Дисперсионный анализ.

1 Пусть ξ_{k_1,k_2} – случайная величина, имеющая распределение Фишера F_{k_1,k_2}. Доказать, что $\xi_{k_1,k_2} \xrightarrow{d} 1$ при $k_1, k_2 \to \infty$.

2 Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ – выборки из функций распределения $F(x)$ и $G(x) = F(x - \theta)$ соответственно. Рассмотрим основную гипотезу $H_0 : \theta = 0$ и альтернативу $H_1 : \theta \neq 0$. Доказать, что критерий Уилкоксона является состоятельным критерием проверки H_0 против H_1.

3 Выданы выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$. Определив, являются они парными или независимыми, нормальными или произвольными, проверить гипотезу об их однородности (если они признаны не гауссовскими – то гипотезу об отсутствии сдвига) с помощью статистической процедуры, контролирующей FWER на уровне 0.1.

4 Выданы выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$. Определив, являются ли они парными или независимыми, нормальными или произвольными, проверить гипотезу об их однородности с помощью статистической процедуры, контролирующей FDR на уровне 0.1.

5 Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ – две независимые выборки из распределения Стьюдента t_{10}. Рассмотрим критерий Стьюдента $\{|T| > u_{1-\alpha/2}\}$, где $u_{1-\alpha/2}$ – $(1-\alpha/2)$-квантиль из распределения t_{2n-2}. Можно ли пользоваться критерием Стьюдента для проверки гипотезы об однородности данных выборок? С помощью моделирования определить, как ведёт себя уровень значимости данного критерия при $n \to \infty$.

6 Используя то, что при верности гипотезы $H_0 : F = G$ статистика критерия Смирнова не зависит от F, и предполагая, что размер выборок равный и равен n, найти такое минимальное n, что распределение статистики критерия Смирнова неотличимо от распределения Колмогорова на уровне значимости $\alpha = 0.05$ (количество выборок для каждого n взять равным 100). Т.е. фактически нужно выяснить минимальное n, при котором можно пользоваться приближением в теореме Смирнова.

Замечание. Возможно, вам пригодится явный вид статистики критерия Смирнова:

$$D_{n,m} = \max\left\{ \max_{i=1,\ldots,n}\{\frac{i}{n} - G_m^*(X_i)\}, \max_{i=1,\ldots,n}\{\frac{j}{m} - F_n^*(Y_j)\} \right\},$$

где выборки $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ из функций распределения F и G соответственно, а F_n^* и G_m^* – их эмпирические функции распределения.