

AIC ANALYTICS DAY '18



Об опыте внедрения А/Б
экспериментов в социальной
сети ОК

Социальная сеть
Одноклассники

2018

Dmitry Bugaychenko

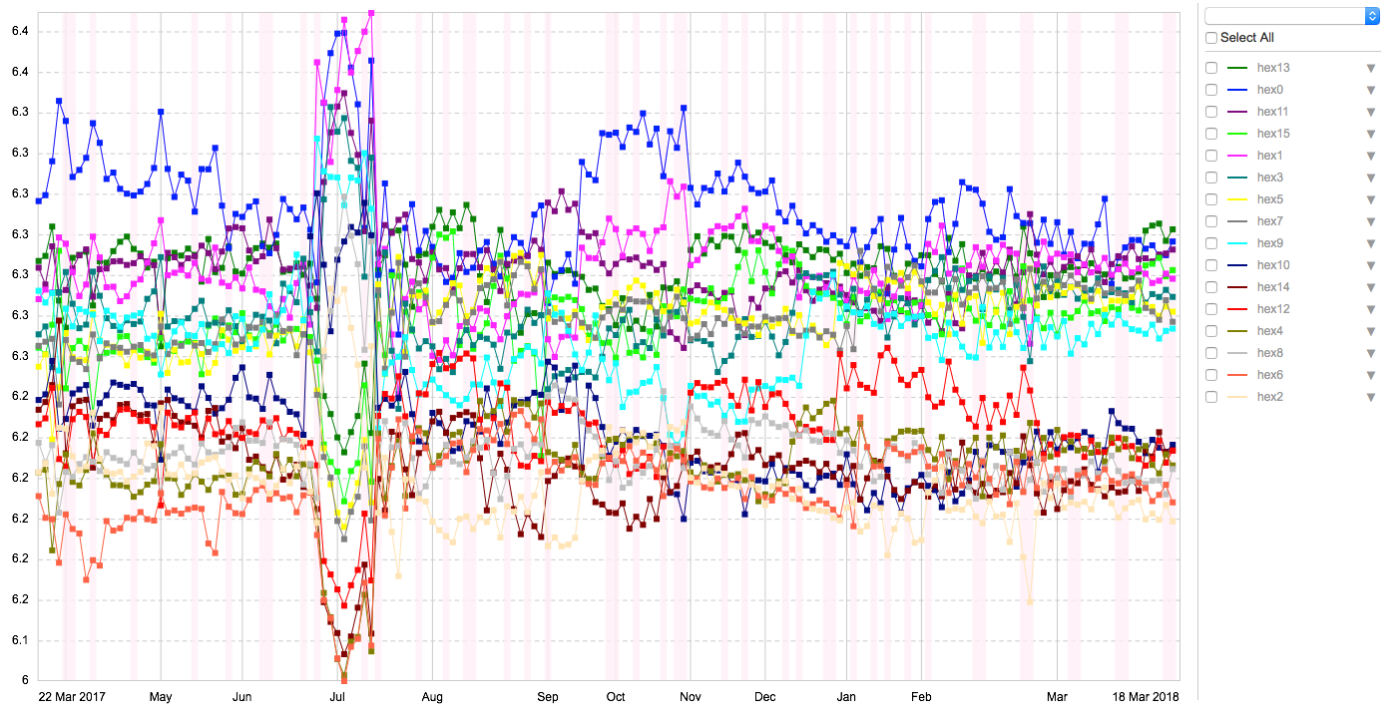




Хочешь сделать ОК лучше?

- Запили фичу
- Выкати под рубильником
- Включи в рамках эксперимента на 1/16
- Оцени эффект

Аудитория значительно отличается между 1/16



30+ экспериментов на вебе по фильтру «64-»



app.feed.discoveryInTabUsers	2,64-127
app.feed.likeFeedback.likeFeeds	64-79,240-255
app.friendships.config	toolbarIconEnabled=0-255 burningCounterEnabled=0-255 burni...
app.groups.active-topics	enabled=64-111,5 5...
app.groups.friendsGroupsPortletConfiguration	showGroupNameInTwoLines=0-255 showJoinedFriendNameBelow=64-...

