

Критерий Колмогорова-Смирнова (Kolmogorov–Smirnov test)

Синонимы: K–S test, KS test

Разделы: [Метрики](#)

В [математической статистике](#) критерий Колмогорова-Смирнова — непараметрический критерий равенства непрерывных одномерных вероятностных распределений, используемый для сравнения эмпирического выборочного распределения с теоретическим (одномерной критерий), или для сравнения двух [выборок](#) (двумерный критерий). Критерий назван в честь советских математиков [Андрея Колмогорова](#) и [Николая Смирнова](#)

Часто Критерий Колмогорова-Смирнова рассматривают как два отдельных критерия.

1. [Критерий согласия](#) Колмогорова — служит для проверки [гипотезы](#) о принадлежности значений выборки к определенному теоретическому закону распределения.
2. Критерий однородности Смирнова — применяется для проверки гипотезы о принадлежности значений двух независимых выборок к одному и тому же закону распределения.

В основе критерия лежит **статистика Колмогорова-Смирнова**, которая является оценкой расстояния между эмпирической выборочной функцией распределения и кумулятивной функцией теоретического распределения, либо между эмпирическими функциями распределения двух выборок. Нулевое распределение данной статистики определяется при [нулевой гипотезе](#), что выборка извлечена из теоретического распределения (в одномерном случае), или две выборки извлечены из одного и того же распределения (в двумерном случае).

В случае одномерной выборки распределение, рассматриваемое при нулевой гипотезе, может быть [непрерывным](#), полностью [дискретным](#) или смешанным. В случае двух выборок распределение, рассматриваемое в рамках нулевой гипотезы, является непрерывным, но в остальном оно не ограничено.

Тест Колмогорова-Смирнова с двумя выборками является одним из наиболее полезных и общих непараметрических методов сравнения двух выборок, поскольку он чувствителен к различиям как в положении, так и в форме эмпирических кумулятивных функций распределения двух выборок.

Критерий Колмогорова-Смирнова может применяться в практике анализа данных для различных задач. Например, в очистке данных при заполнении пропусков, с помощью критерия может быть оценен теоретический закон распределения, которому подчиняются значения выборки, после чего пропуски могут быть заполнены значениями из этого же распределения.

Если при анализе требуется сформировать несколько выборок из одного и того же набора данных (например, обучающую и тестовую), то с помощью критерия Колмогорова-Смирнова будет полезно убедиться, что они подчиняются одному закону распределения и могут быть корректно использованы для построения модели.