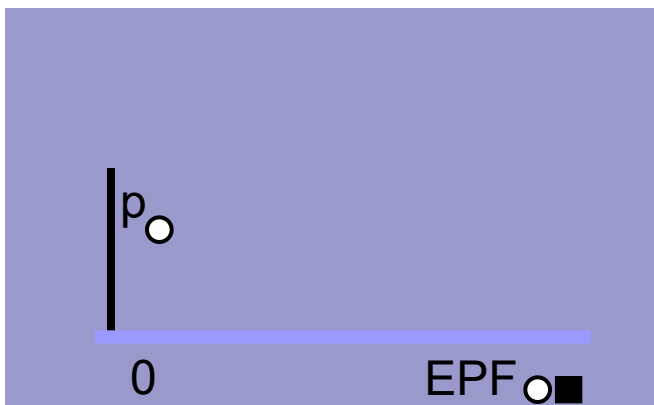
Primary Decision ■

Primary
Outcome



$$EPC_{\blacksquare} = ?$$

Secondary
Outcome

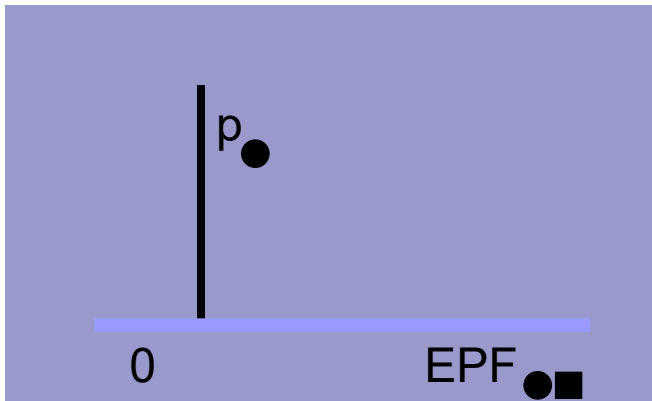


Оценка

Ожидаемый
Выигрыш

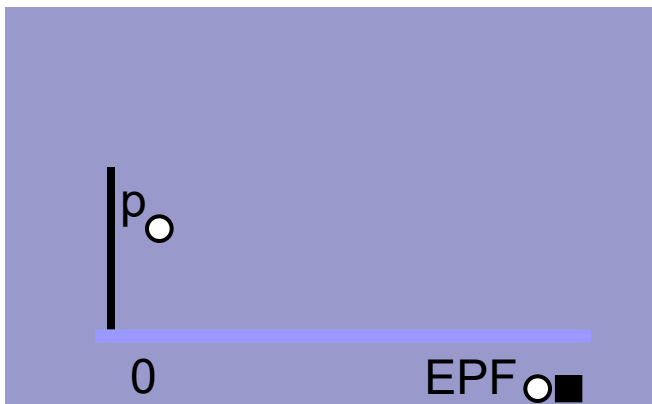
Primary Decision ■

Primary
Outcome



$$EPC_{\blacksquare} = EPF_{\bullet\blacksquare} \cdot p_{\bullet} + EPF_{\circ\blacksquare} \cdot p_{\circ}$$

Secondary
Outcome

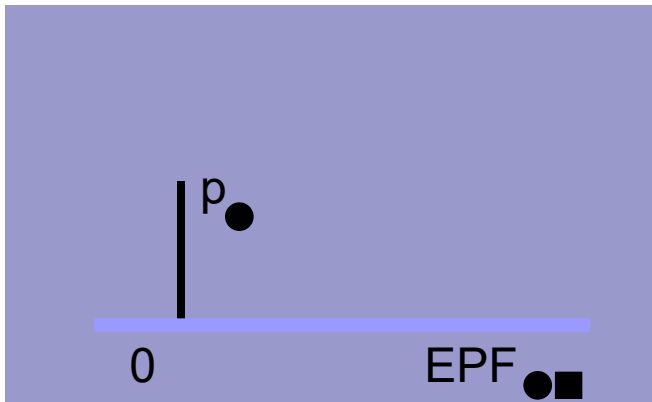


Оценка

Ожидаемый
Выигрыш

Primary Decision ■

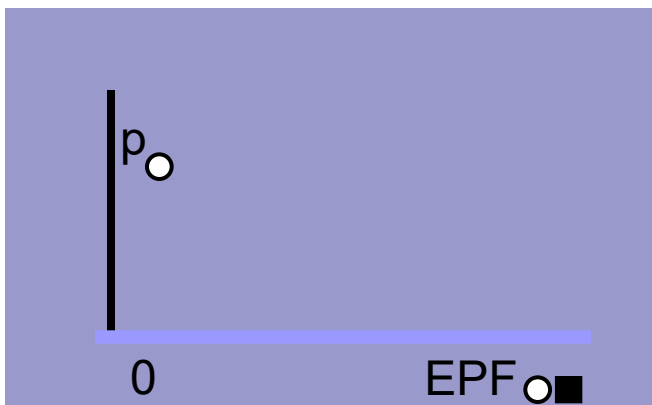
Primary
Outcome



$$EPC_{\blacksquare} = EPF_{\bullet\blacksquare} \cdot p_{\bullet} + EPF_{\circ\blacksquare} \cdot p_{\circ}$$

$$EPC_{\blacksquare} = EPF_{\bullet\blacksquare} \cdot p_{\bullet} + EPF_{\circ\blacksquare} \cdot p_{\circ}$$

