**Кандидат: Арбузина Арина Александровна**

**Решение тестового задания**

**Первое задание**.

В задании анализируется использование пользователями фильтров на сайте.

При решении задания я отвечала на вопросы:

1. *Сколько людей пользуются фильтрами?*
2. *В каких городах фильтрами пользуются больше? в каких меньше?*
3. *Какие разделы фильтров наиболее востребованы? “фильтры”, “сортировка”, “категории”?*
4. *Как часто люди пользуются выбором цены?*

Подробные шаги выполнения задания, ответы на вопросы, все возникшие мысли и обнаруженные инсайты приведены в ноутбуке “Задание 1” ([ссылка](https://drive.google.com/file/d/1veqkWw1dY5qVho8H3BsdzH38wjA_UQNy/view?usp=sharing)).

**Второе задание.**

В задаче исследуется эффект добавления нового фильтра на количество бронирований.

Код решения задачи представлен в ноутбуке “Задание 2” ([ссылка](https://drive.google.com/file/d/16fjKwvVRrdFC7CV4Ziai_p8psVTbLyp0/view?usp=sharing)).

**Решение:**

По данным за неделю, количество бронирований в группе Б (с новым фильтром) больше, чем в группе А (без нового фильтра) (450 > 300). Необходимо проверить, что это различие является статистически значимым и новый фильтр увеличивает количество бронирований.

1. *Опиши, какой результат мы хотим достичь с помощью нового фильтра, на какие метрики стоит смотреть.*

Результат, который мы хотим достичь с помощью нового фильтра, - это увеличение количества бронирований.

Целевая метрика, которую я буду сравнивать в двух группах, - это конверсия из перехода на сайт в бронирование, так как количество переходов на сайт одинаковое в группах.

1. *Подсчитай конверсию в бронирование для каждой группы.*

Значение конверсии в каждой группе:

* В группе Б: *p1* = 450 / 5000 = 0.09
* В группе А: *p2* = 300 / 5000 = 0.06

1. *Оцени, насколько новый фильтр повлиял на увеличение бронирований.*

В данном эксперименте один заход на сайт - это случайная величина. Она имеет биномиальное распределение: каждое оформление бронирования пользователем при заходе на сайт - это успех, каждое “не оформление” бронирования пользователем при заходе на сайт - это неудача. По сути, конверсия отражает относительное число успехов, и оно в каждой группе свое (*p1* и *p2*).

Можно составить такую таблицу успехов и неудач для каждой из групп:

|  | Кол-во успехов | Кол-во неудач |
| --- | --- | --- |
| Группа А | 300 | 4700 |
| Группа Б | 450 | 4550 |

И, если перейти к вероятностям (относительным величинам), можно получить закон распределения случайной величины X для каждой из групп:

|  | P(X = 1) (успех) | P(X = 0) (неудача) |
| --- | --- | --- |
| Группа А | 0.06 | 0.94 |
| Группа Б | 0.09 | 0.91 |

Сформулирую нулевую и альтернативную гипотезы:

* H0: *p1 <= p2*. Конверсия в группе Б не превышает конверсию в группе А
* H1: *p1 > p2*. Конверсия в группе Б выше, чем конверсия в группе А

Для проверки нулевой гипотезы воспользуюсь z-тестом для сравнения конверсий в двух группах. В функцию передам параметр alternative="larger" для задания вида альтернативной гипотезы и вида p-value (односторонний). Порог для p-value для принятия решения - *alpha = 0.05*.

from statsmodels.stats.proportion import proportions\_ztest

success\_cnts = np.array([450, 300])

total\_cnts = np.array([5000, 5000])

test\_stat, p\_value = proportions\_ztest(count=success\_cnts, nobs=total\_cnts, alternative='larger')

>>print(p\_value)

6.296795408748488e-09

При нулевой гипотезе о том, что новый фильтр не увеличит число бронирований, вероятность получения таких или еще более выраженных отличий, как в наблюдаемых данных, составляет 6.3\*10^(-9), то есть практически нулевая вероятность. Поскольку это значение ниже выбранного порога уровня значимости (6.3\*10^(-9) < 0.05), то есть основания отклонить нулевую гипотезу в пользу принятия альтернативной. Следовательно, конверсия в группе Б выше, чем в группе А, и это различие статистически значимо.

Рассчитаю 95 % доверительные интервалы, чтобы найти, в каких границах лежит конверсия в каждой из групп. Для этого воспользуюсь расчетом доверительных интервалов с помощью модуля api из библиотеки statsmodels:

import statsmodels.api as sm

low\_a, high\_a = sm.stats.proportion\_confint(count=300, nobs=5000, alpha=0.05, method='wilson')

>>> print(low\_a, high\_a)

>>> 0.05374897325570841, 0.06692660445594185

low\_b, high\_b = sm.stats.proportion\_confint(count=450, nobs=5000, alpha=0.05, method='wilson')

>>> print(low\_b, high\_b)

>>> 0.08237915058874333, 0.0982503650062035

Разброс конверсии в группе А составляет: 0.054 - 0.067.

Разброс конверсии в группе Б составляет: 0.082 - 0.098.

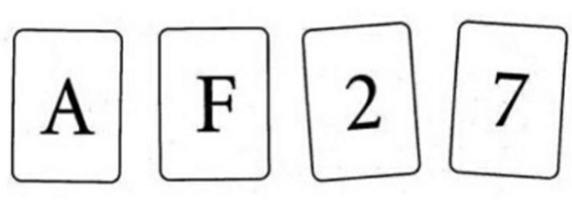
Полученные доверительные интервалы не пересекаются, что еще раз подтверждает утверждение о том, различие конверсий в двух группах является статистически значимым.

1. *Стоит ли внедрять данный фильтр на основании полученных результатов?*

На основании полученных результатов эксперимента, предлагается принять решение о внедрении нового фильтра.

**Третье задание.**

Есть 4 карточки:



Необходимо определить, какие карточки обязательно необходимо перевернуть, чтобы проверить, выполняется ли для этих четырех карточек правило:

**“Если на одной стороне карточки – гласная буква, то на другой стороне этой карточки должно быть четное число”**

**Решение:**

С обратной стороны любой карточки может быть все, что угодно: гласная буква, согласная буква, четное число, нечетное число или вообще что-то другое (рисунок крокодила, например).

Чтобы проверить правило, необходимо понять, какие исходы могут подтвердить или опровергнуть это правило.

Рассмотрю все исходы в зависимости от того, что находится на лицевой стороне карточки:

| Карточка | Лицевая сторона | Обратная сторона | Отношение в правилу |
| --- | --- | --- | --- |
| A | гласная | четное | подтверждает |
| гласная | все что угодно, кроме четного числа | опровергает |
| F | согласная | все что угодно | никак не регулируется правилом |
| 2 | четное | все что угодно | никак не регулируется правилом |
| 7 | нечетное | гласная | опровергает |
| нечетное | все что угодно, кроме гласной | никак не регулируется правилом |

В итоге, на факт выполнения правила влияет обратная сторона двух карточек:

* перевернув карточку “А”, мы опровергаем или подтверждаем правило;
* перевернув карточку “7”, мы можем обнаружить опровержение правила.

**Ответ:**

Необходимо перевернуть всего две карточки для проверки правила: карточку с гласной буквой “A” и карточку с нечетным числом “7”.