Проверка гипотезы о равенстве дисперсий

$$H_0 : \sigma_x^2 = \sigma_y^2 H_1 : \sigma_x^2 \neq \sigma_y^2$$

- **Нулевая гипотеза (H0):** $\sigma_x^2 = \sigma_y^2$ - это нулевая гипотеза, которую мы хотим проверить. Она утверждает, что дисперсии двух выборок X и Y равны.
- **Альтернативная гипотеза (H1):** $\sigma_x^2 \neq \sigma_y^2$ - это альтернативная гипотеза, которая утверждает, что дисперсии выборок X и Y не равны.

Расчет дисперсий:

Для начала, мы вычисляем выборочные дисперсии для обеих выборок:

- s_x^2: Выборочная дисперсия для выборки X вычисляется как среднее квадратов отклонений каждого элемента выборки X от ее выборочного среднего $\overline{x}$, умноженное на $\frac{1}{n_x-1}$, где n_x - размер выборки X.

$$s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \overline{x})^2$$

```
x_var = data.X.var(ddof=1)
print(f'Дисперсия первой выборки: {x_var}')
```

- s_y^2: Аналогично, выборочная дисперсия для выборки Y вычисляется как среднее квадратов отклонений каждого элемента выборки Y от ее выборочного среднего $\overline{y}$, умноженное на $\frac{1}{n_y-1}$, где n_y - размер выборки Y.

$$s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \overline{y})^2$$

```
y_var = data.Y.var(ddof=1)
print(f'Дисперсия второй выборки: {y_var}')
```

Вычисленное значение критерия Фишера

Затем мы вычисляем значение критерия Фишера (F), который является отношением выборочных дисперсий:

$$F = \frac{\sigma_x^2}{\sigma_y^2} \text{если } S_x > S_y, \quad F = \frac{\sigma_y^2}{\sigma_x^2} \text{если } S_x < S_y$$

Это значение F показывает, насколько выборочные дисперсии отличаются друг от друга.

```
f_r = y_var / x_var
Markdown(f'Вычисленное значение критерия: $F={f_r}$')
```

Теоретическое значение критерия Фишера

Далее мы вычисляем теоретическое значение критерия Фишера ($F_{кр}$), которое является критическим значением для заданного уровня значимости α, степени свободы для выборок X и Y (соответственно $n_x - 1$ и $n_y - 1$).

- $F_{кр}$ вычисляется с использованием функции обратного преобразования распределения F (инверсия функции распределения).

$$F_{кр} = F(\alpha/2, n_x - 1, n_y - 1)$$

```
f_t = st.f.ppf(1-a/2, n_x - 1, n_y - 1)
print(f'Теоретическое значение критерия: {f_t}')
```

Вывод о принятии или не принятии гипотезы

Наконец, мы сравниваем вычисленное значение критерия Фишера (F) с теоретическим значением ($F_{кр}$):

- Если $F > F_{кр}$ или $F < \frac{1}{F_{кр}}$, то у нас есть статистически значимые различия между дисперсиями выборок X и Y, и мы отвергаем нулевую гипотезу.
- Если $F \leq F_{кр}$ и $\frac{1}{F_{кр}} \leq F$, то у нас нет статистически значимых различий между дисперсиями выборок X и Y, и нулевая гипотеза не отвергается.

```
if f_r < f_t:
    display(Markdown('Гипотеза о равенстве дисперсий принимается, т.к. $F<F_{кр}$ ($H_0$)'))
else:
    display(Markdown('Гипотеза о равенстве дисперсий отвергается ($H_0$), принимается альтернативная гипотеза ($H_1$)'))
```

Этот анализ позволяет определить, есть ли статистически значимые различия в дисперсиях между двумя выборками.

Проверка гипотезы о равенстве генеральных средних

$$H_0 : \mu_x = \mu_y H_1 : \mu_x \neq \mu_y$$

- **Нулевая гипотеза (H0):** $\mu_x = \mu_y$ - это нулевая гипотеза, которую мы хотим проверить. Она утверждает, что средние значения двух выборок X и Y равны.
- **Альтернативная гипотеза (H1):** $\mu_x \neq \mu_y$ - это альтернативная гипотеза, которая утверждает, что средние значения выборок X и Y не равны.

Вычисление выборочных средних:

Сначала вычисляются выборочные средние для обеих выборок:

- $\overline{x}$: Вычисление среднего значения выборки X как сумма всех элементов выборки, поделенная на количество элементов в выборке n_x.

$$\overline{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$$

```
x_mean = x.mean()
print(f'Выборочное среднее первой выборки: {x_mean}')
```

- $\overline{y}$: Вычисление среднего значения выборки Y аналогично, как сумма всех элементов выборки, поделенная на количество элементов в выборке n_y.

$$\overline{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$$

```
y_mean = y.mean()
print(f'Выборочное среднее второй выборки: {y_mean}')
```

Эти значения представляют собой выборочные оценки средних генеральных совокупностей.