Дальше больше: 37 БАЗОВЫХ показателей



Num Uniq Users, Относительные уникалы по платформам, Avg day minutes, Avg day minutes относительные по платформам, Число сессий с кликами, Num Sessions, Uniq Seen users, Uniq Clicked Users, Uniq Liked Users, Likes for users, Uniq Seen groups, Uniq Clicked Groups, Uniq Liked Groups, Likes for group, Avg day's max scroll length, Avg session scroll length, Num session with scroll > 100 feeds, Num session with scroll > 50 feeds, Отправлено сообщений, Походы в гости, Приглашения дружить, Вступление в группы, Классы, Классящие (не только лента), Среднее классов на классящего (не только лента), Комментарии, Отправлено подарков, Avg session minutes (by platform), Avg day's max scroll length (platform), Комментарии (юзерские статусы и фото), Комментарии WEB, Num Uniq likes in feed, DAU лайкающей аудитории, DAU комментирующей аудитории, DAU by shown feeds relative by platform, DAU by action type relative, Num shown feeds relative by platform



И что с того?

- Наблюдая эффект, мы не знаем
 - Можно ли считать его новым
 - Вызван ли он нашим экспериментом



И что с того?

- Наблюдая эффект мы не знаем
 - Можно ли считать его новым
 - Вызван ли он нашим экспериментом

Мы не можем оценить эффект
нашего эксперимента!



В итоге

- Более половины времени уходит на то, чтобы отследить что и куда двинулось
- Остальная половина - на то чтобы найти почему

ТАК ДАЛЬШЕ

ЖИТЬ НЕЛЬЗЯ!!!





План Б

- Узнать «что там у больших»
- Адаптировать под наши реалии
- Погонять прототип
- Внедрить



План Б

- Узнать «что там у больших»
- Адаптировать под наши реалии
- Погонять прототип
- Внедрить



Что там у больших?

- Хешированные с солью ИД
- Иерархия сертифицированных метрик
- Статистическая обработка результатов
- Высокий уровень автоматизации анализа
- Собственные системы менеджмента экспериментов
- Высокий уровень культуры эксперимента



Наши реалии

- Эксперименты по фиксированным партициям
- Графики в лучшем случае по 1/16-м
- Все надо отсматривать вручную

НЕЛЬЯ ПРОСТО ТАК ВЗЯТЬ



И ПЕРЕЙТИ НА СОЛЕННЫЕ ХЕШИ

risovuch.ru



Что нужно

- Расширение функционала рубильников
- Возможность оперативного контроля эффекта по графикам
- Автоматизация анализа эффекта



В итоге

- На одной ГРУППЕ может быть запущено несколько экспериментов
- Деление на ПОДГРУППЫ у каждого свое
 - Эксперименты не влияют на А/Б результаты друг друга
- Деление на ПОДГРУППЫ одноразовое
 - Эксперименты в прошлом не влияют на А/Б результаты новых экспериментов



Оперативный контроль эффекта

- Эффект на ГРУППЕ в целом видим на штатных графиках
- Базовые метрики на ПОДГРУППАХ считаются в системе потокового анализа
- Детальный эффект на ПОДГРУППАХ оценивается автоматически на следующий день



Автоматический анализ эффекта

- В зрелом проекте часто эффект $< 1\%$
- Делая вывод об эффекте «на глаз» будем часто ошибаться
- **Нужны статистические методы**



