Галанов Ю.И.

Лабораторный практикум по мат. статистике

Проверка гипотезы о нормальности распределения. Критерий Колмогорова

Задание.

- **Смоделировать две выборки одинакового объема: одну с нормальным распределением, а другую с равномерным распределением. Мат. ожидание и дисперсию для обеих выборок задать одинаковыми.**
- **Провести визуальную проверку гипотезы о нормальности распределений. Для чего представить стандартизованные значения вариант распределений как функции соответствующих квантилей стандартного распределения.Если распределение нормально, то должна получиться прямая линия. В противном случае экспериментальные точки не будут укладываться на прямую линию.** ***Визуальный метод проверки нормальности основан на способности человека отличать прямую линию от непрямой.***
- **Проанализируйте, как влияет объем выборок на на возможность их визуального распознавания.**
- **Проверьте гипотезу о нормальности выборок с помощью критерия Колмогорова. Для этого задайте и подсчитайте значения статистик критерия для обеих выборок. Критические точки возьмите из прилагаемых таблиц.**
- **Проверьте экспериментально, как влияет объем выборки и значение уровня значимости на возможность идентифицировать нормальную выборку.**
- **При составлении отчета опишите пять шагов проверки гипотезы применительно к Вашей задаче.**

Контрольные вопросы

- **Что называют статистической гипотезой?**
- **Какая гипотеза является простой?**
- **Какие ошибки возможны при проверке гипотезы?**
- **Что такое уровень значимости и каков диапазон его значений?**
- **Что такое статистический критерий и статистика критерия?**
- **Что такое мощность критерия?**
- **Что такое критическая область. Какими бывают критические области?**
- **Какова критическая область для данной задачи?**
- **От чего зависит надежность статистических выводов?**

Примерная программа проведения работы

Моделируем две выбоки с различными распределениями, но одинаковыми параметрами

Нормальное

$N := 500 \qquad i := 0 .. N - 1$

$m := \frac{a+b}{2} \qquad m = 3$

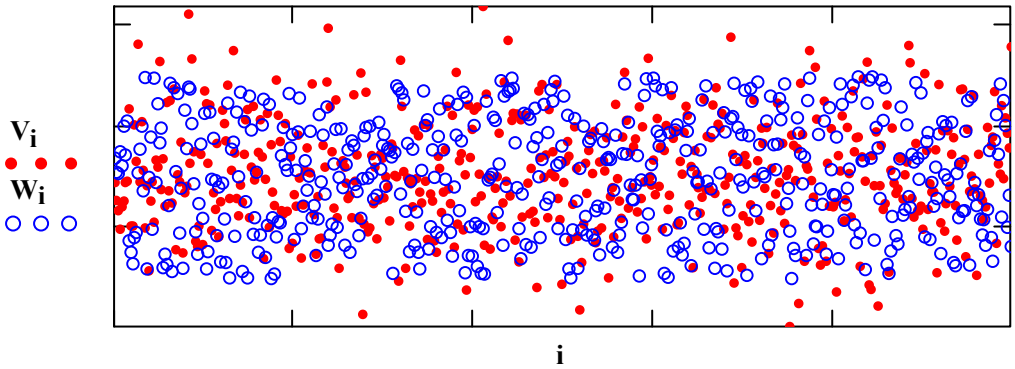
$V_i := m + s \cdot \cos(2 \cdot \pi \cdot rnd(1)) \cdot \sqrt{-2 \cdot \ln(rnd(1))}$

Равномерное

$a := 1 \qquad b := 5$

$s := \frac{b-a}{2 \cdot \sqrt{3}} \qquad s = 1.154701$

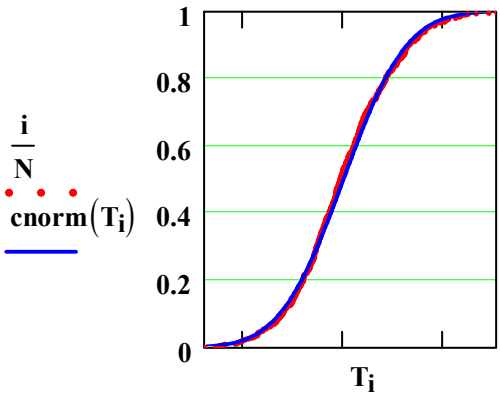
$W_i := a + rnd(b - a)$

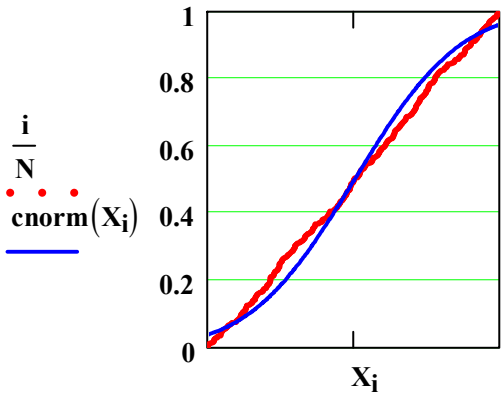


Преобразуем обе выборки в стандартные, предполагая, что справедлива гипотеза о нормальности выборок

