

Анализ рекламных креативов для ООО “Семечки”.

Цель:

Сравнить 3 рекламных креатива для продукта “Орехи со вкусом салями” и выбрать самый эффективный.

Задача:

Провести A/B-тест 3 креативов, один из которых уже запускался и показывал результат в 2,1% CTR, проанализировать данные и выбрать лучший.

Формализованные задачи:

1. Рассчитать необходимую для A/B-теста аудиторию.
2. Составить медиаплан(сохранить общий бюджет (165 000 руб.) и охват аудитории (мужчины 20–60 лет, Москва, Екатеринбург, СПб).
3. Подсчитать промежуточные результаты.
4. После получения итоговых результатов подсчитать статистически значимую разницу(успешным будет считаться изменение метрики на 0,1 процентных пункта (имеется в виду абсолютное увеличение),мощность 85%, $\alpha=5\%$)
5. Сделать выводы по полученным данным.

Расчет необходимой для A/B-теста аудитории.

Нам нужно посчитать, какой аудитории будет достаточно для теста второго и третьего креатива.

Для этого мы воспользуемся калькулятором.

Введем следующие значения



baseline conversion rate — базовая конверсия, которую мы принимаем за плановую(2,1%)
minimum Detectable Effect — минимальный эффект, который мы сможем заметить(0,1%)

statistical power $1-\beta$ — статистическая мощность(85%)
significance level α — уровень доверия(5%)

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: How many subjects are needed for an A/B test?

| | | | | | |
|---|----------------------------------|---|---|-----------|--------------------------|
| Baseline conversion rate: | <input type="text" value="2.1"/> | % |  | 2.1% | [link] |
| Minimum Detectable Effect: | <input type="text" value="0.1"/> | % |  | 2% – 2.2% | |
| <small>The Minimum Detectable Effect is the smallest effect that will be detected (1-β)% of the time.</small> | | | <input checked="" type="radio"/> Absolute <input type="radio"/> Relative | | |
| | | | <small>Conversion rates in the gray area will not be distinguishable from the baseline.</small> | | |

Sample size:
372,135
per variation

Statistical power $1-\beta$: 85% Percent of the time the minimum effect size will be detected, assuming it exists
Significance level α : 5% Percent of the time a difference will be detected, assuming one does NOT exist

Мы рассчитали в калькуляторе размер необходимой нам аудитории для проведения второго и третьего креативов.
Нам необходимо иметь аудиторию в размере 372135 человек. Округлим эту цифру до **373000**.

Медиаплан.

Общий бюджет составляет 165000 рублей, на старый, основной креатив мы выделим около 45% бюджета. Это составляет 75480 рублей. Оставшуюся сумму мы разделим между 2м и 3м креативами - по 44760 рублей.

| | |
|--------------|--|
| Рекламодат | ООО Семечки |
| Бренд | Снеки |
| Продукт | Орехи со вкусом салями |
| Целевая ауд | Мужчины 20-60, интересуются просмотром футбольных матчей |
| Длительность | Месяц |
| Начало | 01.04.2025 |
| Конец | 30.04.2025 |

| Номер кампании | Площадка | Таргетинг | Формат | Динамическая цена или фиксированна | Период | Единица заку | Цена закупки | Кол-во закупок | Бюджет, без НДС | Количество показов | Прогнозируемый охват | Прогнозируемая средняя | CPT, руб / прогнозируемый | CTR, % / прогнозируемый | Клики прогнозируемые | CPC, руб./ прогнози |
|----------------|-----------|--------------------------------|--------|------------------------------------|--------------------------|--------------|--------------|----------------|-----------------|--------------------|----------------------|------------------------|---------------------------|-------------------------|----------------------|---------------------|
| 1 | Майтаргет | Веб и мобайл, муж. 20 - 60 GEO | Баннер | Dynamic | 01.04.2025 30.04.2025 | 1000 impr. | 60,00 Р | 1 258 | 75 480,00 RUR | 1 258 000 | 629 000 | 2 | 60,00 RUR | 2,10% | 26 418 | 2,86 RUR |
| 2 | Майтаргет | Веб и мобайл, муж. 20 - 60 GEO | Баннер | Dynamic | 01.04.2025 30.04.2025 | 1000 impr. | 60,00 Р | 746 | 44 760,00 RUR | 746 000 | 373 000 | 2 | 60,00 RUR | 2,20% | 16 412 | 2,73 RUR |
| 3 | Майтаргет | Веб и мобайл, муж. 20 - 60 GEO | Баннер | Dynamic | 01.04.2025 30.03.2025 | 1000 impr. | 60,00 Р | 746 | 44 760,00 RUR | 746 000 | 373 000 | 2 | 60,00 RUR | 2,20% | 16 412 | 2,73 RUR |
| Total: | | | | | | | | | 165 000,00 RUR | 2 750 000 | 1 375 000 | 2 | 60,00 RUR | 2,15% | 59 242 | 2,79 RUR |

| | |
|--------|----------------|
| Итого: | |
| Бюджет | 165 000,00 RUR |
| Бюджет | 206 250,00 RUR |

Подсчет промежуточных результатов.

