

Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 4_1

Критерии однородности Колмогорова-Смирнова и Вилконсона

Направление 01.04.02: «Прикладная математика и информатика»
Образовательная программа ВМ.5505.2021: «Математическое и информационное
обеспечение экономической деятельности»

Преподаватель:
доктор технических наук,
профессор Буре Владимир Мансурович

Санкт-Петербург
2022 г.

1 Постановка задачи

В таблице ниже приводятся сведения об экспорте и импорте Германии.

Исходные данные

1985: 184.71, 158.33
1986: 243.42, 191.28
1987: 294.55, 228.73
1988: 323.04, 280.92
1989: 341.88, 270.18
1990: 410.25, 346.72
1991: 403.18, 390.87
1992: 422.73, 402.15
1993: 382.31, 346.41
1994: 430.27, 385.14
1995: 524.51, 464.75
1996: 521.04, 456.83

Необходимо проверить однородность первых разностей исходных данных на уровне значимости 0.05 по критериям Колмогорова-Смирнова и Вилкоксона.

2 Ход работы

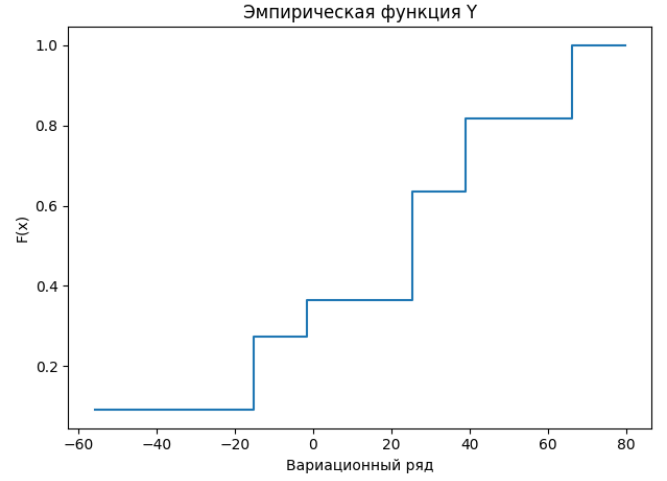
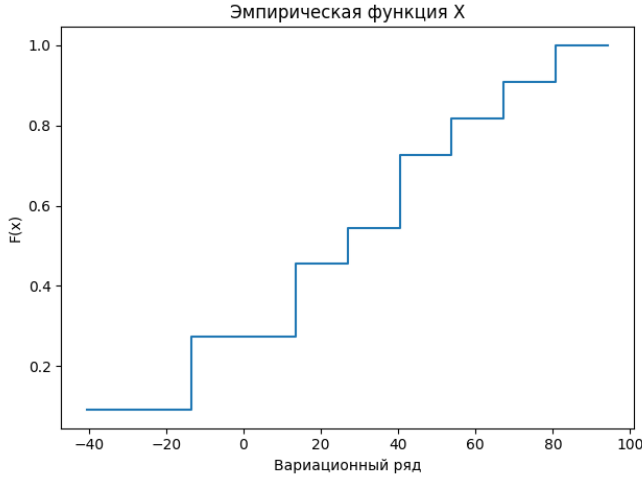
Изначально были изучены разделы 2.3.1 - 2.3.4 из книги Буре В.М., Парилина Е.М., Седаков А.А. «Методы прикладной статистики в R и Excel».

Была произведена обработка данных: вычислены первые разности, исключены повторения и выборки отсортированы.

Первые разности

58.71, 32.95
51.13, 37.45
28.49, 52.19
18.84, -10.74
68.37, 76.54
-7.07, 44.15
19.55, 11.28
-40.42, -55.74
47.96, 38.73
94.24, 79.61
-3.47, -7.92

Также, на основе полученных вариационных рядов построены эмпирические функции распределения.



Дальнейшие вычисления по критериям производились в соответствии с алгоритмами, описывающие критерии.