$Z := \left(\frac{V - mean(V)}{stdev(V)}\right) \qquad T := sort(Z) \qquad Y := \left(\frac{W - mean(W)}{stdev(W)}\right) \qquad X := sort(Y)$

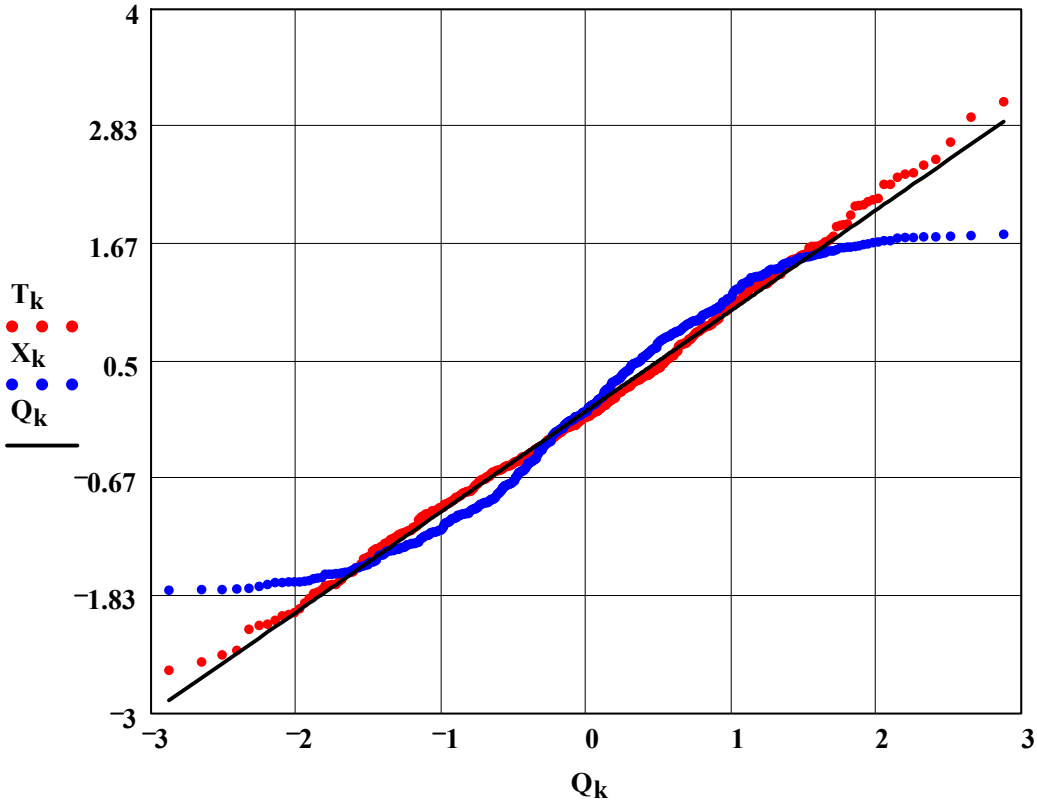
Построим выборочные функции распределения, используя вариационный ряд и сравним их с теоретическим (cnorm(x))





Визуальная проверка нормальности

$k := 1 .. N - 1$ $\qquad F_k := \frac{k}{N}$ $\qquad Q_k := qnorm(F_k, 0, 1)$



Подсчитаем значения статистик критерия Колмогорова

Нормальное

$TT_i := \left| \frac{i}{N} - cnorm(T_i) \right|$

$tT := \sqrt{N} \cdot max(TT)$

$tT = 0.709174$

Равномерное

$TX_i := \left| \frac{i}{N} - cnorm(X_i) \right|$

$tX := \sqrt{N} \cdot max(TX)$

$tX = 1.578371$

Делаем выводы о принадлежности или непринадлежности выборок к нормальному распределению

Таблица квантилей распределения Колмогорова

Уровень значимости	критическая точка
0.05	1.358
0.01	1.6276

$j := 0 .. 1001$

$$K(t) := \sum_{i=-2001}^{2001} \left[\left[(-1)^{-i} \right] \cdot e^{-\left(2 \cdot i^2 \cdot t^2\right)} \right]$$

$K(1.358) = 0.949973$ $\qquad$ $t := 0, 0.1 .. 2$

Расчет квантилей распределения Колмогорова

$qK(t, alfa) := root(K(t) - alfa, t)$ $\qquad$ $qK(1, 0.95) = 1.354618$

$qK(1, 0.9) = 1.223734$ $\qquad$ $qK(1, 0.95) = 1.354618$ $\qquad$ $qK(1, 0.99) = 1.621647$

$qK(1, 0.995) = 1.704622$ $\qquad$ $qK(1, 0.999) = 1.871257$ $\qquad$ $qK(1, 0.9995) = 1.898543$