Для кампаний получены следующие предварительные результаты:

Первая кампания: 45346 показов, 861 клик;

Вторая кампания: 38437 показов, 663 клика;

Третья кампания: 44836 показов, 762 клика.

Мы не можем сейчас принять решение по отключению кампаний. Выводы делать рано: мы ещё не достигли выборки, которую заранее спланировали для этого эксперимента. Мы можем совершить **ошибку подглядывания** и сделать неправильные выводы.

Ошибка подглядывания в A/B-тестах — это преждевременная проверка статистической значимости результатов во время проведения эксперимента, что может привести к ложным выводам и принятию неверных решений.

Почему возникает ошибка подглядывания?

В A/B-тестах данные собираются постепенно, и если аналитик или система часто проверяют p-value или другие метрики до завершения теста, высока вероятность:

Ложноположительных результатов (False Positive) — увидеть "значимый" эффект там, где его нет (из-за случайных колебаний).

Неправильной остановки теста — раннее завершение теста при кажущемся "успехе" или "провале", хотя реальный эффект мог бы проявиться позже.

Пример ошибки подглядывания

Допустим, вы запустили A/B-тест длительностью 7 дней, но начинаете проверять результаты каждый день:

День 1: $p\text{-value} = 0.03$ (кажется, есть значимый эффект!).

День 3: $p\text{-value} = 0.4$ (эффект "исчез").

День 7: $p\text{-value} = 0.1$ (граница значимости не достигнута).

Если бы вы остановились на Дне 1, то ошибочно решили бы, что вариант В лучше, хотя это был просто случайный шум.

Как избежать подглядывания в A/B-тестах?

1. Фиксируйте размер выборки заранее

Рассчитайте необходимый объем данных до старта теста.

Не останавливайте тест, пока не наберете нужное количество наблюдений (если нет форс-мажора).

2. Используйте методы корректировки

Поправка на множественные сравнения (например, поправка Бонферрони).

Последовательный анализ (Sequential Testing):

Алгоритмы типа SPRT (Sequential Probability Ratio Test) или Bayesian A/B-тесты, которые позволяют проверять данные по мере поступления, но с корректировкой значимости.

3. Остановка только по правилам

Предопределите правила остановки (например, "останавливаем тест только если $p\text{-value} < 0.01$ после 80% данных").

Избегайте ручных остановок из-за "интересных" промежуточных результатов.

4. Следите за временными трендами

Если метрика зависит от дня недели (например, конверсия выше в выходные), не сравнивайте первые 2 дня с последними 2 дня — дождитесь полного цикла.

Анализ результатов.

После месяца тестирования были получены результаты. Необходимая аудитория набрана по всем креативам.

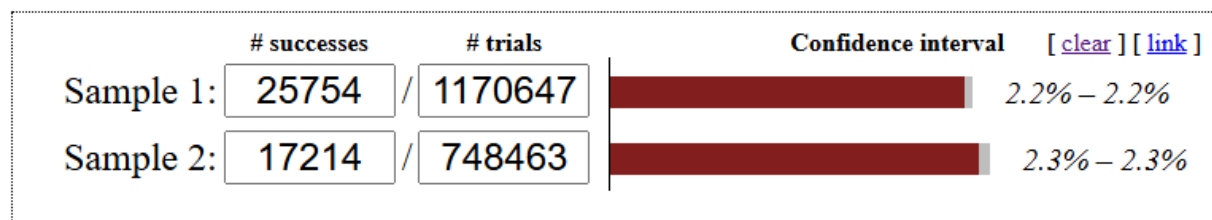
| Period | Views | Clicks | Campaign number |
|-------------------------|---------|--------|-----------------|
| 01.04.2025 - 30.04.2025 | 1170647 | 25754 | 1 |
| 01.04.2025 - 30.04.2025 | 748463 | 17214 | 2 |
| 01.04.2025 - 30.04.2025 | 749745 | 12745 | 3 |

Проверим статистическую значимость результатов, загрузив наши данные в калькулятор и воспользовавшись методом «Критерий хи-квадрат».

Сравнение первого и второго креативов..

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

Sample 2 is more successful



($p < 0.001$)

Confidence level: 95%

Как мы видим, второй креатив показывает лучший результат, чем первый (CTR 2,3% и 2,2% соответственно). Этот результат является статистически значимым ($p\text{-value} < 0.001$).

Сравнение первого и третьего креативов.