О допущениях

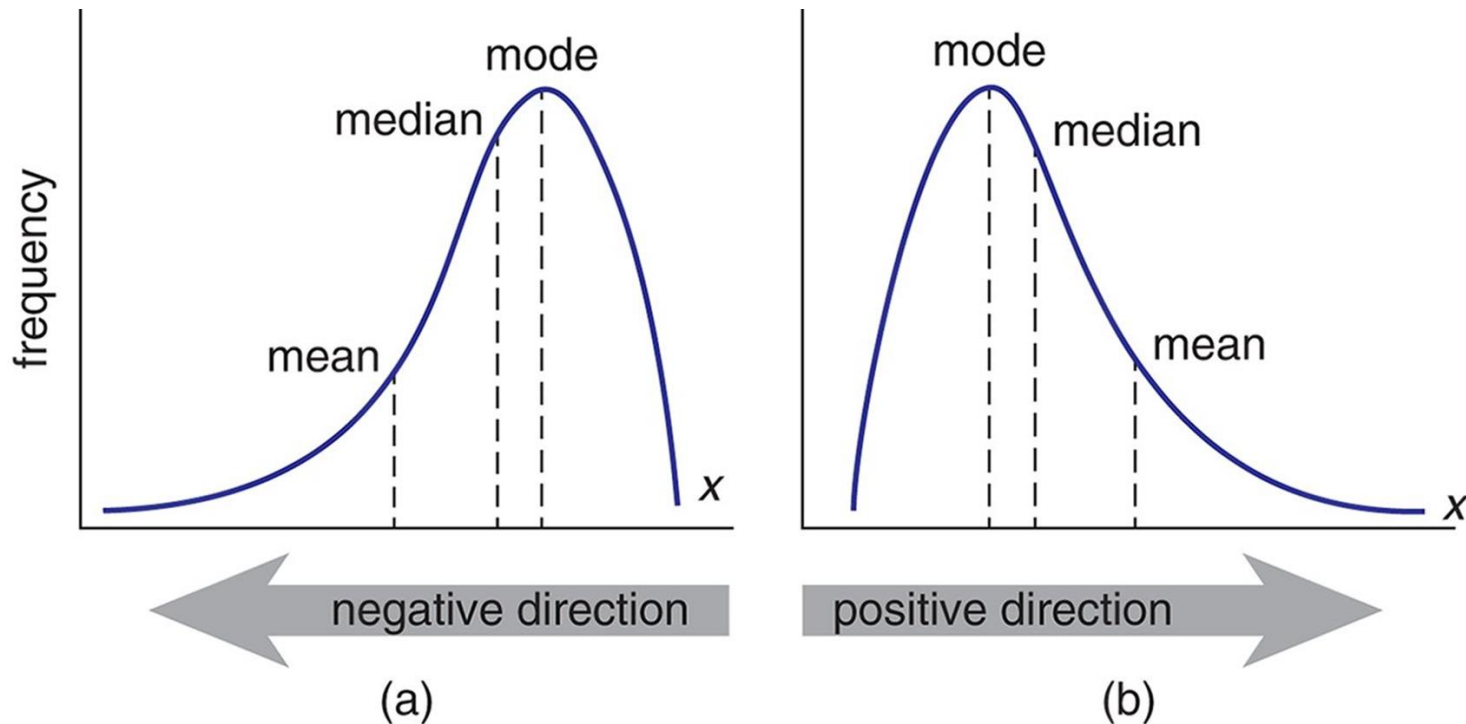
- Все методы содержат ряд допущений
- При нарушении допущений дают ошибочный результат
- Большинство методов требуют **независимости измерений**
- Многие требуют **нормальности распределения**



Пример нарушения независимости

- Поделим юзеров на $A1$, $A2$, $B1$, $B2$
- Проведем замеры на днях $D1$ и $D2$
- Сложим замеры $A1D1$, $A1D2$, ..., $B2D2$ в тест
- Получим ересь
 - $A1D1$ и $A1D2$ зависимы по $A1$
 - $B1D1$ и $B2D1$ зависимы по $D1$

Пример нарушения нормальности





Общая схема анализа

- Проверить наличие отличия тестом
- Оценить масштаб доверительным интервалом
- Убедится в новизне сравнив с прошлым

Шаг 1: Проверяем на отличие в рамках дня



- В рамках дня делим юзеров в ПОДгруппах на «корзины» случайным образом
- Считаем по каждой «корзине» сумму/уников/среднее
- Проверяем на нормальность (тест Колмогорова-Смирнова)
- Проводим тест на отличие (Т-тест или тест Манна-Уитни)
- Оцениваем доверительный интервал (среднего или медианы)
- В зависимости от знака эффекта выбираем (+ => нижнюю, - => верхнюю) как результат дня



Шаманские трюки

- Размер «корзины» определяем по формуле:
 $\text{MAX}(20, \text{SQRT}(N) / 5)$
- Если тест показал отличие, повторяем 10 раз и выбираем результат с наименьшей значимостью



Шаг 2: Агрегируем между днями

- Рассматриваем день как наблюдение 0/1 (нет/есть отличие)
- Проводим второй тест на не-случайность срабатывания (биномиальный тест)
- Делаем тест на отличие в периоде до и после включения (тест Манна-Уитни)
- Усредняем эффект по дням (статистически бессмысленно)

