К вопросу аналитического вычисления мощности F-критерия однофакторного дисперсионного анализа.

Автор: **Егор Личак ПМ21-5**
Научный руководитель: **д.ф.-м.н, профессор, П. Е. Рябов**

Финансовый университет при Правительстве Российской Федерации

20 марта 2023 г.

Цели

1. Вывести аналитическое выражение для мощности F-критерия однофакторного дисперсионного анализа.
2. Разработать функционал для проведения однофакторного дисперсионного анализа в *Jupyter Notebook*.
3. Сформулировать предположение о состоятельности и несмещенности критерия.

Модель

Случайные выборки

$\overrightarrow{X_1}$	$\overrightarrow{X_2}$	...	$\overrightarrow{X_j}$	...	$\overrightarrow{X_k}$
1	2	...	j	...	k
X_{11}	X_{12}	...	X_{1j}	...	X_{1k}
X_{21}	X_{22}	...	X_{2j}	...	X_{2k}
$\vdots$	$\vdots$	$\vdots$	$\cdot$ X_{ij} $\cdot$	$\vdots$	$\vdots$
...	...	...	$X_{n_j j}$	...	$X_{n_j k}$
$\vdots$	$\vdots$	$\vdots$	$\cdot$ X_{ij} $\cdot$	$\vdots$	$\vdots$
...	...	...	...	...	$X_{n_k k}$

Гипотеза однородности:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k = \mu$$
$$H_1 : \exists i, j : \mu_i \neq \mu_j$$

Критерий

Обозначения

$SSE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2$ – внутригрупповая сумма квадратов;

$SSTR = \sum_{j=1}^{k} (\overline{X}_j - \overline{X})^2 \cdot n_j$ – межгрупповая сумма квадратов;

$\mathbb{F} = \frac{SSTR/(k-1)}{MSE/(n-k)}$ – статистика критерия с правосторонней критической областью

$$K_\alpha = \{x_{ij} : \mathbb{F} > f_\alpha(k-1, n-k)\}.$$

Доказательство

Лемма 1: Вне зависимости от верности гипотез H_0 или H_1 случайная величина $\frac{SSE}{\sigma^2} \sim \chi^2(n-k)$.
Лемма 2: Вне зависимости от верности гипотез H_0 или H_1 статистика $\frac{SSTR}{\sigma^2} \sim \chi'^2(l,\lambda)$ – ***нецентральное*** распределение хи-квадрат с l степенями свободы и параметром нецентральности λ.

Лемма 3: $SSTR = \sum\limits_{j=1}^{k} (\overline{X}_j - \mu_w)^2 \cdot n_j - n \cdot (\overline{X} - \mu_w)^2$,

где через $\mu_w = \frac{1}{n}\sum_j^k n_j \cdot \mu_j$ обозначено взвешенное среднее по всем выборкам.

Лемма 4: $E\left(\frac{SSTR}{\sigma^2}\right) = (k-1) + \frac{\sum\limits_{j=1}^{k}(\mu_j - \mu_w)^2 n_j}{\sigma^2}$.
Утверждение: $E[\chi'^2(k,\lambda)] = k + \lambda$.

Основные теоремы

Теорема 1: Если верна гипотеза H_1, то $\frac{SSTR}{\sigma^2} \sim \chi'^2(k-1, \lambda)$, где $\lambda = \frac{\sum\limits_{j=1}^{k}(\mu_j - \mu_w)^2 \cdot n_j}{\sigma^2}$,

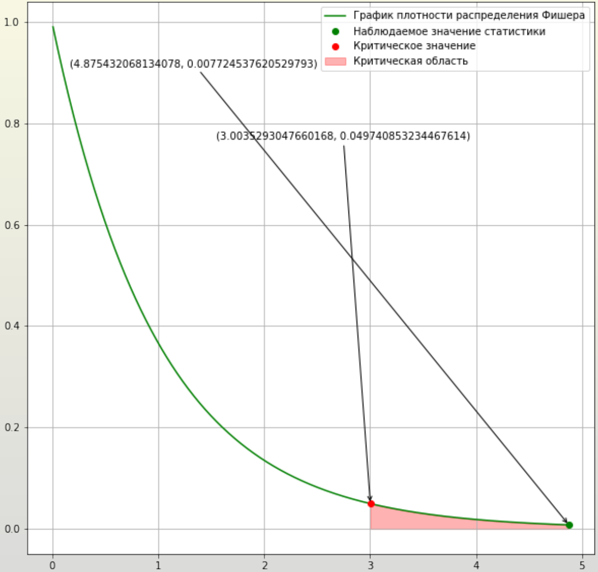
Аналитическое выражение для мощности критерия

