

САМОСТОЯТЕЛЬНАЯ РАБОТА N 3

Проверка непараметрических статистических гипотез. Критерий согласия Колмогорова.

1. Получите выборку объемом 100 значений из экспоненциально распределенной генеральной совокупности с параметром $\lambda = a/b$.
2. Проверьте простую гипотезу согласия по критерию Колмогорова для смоделированных выборок.

$$H_0: F(x) = 1 - e^{-\lambda x}, x \geq 0$$

Альтернативная гипотеза H_1 заключается в том, что нулевая гипотеза H_0 неверна. В силу теоремы Гливленко-Кантелли эмпирическая функция распределения представляет собой состоятельную оценку теоретической функции распределения, то есть при условии истинности гипотезы H_0 и непрерывности функции распределения $F(x)$ справедливо

$$|F_n(x) - F(x)| \rightarrow 0, n \rightarrow \infty.$$

Статистика Колмогорова имеет вид $D_n = \max_x |F_n(x) - F(x)|$ (расстояние между эмпирическим и теоретическим распределением в равномерной метрике). Если нулевая гипотеза верна и функция распределения $F(x)$ непрерывна, то закон распределения статистики Колмогорова оказывается одним и тем же для всех непрерывных функций распределения, он зависит только от объема выборки.

Для малых n для статистики D_n при справедливости гипотезы H_0 составлены таблицы процентных точек. Для больших n распределение статистики D_n при справедливости гипотезы H_0 указывает теорема Колмогорова.

Теорема.

При справедливости гипотезы H_0 и непрерывности функции распределения $F(x)$ справедливо

$$P(\sqrt{n} D_n < z) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 z^2}$$

Алгоритм проверки нулевой гипотезы по критерию Колмогорова выглядит следующим образом. По исходной выборке надо вычислить значение статистики D_n . Для этого можно воспользоваться формулой

$$\max_k \left[\frac{k}{n} - F(x_{(k)}), F(x_{(k)}) - \frac{k-1}{n} \right],$$

здесь $x_{(1)}, \dots, x_{(n)}$ — элементы вариационного ряда.

Полученную величину D_n сравнивают с критическим значением, взятым из статистической таблицы, для выбранного уровня значимости 0.95. Нулевую гипотезу отвергают, если значение статистики превосходит выбранное критическое значение, в противном случае нулевую гипотезу принимают.

3. Смоделируйте 15 независимых выборок из равномерного распределения с постоянным математическим ожиданием a и постоянной дисперсией b , по 40 наблюдений в каждой выборке. Проверьте их соответствие нормальному распределению с мат. ожиданием a и дисперсией b по критерию Колмогорова.

4. Вычислите выборочное среднее по каждой из 15 выборок. Для получившейся совокупности значений проверьте гипотезу о том, что эта совокупность представляет собой выборку из нормального распределения со средним a и дисперсией $b/40$ на уровне значимости 0,05. Дайте теоретическое объяснение результатам п. 3 и 4.