



СТАТИСТИКА. T-TEST

Варламов Никита

<https://github.com/NikitaVarlamov>

ИСПОЛЬЗУЕМЫЕ БИБЛИОТЕКИ И ИНСТРУМЕНТЫ





ГИПОТЕЗА №1

Менеджер сайта, предоставляющего независимым продавцам площадку для продаж, решил провести АБ-тест, выбрав в качестве метрики время обработки заказов продавцами.

Для контрольной группы продавцов оставили предыдущий интерфейс работы с заказами, а для тестовой группы внедрили интерактивный дэшборд заказов.

Представлены результаты времени обработки заказов в часах для тестовой и контрольной групп.

Необходимо проверить гипотезу о том, что использование интерактивного дэшборда уменьшило время обработки заказов.

ФОРМУЛИРОВКА ГИПОТЕЗЫ

НУЛЕВАЯ ГИПОТЕЗА

$$H_0: \mu_1 = \mu_2$$

Использование альтернативного дашборда не оказало никакого влияния на время обработки заказов продавцом.

АЛЬТЕРНАТИВНАЯ ГИПОТЕЗА

$$H_1: \mu_1 > \mu_2$$

Использование изменений в интерфейсе площадки агрегатора уменьшило время обработки заказов.

Уровень значимости альфа определим равным $\alpha = 0.05$



АЛГОРИТМ РЕШЕНИЯ

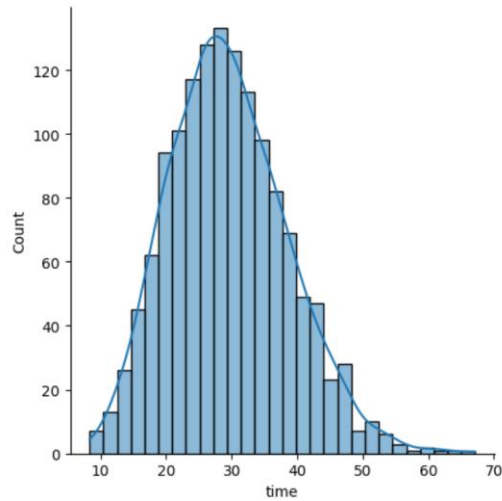
ОБЗОР ДАННЫХ

- Первичный обзор данных
- Вывод основных статистических сведений
- Визуальная оценка плотности распределения данных (гистограмма + KDE/ЯОП)

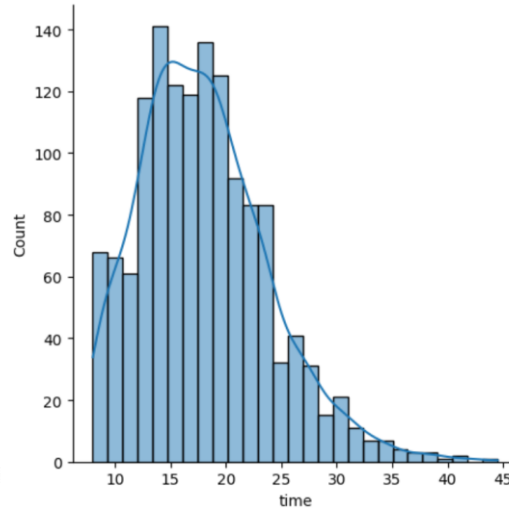
СТАТИСТИЧЕСКАЯ ОЦЕНКА ДАННЫХ

- Сравнение стандартные отклонений выборок
- Проведение Т-теста для независимых выборок
- Интерпретация результатов

1. Kernel Density Estimation, KDE или ядерная оценка плотности, ЯОП - один из непараметрических способов оценки плотности случайной величины, особенностью которого является сглаживание данных.



Плотность распределения данных.
Контрольная группа



Плотность распределения данных.
Тестовая группа

3. Т-тест, интерпретация результатов

Результаты t-теста:

t-статистика: 40.01056

p-значение: 2.9949647128013707e-277

`p_value < alpha`

Отвергаем нулевую гипотезу. Использование интерактивного дэшборда значительно уменьшило время обработки заказов.

2. Сравним стандартные отклонения выборок, для определения необходимости применения теста Уэлча.

```
control_std = control['time'].std( )  
test_std = test['time'].std( )
```

Стандартное отклонение контрольной выборки: 8.9799
Стандартное отклонение тестовой выборки: 5.8578

3. Разница между стандартными отклонениями двух выборок, очевидно и ожидаемо, не существенна.

Применим двухвыборочный Т-тест для независимых выборок.