Question: Does the rate of success differ across two groups?

| | # successes | # trials | Confidence interval | [clear] [link] |
|-----------|-------------|----------|--|--|
| Sample 1: | 25754 | 1170647 |  | 2.2% – 2.2% |
| Sample 2: | 12745 | 749745 |  | 1.7% – 1.7% |

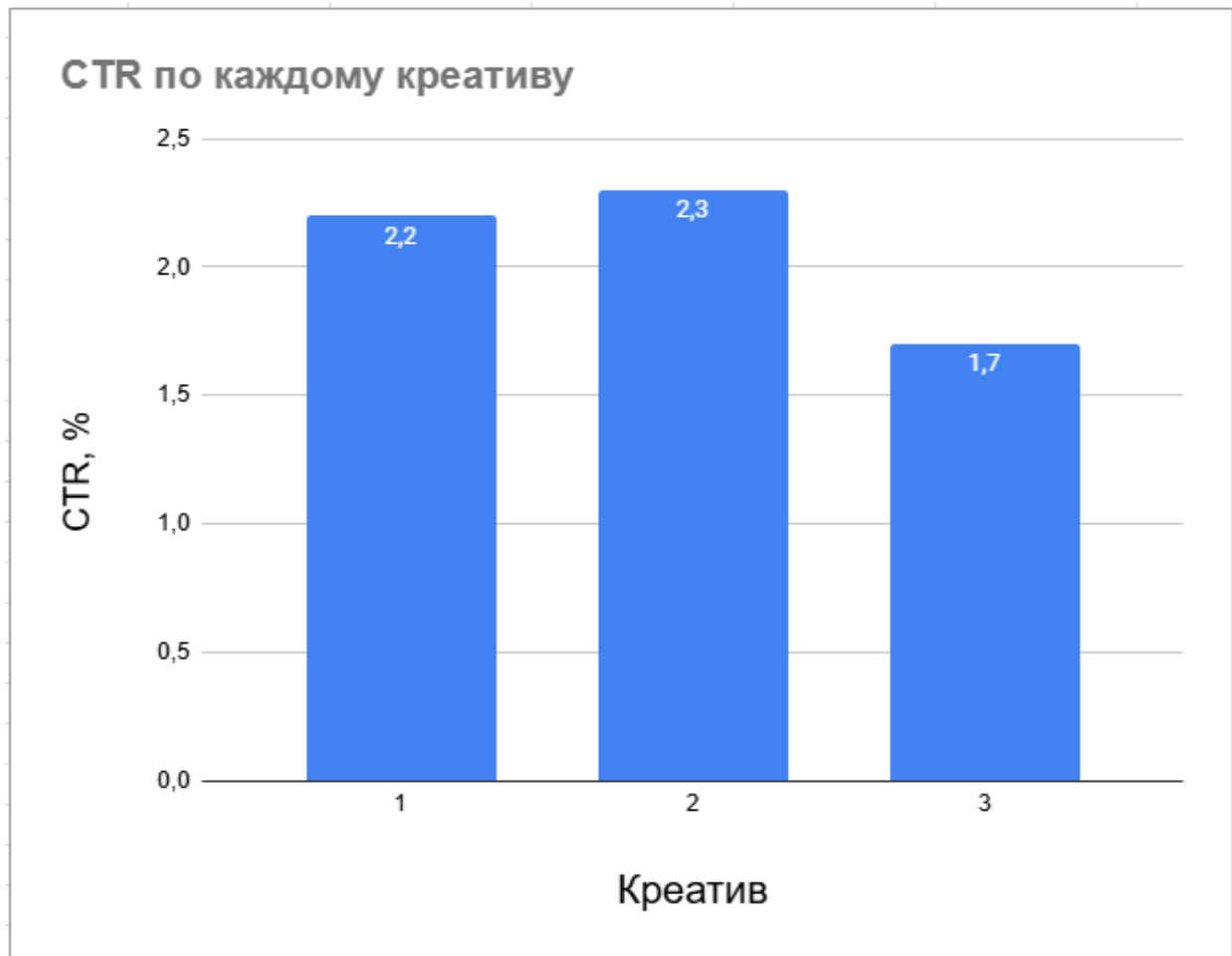
Verdict:

Sample 1 is more successful

($p < 0.001$)

Confidence level: 95%

Как мы видим, третий креатив значительно отстает от первого со статистически значимой разницей(CTR 1.7% и 2,2% соответственно, $p\text{-value} < 0.001$).



Выводы и рекомендации.

При проведении A/B-теста была выявлена статистически значимая разница между всеми креативами, тест можно считать успешным. Самым эффективным оказался второй креатив(CTR 2,3%), старый креатив уступает ему в эффективности со статистически значимой разницей(CTR 2.2%), третий креатив оказался самым неэффективным(CTR 1.7%).

Рекомендуем:

1. Выделить основную часть бюджета(80%) на второй креатив ввиду его максимальной эффективности.
2. Сохранить первый креатив с бюджетом 20%.
3. Закрывать третий креатив по причине его неэффективности.

