Лабораторная работа 5.

Имеются две выборки X1 и X2 (функции распределения генеральных совокупностей непрерывны). Проверьте гипотезу об однородности выборок с помощью критерия Колмогорова-Смирнова.

Критерий однородности Колмогорова-Смирнова

Проверяем гипотезу о равенстве функций распределения $H_0 : F(\cdot) = G(\cdot)$ при альтернативной гипотезе $H_1 : F(\cdot) \neq G(\cdot)$, функции непрерывны, $m \leq n$

Статистика Смирнова определяется следующей формулой $D_{m,n} = \sup\limits_{|x|<\infty} \left| F_n(x) - G_m(x) \right|$

На практике значение статистики рекомендуется вычислять по формулам:

$$D^+{}_{m,n} = \max_{1\leq r\leq m} \left[\frac{r}{m} - F_n\left(y_{(r)}\right) \right] = \max_{1\leq s\leq n} \left[G_m\left(x_{(s)}\right) - \frac{s-1}{n} \right] \quad (1)$$

$$D^-{}_{m,n} = \max_{1\leq r\leq m} \left[F_n\left(y_{(r)}\right) - \frac{r-1}{m} \right] = \max_{1\leq s\leq n} \left[\frac{s}{n} - G_m\left(x_{(s)}\right) \right] \quad (2)$$

$$D_{m,n} = \max\left[D^+{}_{m,n}, D^-{}_{m,n} \right] \quad (3)$$

При справедливости нулевой гипотезы и неограниченном увеличении объемов выборок статистика

$$\sqrt{\frac{mn}{m+n}} D_{m,n} \qquad (4)$$

асимптотически подчиняется распределению Колмогорова.

Алгоритм:

Строим эмпирические функции распределения $F_n(x), G_m(x)$

Вычисляем статистики по формулам $(1),(2),(3)$

Находим значение исправленной статистики (4)

Находим критическую область – интервал $\left(k_{1-\alpha}, \infty\right)$

Если значение статистики попадает в критическую область, нулевая гипотеза отвергается.

Ранговый критерий однородности Вилкоксона

Проверяем гипотезу о равенстве непрерывных функций распределения $H_0 : F(\cdot) = G(\cdot)$
Возможны альтернативные гипотезы:

$$H_1 : F(\cdot) \geq G(\cdot)$$
$$H_1 : F(\cdot) \leq G(\cdot)$$
$$H_1 : F(\cdot) \neq G(\cdot)$$

Алгоритм

1.Объединим выборки $X_{[n]}, Y_{[m]}$ в общий вариационный ряд.

Припишем каждому элементу ранг, т.е. номер в вариационном ряду.
Обозначим ранги элементов выборки $Y_{[m]}$ через $r_1, \dots r_m$.

Вычислим статистику Вилкоксона по следующей формуле:

$$W\left(X_{[n]}, Y_{[m]}\right) = r_1 + \ldots + r_m$$

3. Задаем уровень значимости α.
4. Найдем критическую область для критерия:

- При альтернативе $H_1 : F(\cdot) \leq G(\cdot)$, критическая область
$$\left[0, w_{\alpha,m,n}\right],$$
$w_{\alpha,m,n}$ - квантиль предельного распределения статистики Вилкоксона.
- При альтернативе $H_1 : F(\cdot) \geq G(\cdot)$, критическая область
$$\left[m(n+m+1) - w_{\alpha,m,n}, \infty\right)$$
- При альтернативе $H_1 : F(\cdot) \neq G(\cdot)$, критическая область
$$\left[0, w_{\frac{\alpha}{2},m,n}\right] \cup \left[m(n+m+1) - w_{\frac{\alpha}{2},m,n}, \infty\right)$$

Для больших выборок можно использовать следующую аппроксимацию

$$W^*\left(X_{[n]}, Y_{[m]}\right) = \frac{(r_1 + \ldots + r_m) - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}$$

- При альтернативе $H_1 : F(\cdot) \leq G(\cdot)$, критическая область $\left(-\infty, u_\alpha\right]$, квантиль нормального стандартного распределения.
- При альтернативе $H_1 : F(\cdot) \geq G(\cdot)$, критическая область $\left[u_{1-\alpha}, \infty\right)$
- При альтернативе $H_1 : F(\cdot) \neq G(\cdot)$, критическая область
$$\left(-\infty, u_{\frac{\alpha}{2}}\right] \cup \left[u_{1-\frac{\alpha}{2}}, \infty\right)$$

Ранговый критерий однородности Манна-Уитни

Статистика Манна-Уитни это число успехов в парных сравнениях.

$$I\left\{x_i < y_j\right\} = \begin{cases} 1, x_i < y_j \\ 0, x_i \geq y_j \end{cases}$$

Статистика вычисляется по формуле:

$$U\left(X_{[n]}, Y_{[m]}\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} I\left\{x_i < y_j\right\}$$

Статистика Манна-Уитни и Вилкоксона связаны следующим образом:

$$U_Y\left(X_{[n]}, Y_{[m]}\right) = mn + \frac{m(m+1)}{2} - W\left(X_{[n]}, Y_{[m]}\right)$$

$$U_X\left(X_{[n]}, Y_{[m]}\right) = mn + \frac{n(n+1)}{2} - W\left(Y_{[m]}, X_{[n]}\right)$$

Статистика Манна-Уитни может быть вычислена следующим образом:

$$U\left(X_{[n]}, Y_{[m]}\right) = \min\left\{U_Y\left(X_{[n]}, Y_{[m]}\right), U_X\left(X_{[n]}, Y_{[m]}\right)\right\}$$

При $m, n > 8$ можно применить аппроксимацию

$$U^*\left(X_{[n]}, Y_{[m]}\right) = \frac{U\left(X_{[n]}, Y_{[m]}\right) - mn/2}{\sqrt{mn(m+n+1)/12}}$$

статистика подчиняется стандартному нормальному закону.