Уровень значимости определим равным $\alpha = 0.05$

```
t_statistic, p_value = st.ttest_ind(  
    control['time'],  
    test['time'],  
    equal_var=True,  
    alternative = "greater")
```

```
print("Результаты t-теста:")  
print(f"t-статистика: {t_statistic:.5f}")  
print(f"p-значение: {p_value}")
```

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

```
# Зададим доверительный уровень
confidence_level = 0.9; alpha = 0.1
```

```
# Средние и стандартные отклонения для двух выборок
mean_control_group = control['time'].mean()
std_control_group = control['time'].std()
```

```
mean_test_group = test['time'].mean()
std_test_group = test['time'].std()
```

```
# Размер выборок
n_control_group = len(control)
n_test_group = len(test)
```

```
# Определение Zα/2 для 90% уровня доверия
z_control_group = 1.65
z_test_group = 1.65
```

Расчёт доверительного интервала производится по формуле:

$$\bar{X} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

```
# Расчёт доверительных интервалов выборок
confidence_interval_low_control = mean_control_group - z_control_group *
(std_control_group / np.sqrt(n_control_group))
confidence_interval_high_control = mean_control_group + z_control_group *
(std_control_group / np.sqrt(n_control_group))
```

```
confidence_interval_low_test_group = mean_test_group - z_test_group *
(std_test_group / np.sqrt(n_test_group))
confidence_interval_high_test_group = mean_test_group + z_test_group *
(std_test_group / np.sqrt(n_test_group))
```

Доверительный интервал для среднего времени обработки заказа без использования интерактивного дашборда: (29.106, 29.900)

Доверительный интервал для среднего времени обработки заказа с использованием интерактивного дашборда: (17.750, 18.268)

Инструменты статистики scipy позволяют произвести подобный расчет в одну формулу, которая принимает в себя три значения:

- доверительный уровень (90% в нашем случае)

- среднее значение выборки (Mean)

- стандартную ошибку среднего (Standard Error of Mean - SEM)

Размер выборок контрольной и тестовой групп достаточно велик, поэтому, согласно ЦПТ, построим доверительный интервал с использованием нормального распределения.

```
interval_con = st.norm.interval(
    confidence=0.9,
    loc=np.mean(control['time']),
    scale=st.sem(control['time']))
```

```
interval_test = st.norm.interval(
    confidence=0.9,
    loc=np.mean(test['time']),
    scale=st.sem(test['time']))
```

Для контрольной выборки 90% доверительный интервал составляет: (29.1071 , 29.8986)

Для тестовой выборки 90% доверительный интервал составляет: (17.7510 , 18.2673)

То есть, в 90 процентах случаев наш доверительный интервал будет включать истинное среднее значение времени обработки заказа.



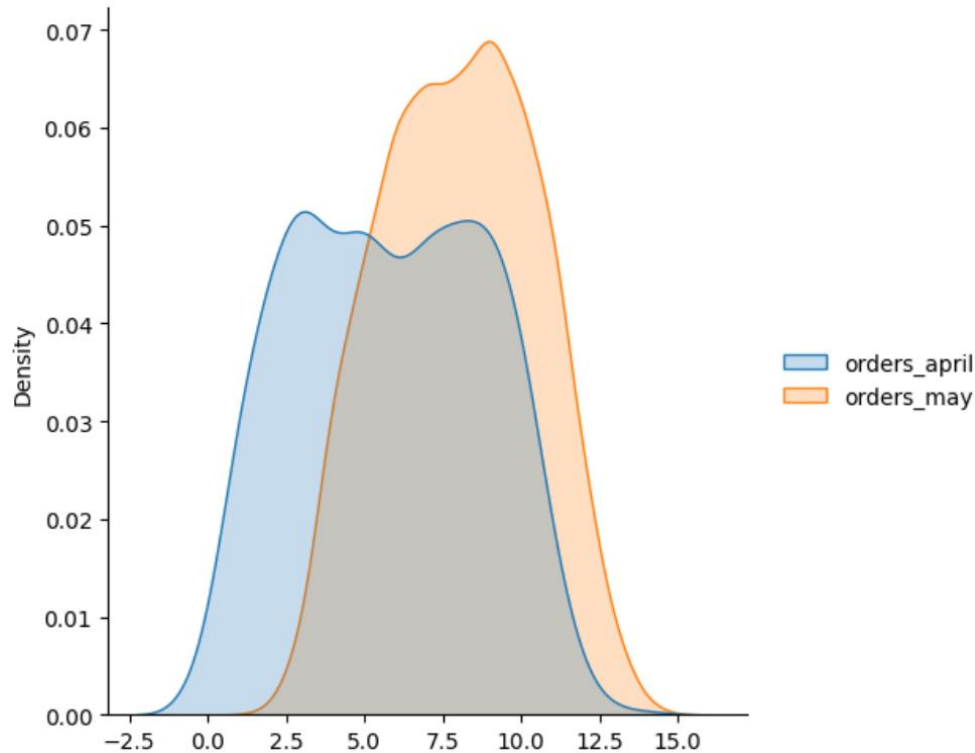
ГИПОТЕЗА №2

Маркетплейс предоставил с 1 мая часто заказывающим клиентам бесплатную доставку, действующую для всех заказов до конца календарного месяца, если в этом месяце клиент уже сделал пять заказов (то есть начиная с шестого заказа).