Вычисление дисперсий выборок:

Затем вычисляется объединенная выборочная дисперсия S^2, которая используется в дальнейшем для вычисления t-критерия:

- S^2: Вычисление объединенной выборочной дисперсии S^2 как взвешенное среднее дисперсий выборок X и Y, учитывая их размеры и выборочные дисперсии.

$$S^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

```
S_pow_2 = ((n_x - 1) * x.var(ddof=0) + (n_y - 1) * y.var(ddof=0))/(n_x + n_y - 2)
```

```
Markdown(f"$S^2 = {S_pow_2}$")
```

Вычисление t-критерия Стьюдента:

После вычисления выборочных средних и объединенной выборочной дисперсии, мы вычисляем t-критерий Стьюдента (t):

- t: Вычисление t-критерия, который является мерой того, насколько выборочные средние $\overline{x}$ и $\overline{y}$ различаются относительно объединенной выборочной дисперсии S^2
.

$$t = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

```
t_r = (x_mean - y_mean) / ((S_pow_2 * (1 / n_x + 1 / n_y) ) ** 0.5)
print(f'Вычисленное значение критерия: {t_r}')
```

Вычисление табличного t-криетрия Стьюдента:

Затем вычисляется табличное значение t-критерия Стьюдента ($t_{кр}$):

- $t_{кр}$: Это критическое значение, которое зависит от заданного уровня значимости α и числа степеней свободы, равного сумме степеней свободы выборок X и Y минус 2.

$$t_{кр} = t_{1-\frac{\alpha}{2}, n_x + n_y - 2}$$

```
t_t = st.t.ppf(1 - a / 2, n_x + n_y - 2)
print(f'Табличное значение критерия: {t_t}')
```

Вычисление Z:

В вашей процедуре также включено вычисление Z-критерия для проверки гипотезы, если известны дисперсии генеральных совокупностей:

- Z: Это Z-критерий, который используется для проверки гипотезы о различии средних значений, учитывая известные дисперсии σ_x^2 и σ_y^2.

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

```
z = (x_mean - y_mean) / ((x_var / n_x + y_var / n_y) ** 0.5)
print(f'Вычисленное значение критерия: {z}')
```

Вычисление табличного Z

Наконец, вычисляется табличное значение Z-критерия ($z_{кр}$):

- $z_{кр}$: Это критическое значение Z-критерия, которое зависит от заданного уровня значимости α и используется для сравнения с вычисленным Z-критерием. Если вычисленный Z-критерий превышает по абсолютной величине это табличное значение, то нулевая гипотеза отвергается.

```
z_critical = st.norm.ppf(1 - a/2)
print(f'Табличное значение критерия: {z_critical}')
```

Этот анализ позволяет определить, есть ли статистически значимые различия между средними значениями двух выборок X и Y на основе выборочных данных и известных дисперсий (в случае использования Z-критерия) или объединенной выборочной дисперсии (в случае использования t-критерия Стьюдента).

Проверка гипотезы о равенстве вероятностей благоприятного исхода в двух сериях

$$H_0 : p_1 = p_2 H_1 : p_1 \neq p_2$$

- **Нулевая гипотеза (H0):** $p_1 = p_2$ - это нулевая гипотеза, которую мы хотим проверить. Она утверждает, что вероятности благоприятного исхода в двух сериях p_1 и p_2 равны.
- **Альтернативная гипотеза (H1):** $p_1 \neq p_2$ - это альтернативная гипотеза, которая утверждает, что вероятности благоприятного исхода в двух сериях p_1 и p_2 не равны.

Вычисление вероятностей

Сначала вычисляются вероятности благоприятного исхода в каждой серии:

- W_1: Вычисление вероятности благоприятного исхода в первой серии как отношение количества благоприятных исходов m_1 к общему количеству наблюдений n_1 в первой серии.

$$W_1 = \frac{m_1}{n_1}$$

```
w1 = m1 / n1
Markdown(f'Вероятность благоприятного исхода в первой серии: $W_1 = {w1}$')
```

- W_2: Аналогично, вычисление вероятности благоприятного исхода во второй серии как отношение количества благоприятных исходов m_2 к общему количеству наблюдений n_2 во второй серии.

$$W_2 = \frac{m_2}{n_2}$$

```
w2 = m2 / n2
Markdown(f'Вероятность благоприятного исхода в первой серии: $W_2 = {w2}$')
```

Также вычисляется общая вероятность p для обеих серий, объединяя количество благоприятных исходов и общее количество наблюдений:

- p: Вычисление общей вероятности благоприятного исхода в обеих сериях как отношение суммы количества благоприятных исходов в обеих сериях $(m_1 + m_2)$ к суммарному общему количеству наблюдений в обеих сериях $(n_1 + n_2)$.

$$p = \frac{(m1 + m2)}{(n1 + n2)}$$

```
p = (m1 + m2) / (n1 + n2)
Markdown(f'Вероятность благоприятного исхода в обеих сериях: $p = {p}$')
```

Вычисление критерия

Затем мы вычисляем значение Z-критерия (z), который является мерой того, насколько различаются вероятности благоприятного исхода в двух сериях относительно общей вероятности p:

- z: Вычисление Z-критерия, который определяется как разница между вероятностями W_1 и W_2, деленная на стандартное отклонение этой разницы. Стандартное отклонение рассчитывается с использованием общей вероятности p и размеров выборок n_1 и n_2.

$$z = \frac{W_1 - W_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

```
z = (w1 - w2) / np.sqrt(p * (1 - p) * (1 / n1 + 1 / n2))
print(f'Вычисленное значение критерия: {z}')
```

Табличное значение критерия

Наконец, мы вычисляем табличное значение Z-критерия ($z_{кр}$), которое используется для сравнения с вычисленным Z-критерием:

- $z_{кр}$: Это критическое значение Z-критерия, которое зависит от заданного уровня значимости α. Если вычисленное значение Z-критерия превышает по абсолютной величине это табличное значение, то нулевая гипотеза отвергается.

$$z_t = \Phi^{-1}(1 - \frac{a}{2})$$

```
z_critical = st.norm.ppf(1 - a / 2)
print(f'Табличное значение критерия: {z_critical}')
```

Этот анализ позволяет определить, есть ли статистически значимые различия между вероятностями благоприятного исхода в двух сериях на основе выборочных данных и заданного уровня значимости.е