Запуск эксперимента глазами Юзера





Experiments / XPRM-10750

[feed] Сравнение ширины SVD

[Edit](#) [Comment](#) [Assign](#) [More ▾](#) [Configure A/B Test](#) [Add milestone](#) [Make ready](#) [Workflow ▾](#)




Details

Type:  Experiment
Priority:  Major
Labels: None
Project affected: Лента

Status: **IN PROGRESS**
Resolution: Unresolved

Description

Запуск эксперимента глазами Юзера

Summary*	<input type="text" value="[feed] Сравнение ширины SVD"/>
Platform*	<input type="text" value="All"/>
Partitions*	<input type="text" value=""/> <small>Range of partitions. For example, 64-127.</small>
Providers	<input type="text"/>
Linked Issues	<input type="text" value="is created for"/>
Issue	<input type="text" value="XPRM-10750"/>   

Begin typing to search for issues to link. If you leave it blank, no link will be made.

Запуск эксперимента глазами Юзера

Выбираем
группу

Summary*

[feed] Сравнение ширины SVD

All

Partitions*

32-63

Range of partitions. For example, 64-127.

Providers

Linked Issues

is created for

Issue

XPRM-10750 x

Begin typing to search for issues to link. If you leave it blank, no link will be made.

Запуск эксперимента глазами Юзера

Summary* [feed] Сравнение ширины SVD

Platform* All

Issues* 32-63

Range of partitions. For example, 64-127.

Providers

dw|

Linked Issues

Issue

dwhAdv

dwhAppMetric

dwhFeed

dwhFeedback



Выбираем
показатели

Запуск эксперимента глазами Юзера



Analytics / AN-2284

[feed] Сравнение ширины SVD

Edit

Comment

Assign



More ▾

Configure

Start work

Workflow ▾

Details

Type:	 Experiment	Status:	OPEN
Priority:	 Major	Resolution:	Unresolved
Component/s:	None		
Labels:	None 		
PMS Filter:	32-63,mmsalt:238369,filter:ID%2=1 Copy		

Запуск эксперимента глазами Юзера



Analytics / AN-2284

[feed] Сравнение ширины SVD

Edit

Comment

Assign




More ▾

Configure

Start work

Workflow ▾

Details

Type:	 Experiment	Status:	OPEN
Priority:	 Major	tion:	Unresolved
Component/s:	None		
Labels:	None 		
PMS Filter:	32-63,mmsalt:238369,filter:ID%2=1 Copy		

«Рубильник»
для включения

Запуск эксперимента глазами Юзера



Analytics / AN-2284

[feed] Сравнение ширины SVD

[Edit](#) [Comment](#) [Assign](#) [More ▾](#) [Configure](#) [Start work](#) [Workflow ▾](#)

Details

Type:  Experiment

Priority:  Major

Component/s: None

Labels: None 

PMS Filter: 32-63,mmsalt:238369,filter:ID%2=1 [Copy](#)

Активация
анализатора

[OPEN](#)

Unresolved

В итоге получаем



Влияние на ключевые метрики

Платформа	Действие	Метрика	Эффект (%)	Уверенность в отличии (%)	Уверенность в новизне (%)	Charts
All	Классы (EJB)	users	0.50%	100.00%	100.00%	details
All	Заклашенные не-друзья	users	0.39%	100.00%	99.66%	details
All	Заклашенные друзья	users	0.20%	100.00%	98.74%	details
All	Установка приложения (EJB)	users	-0.03%	15.01%	21.37%	details
All	Занесенные рубли	money	0.02%	15.01%	21.37%	details
All	Присоединение к группам (EJB)	users	0.01%	15.01%	21.37%	details
All	Принятие приглашения в друзья на той стороне (EJB)	counts	-0.01%	15.01%	21.37%	details
All	Запуск приложения (EJB)	users	0.00%	0.00%	0.00%	details
All	Отправка приглашения в друзья (EJB)	users	0.00%	0.00%	0.00%	details
All	Сообщения (EJB)	users	0.00%	0.00%	0.00%	details
All	Дарение (EJB)	users	0.00%	0.00%	0.00%	details
All	Принятие подарка другой стороной (EJB)	money	0.00%	0.00%	48.48%	details
All	Уники по dwh-cache	users	0.00%	0.00%	0.00%	details