Теорема 2: Если верна гипотеза H_1, то $\mathbb{F} = MSTR/MSE \sim F'(k-1, n-k, \lambda)$ – ***нецентральное распределение Фишера***, где $\lambda = \frac{\sum\limits_{j=1}^{k}(\mu_j - \mu_w)^2 \cdot n_j}{\sigma^2}$ – параметр нецентральности. При этом, ***мощность критерия*** определяется следующей формулой:

$$W = 1 - F'_{\mathbb{F}}(f_\alpha(k-1, n-k)).$$

Здесь через $F'_{\mathbb{F}}(\cdot)$ обозначена функция ***нецентрального распределения Фишера*** статистики критерия $\mathbb{F}$.

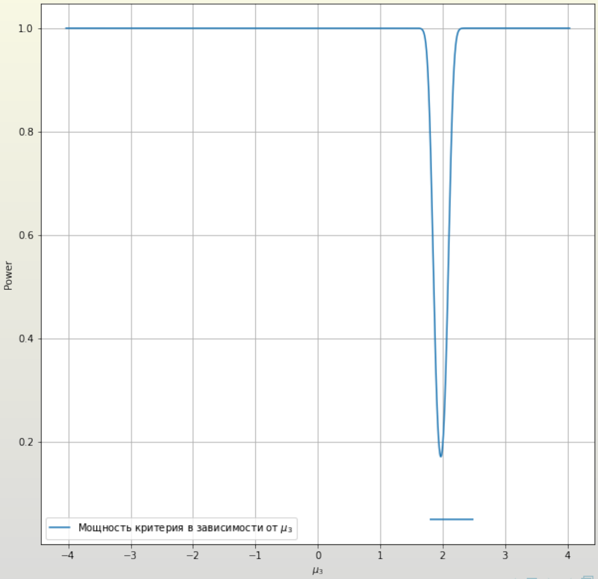
Имплементация на Python



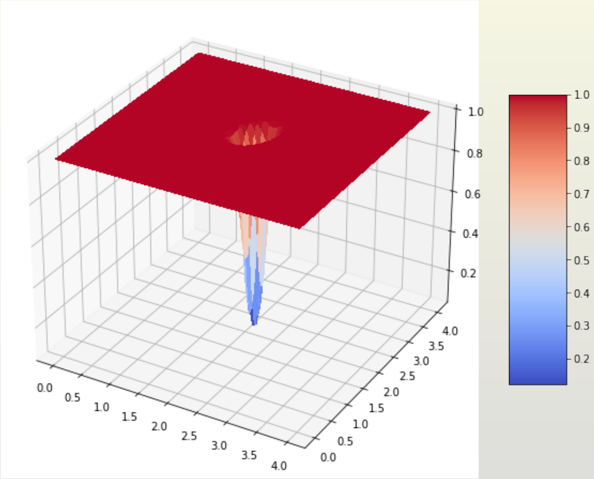
Имплементация на Python

К вопросу аналитического вычисления мощности F-критерия однофакторного дисперсионного анализа.

Егор Личак
ПМ21-5



Имплементация на Python



Список Литературы

1.*Кендалл М., Стьюарт А.* - "Том 2. Статистические выводы и связи" 1973.
2.*Г. Шеффе* "Дисперсионный анализ", 1980.
3.*Douglas C. Montgomery, George C. Runger* "Applied Statistics and Probability for Engineers", 2002
4.Интернет источник:
https://en.wikipedia.org/wiki/Noncentral_chi-squared_distribution
5.*П. Е. Рябов* "Лекции по теории вероятностей и математической статистике", 2022.

Спасибо за внимание