Представлена информация о количестве заказов, которое тысяча случайно отобранных часто заказывающих клиентов сделали в апреле и мае.

Необходимо проверить гипотезу о том, что клиенты стали делать больше заказов после введения бесплатной доставки с шестого заказа.

1. Визуальная оценка плотности распределения данных (гистограмма + KDE/ЯОП)



3. Т-тест, интерпретация результатов

Результаты t-теста:

t-статистика: 18.17200

p-значение: 2.8066025158955254e-64

$p_value < \alpha$

Отвергаем нулевую гипотезу. Клиенты стали делать больше заказов после введения акции на бесплатную доставку.

2. Применим двухвыборочный Т-тест для зависимых выборок.

Апрель: index = 1, май: index = 2

➤ Нулевая гипотеза H_0 :

$\mu_1 = \mu_2$, или $\mu_d = 0$, т.е.

введение бесплатной доставки с 6-ого заказа не повлияло на число заказов за текущий месяц.

➤ Альтернативная гипотеза H_1 :


$\mu_1 < \mu_2$, т.е.

введение нового бонуса увеличило число заказов за месяц.

Уровень значимости $\alpha = 0.05$

3. Т-тест для сравнения среднего значения (зависимые выборки)

```
t_statistic_rel,  
p_value_rel = st.ttest_rel(  
    orders['orders_may'],  
    orders['orders_april'],  
    alternative = "greater")
```



Представлена анонимизированная информация о продавцах маркетплейса: тип продукции (goods) и время осуществления продаж через этот маркетплейс (experience).

Переменная goods принимает значения:

clothes — одежда

electronics — электроника и техника

outdoor — товары для активного отдыха

beauty — косметика и товары для ухода

pets — товары для животных

Переменная experience принимает значения:

0-1 — до одного года продаж

1-3 — от одного до трёх лет продаж

3-5 — от трёх до пяти лет продаж

5 — от пяти лет продаж

ГИПОТЕЗА №3

Необходимо проверить гипотезу о том, что стаж продаж на маркетплейсе не зависит от типа реализуемой продукции.

1. Построение таблицы сопряженности

```
table = pd.crosstab(
    vendors['goods'],
    vendors['experience'])
```

experience	0-1	1-3	3-5	5-
goods				
beauty	40	65	22	26
clothes	104	129	42	46
electronics	68	67	29	31
outdoor	36	46	23	26
pets	13	17	4	5

Degrees of freedom (df)	Significance level (α)							
	.99	.975	.95	.9	.1	.05	.025	.01
1	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217

2. Формулировка гипотезы

- Нулевая гипотеза H_0 : Стаж продаж на маркетплейсе не зависит от типа реализуемой продукции.
- Альтернативная гипотеза H_1 : Стаж продаж на маркетплейсе зависит от типа реализуемой продукции.

Уровень значимости $\alpha = 0.05$

3. Проверка гипотезы путем сравнения Хи-квадрат статистики с критическим значением $\chi^2(5\%) = 21.026$, согласно табличным данным.

```
res = chi2_contingency(table)
```

хи-квадрат статистика: 9.17797

p-value: 0.68766

степеней свободы: 12

рассчитанные ожидаемые значения: [

```
[ 47.59594756  59.08462455  21.88319428  24.43623361]
[ 99.85816448 123.96185936  45.91179976  51.2681764 ]
[ 60.66150179  75.30393325  27.89034565  31.14421931]
[ 40.75208582  50.58879619  18.73659118  20.92252682]
[ 12.13230036  15.06078665  5.57806913  6.22884386]]
```

Не отвергаем нулевую гипотезу. Нет достаточных (значимых) оснований полагать, что стаж продаж на маркетплейсе зависит от типа реализуемой продукции.

Abstract geometric lines in the top-left corner of the slide, consisting of several thin, light brown lines that intersect to form various polygons and shapes.

СПАСИБО

Варламов Никита

<https://github.com/NikitaVarlamov>

varlamov.n.s@gmail.com