2.1 Критерий Колмогорова-Смирнова

Алгоритм критерия однородности двух выборок Колмогорова-Смирнова

1. Выдвинуть нулевую гипотезу $H_0 : F(\cdot) = G(\cdot)$. Сформировать альтернативную гипотезу $H_1 : \sup_{|x|<\infty} |F(x) - G(x)| > 0$
2. Задать уровень значимости критерия α
3. Вычислить значения статистики

$$\sqrt{\frac{mn}{m+n}} D_{m,n}, \quad (1)$$

где n и m - объёмы выборок X и Y соответственно, следующим образом:

- (a) По выборкам $X_{[n]}$ и $Y_{[m]}$ построить эмпирические функции распределения $F_n(x)$ и $G_m(x)$ по формуле

$$F_n(x) = \begin{cases} 0, & x < x_1, \\ \frac{1}{n}, & x_1 \leq x < x_2, \\ \vdots & \\ \frac{k}{n}, & x_k \leq x < x_{k+1}, \\ \vdots & \\ 1, & x \geq x_n \end{cases} \quad (2)$$

$$G_m(x) = \begin{cases} 0, & x < y_1, \\ \frac{1}{m}, & y_1 \leq x < y_2, \\ \vdots & \\ \frac{k}{m}, & y_k \leq x < y_{k+1}, \\ \vdots & \\ 1, & x \geq y_m \end{cases} \quad (3)$$

- (b) Вычислить $D_{m,n}$ по следующим формулам:

$$D_{m,n}^+ = \max_{1 \leq r \leq m} \left[\frac{r}{m} - F_n(y_r) \right] = \max_{1 \leq s \leq n} \left[G_m(x_s) - \frac{s-1}{n} \right] \quad (4)$$

$$D_{m,n}^- = \max_{1 \leq r \leq m} \left[F_n(y_r) - \frac{r-1}{m} \right] = \max_{1 \leq s \leq n} \left[\frac{s}{n} - G_m(x_s) \right] \quad (5)$$

$$D_{m,n} = \max(D_{m,n}^+, D_{m,n}^-) \quad (6)$$

(с) Найти значение критерия в виде формулы (1)

4. Найти критическую область – интервал $(k_{1-\alpha}; \infty)$, где k - квантиль уровня $1 - \alpha$ распределения Колмогорова.
5. Если численное значение статистики критерия, вычисленного по формуле (1), попадает в интервал $(k_{1-\alpha}; \infty)$, то нулевая гипотеза H_0 отвергается, в противном случае нет оснований отвергнуть её при уровне значимости α .

2.2 Критерий Вилкоксона

Алгоритм критерия Вилкоксона

1. Выдвинуть нулевую гипотезу $H_0 : F(\cdot) = G(\cdot)$. Сформулировать альтернативную гипотезу $H_1 : F(\cdot) \neq G(\cdot)$ (двусторонняя альтернатива)
2. Объединить выборки $X_{[n]}$ и $Y_{[m]}$ в общий вариационный ряд и проранжировать. Пусть ранги элементов выборки $Y_{[m]} = \{y_1, \dots, y_m\}$ обозначаются через r_1, \dots, r_m
3. Вычислить статистику Вилкоксона по следующей формуле:

$$W(X_{[n]}, Y_{[m]}) = r_1 + \dots + r_m \quad (7)$$

4. Задать уровень значимости α
5. Найти критическую область критерия:

$$\left[\frac{m(m+1)}{2}, w_{\frac{\alpha}{2}, m, n} \right] \cup \left[m(n+m+1) - w_{\frac{\alpha}{2}, m, n}, mn + \frac{m(m+1)}{2} \right], \quad (8)$$

где $w_{\frac{\alpha}{2}, m, n}$ – квантиль статистики Вилкоксона.

6. Если численное значение статистики $W(X_{[n]}, Y_{[m]})$ попадает в критическую область, то нулевая гипотеза H_0 отвергается, в противном случае нет оснований её отвергнуть при уровне значимости α .

Результаты работы обоих критериев приведены ниже.

Критерий Колмогорова-Смирнова

Гипотеза H_0 принимается на уровне значимости 0.05, так как значение 0.4264 не принадлежит интервалу (1.36; +inf)

Критерий Вилкоксона

Гипотеза H_0 принимается на уровне значимости 0.05, так как значение $w = 123.0$ не принадлежит критической области [66.0; 96.0] \cup [157.0; 187.0]