В итоге получаем



Влияние на ключевые метрики

Платформа	Действие	Метрика	Эффект (%)	Уверенность в отличии (%)	Уверенность в новизне (%)	Charts
All	Классы (EJB)	users	0.50%	100.00%	100.00%	details
All	Заклашенные не-друзья	users	0.39%	100.00%	99.66%	details
All	Заклашенные друзья	users	0.20%	100.00%	98.74%	details
All	Установка приложения (EJB)	users	-0.03%	15.16%	21.37%	details
All	Занесенные рубли	money	0.02%	15.16%	21.37%	details
All	Присоединение к группам (EJB)	users	0.01%	15.16%	21.37%	details
All	Принятие приглашения в друзья на той стороне (EJB)	counts	-0.01%	15.16%	21.37%	details
All	Запуск приложения (EJB)	users	0.00%	0.00%	0.00%	details
All	Отправка приглашения в друзья (EJB)	users	0.00%	0.00%	0.00%	details
All	Сообщения (EJB)	users	0.00%	0.00%	0.00%	details
All	Дарение (EJB)	users	0.00%	0.00%	0.00%	details
All	Принятие подарка другой стороной (EJB)	money	0.00%	0.00%	48.48%	details
All	Уники по dwh-cache	users	0.00%	0.00%	0.00%	details

Отличие
статистически
значимо

В итоге получаем



Влияние на ключевые метрики

Эффект +0.2%

Платформа	Действие	Категория	Изменение (%)	Уверенность в отличии (%)	Уверенность в новизне (%)	Charts
All	Классы (EJB)			100.00%	100.00%	details
All	Заклашенные не-друзья	users	0.00%	100.00%	99.66%	details
All	Заклашенные друзья	users	0.20%	100.00%	98.74%	details
All	Установка приложения (EJB)	users	-0.03%	15.15%	21.37%	details
All	Занесенные рубли	money	0.02%		21.37%	details
All	Присоединение к группам (EJB)	users	0.01%		1.37%	details
All	Принятие приглашения в друзья на той стороне (EJB)	counts	-0.01%		1.37%	details
All	Запуск приложения (EJB)	users	0.00%		0.00%	details
All	Отправка приглашения в друзья (EJB)	users	0.00%	0.00%	0.00%	details
All	Сообщения (EJB)	users	0.00%	0.00%	0.00%	details
All	Дарение (EJB)	users	0.00%	0.00%	0.00%	details
All	Принятие подарка другой стороной (EJB)	money	0.00%	0.00%	48.48%	details
All	Уники по dwh-cache	users	0.00%	0.00%	0.00%	details

Отличие
статистически
значимо

В итоге получаем



Влияние на ключевые метрики

Платформа	Действие	Метрика	Изменение	Доверенность в %	Отличия в %	Charts
All	Классы (EJB)			100.00%		details
All	Заклашенные не-друзья	users	0.00%	100.00%	99.66%	details
All	Заклашенные друзья	users	0.20%	100.00%	98.74%	details
All	Установка приложения (EJB)	users	-0.03%	15.00%	21.37%	details
All	Занесенные рубли	money	0.02%		21.37%	details
All	Присоединение к группам (EJB)	users	0.01%		1.37%	details
All	Принятие приглашения в друзья на той стороне (EJB)	counts	-0.01%		1.37%	details
All	Запуск приложения (EJB)	users	0.00%		0.00%	details
All	Отправка приглашения в друзья (EJB)	users	0.00%	0.00%	0.00%	details
All	Сообщения (EJB)	users	0.00%	0.00%	0.00%	details
All	Дарение (EJB)	users	0.00%	0.00%	0.00%	details
All	Принятие подарка другой стороной (EJB)	money	0.00%	0.00%	48.48%	details
All	Уники по dwh-cache	users	0.00%	0.00%	0.00%	details

Эффект +0.2%

Отличие
является новым

Отличие
статистически
значимо



Открытые вопросы

- Осознанный выбор размера аудитории
- Анализ на других временных интервалах (час/неделя)
- Сертификация метрик
- Стабилизация тренда в эффекте
- Анализ экспериментов с вирусным эффектом
- Снижение ложных срабатываний
- Гетерогенный эффект

Спасибо за внимание!



одноклассники

Мы ищем новых коллег ;)

cv@odnoklassniki.ru