Secondary
Outcome



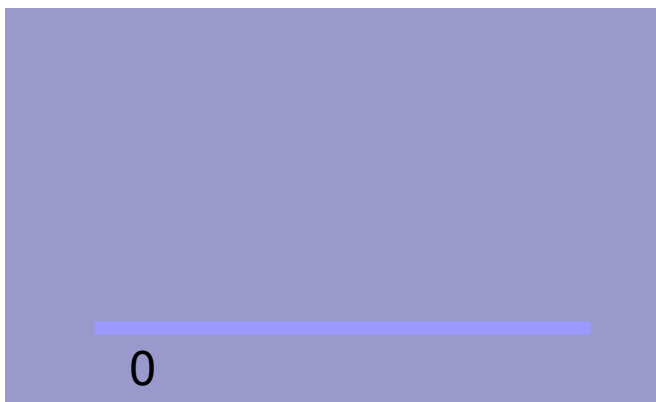
Оценка

Primary Decision



EPC

Primary Outcome



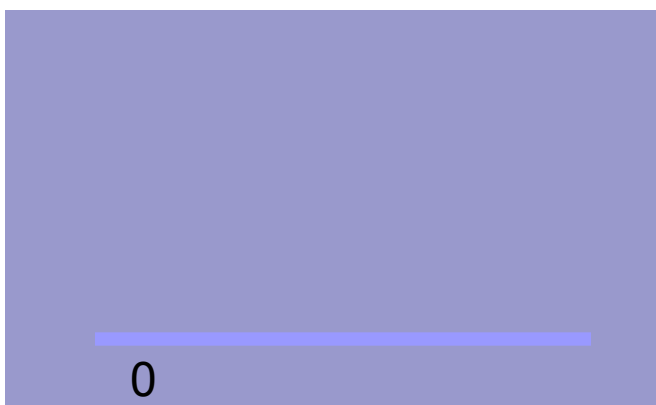
EPC 

EPC 

EPC 

EPC 

Secondary Outcome



EPC 

EPC 

EPC 

EPC 

Оценка

EPC

Сортировка по
EPC.

EPC 

EPC 

EPC 

EPC 

EPC 

EPC 

EPC 

EPC 

Накопленный ожидаемый выигрыш

Агрегация EPC
(по порогу).

EPC Σ EPC

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\blacksquare} \quad \sum_{\blacksquare \geq \blacksquare} EPC_{\square}$$

$$EPC_{\square} \quad \sum_{\blacksquare \geq \square} EPC_{\square}$$

Наблюдаемый (реальный) выигрыш

Реальный выигрыш
(observed profits).

EPC ΣEPC OP

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\blacksquare}

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\blacksquare}

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\square}

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\blacksquare}

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\square}

EPC_{\blacksquare} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\blacksquare}

EPC_{\square} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\blacksquare}

EPC_{\square} $\Sigma_{\blacksquare} EPC_{\square}$ OP_{\square}

Общий реальный выигрыш

Суммируем ОР для
Всех наблюдений с
EPCs выше порога.

EPC	Σ EPC	OP	Σ OP
EPC ₁ ■	$\Sigma_{\geq 1}$ EPC ₁ ■	OP ₁ ■	$\Sigma_{\geq 1}$ OP ₁ ■
EPC ₂ ■	$\Sigma_{\geq 2}$ EPC ₂ ■	OP ₂ ■	$\Sigma_{\geq 2}$ OP ₂ ■
EPC ₃ ■	$\Sigma_{\geq 3}$ EPC ₃ ■	OP ₃ ■	$\Sigma_{\geq 3}$ OP ₃ ■
EPC ₄ ■	$\Sigma_{\geq 4}$ EPC ₄ ■	OP ₄ ■	$\Sigma_{\geq 4}$ OP ₄ ■
EPC ₅ ■	$\Sigma_{\geq 5}$ EPC ₅ ■	OP ₅ ■	$\Sigma_{\geq 5}$ OP ₅ ■
EPC ₆ ■	$\Sigma_{\geq 6}$ EPC ₆ ■	OP ₆ ■	$\Sigma_{\geq 6}$ OP ₆ ■
EPC ₇ ■	$\Sigma_{\geq 7}$ EPC ₇ ■	OP ₇ ■	$\Sigma_{\geq 7}$ OP ₇ ■
EPC ₈ ■	$\Sigma_{\geq 8}$ EPC ₈ ■	OP ₈ ■	$\Sigma_{\geq 8}$ OP ₈ ■

Общий реальный выигрыш

Суммируем ОР для
Всех наблюдений с
EPCs выше порога.

EPC	Σ EPC	OP	Σ OP
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _■	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _■	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$
EPC _□	$\Sigma_{\geq \blacksquare} \text{EPC}_{\blacksquare}$	OP _□	$\Sigma_{\geq \blacksquare} \text{OP}_{\blacksquare}$

