

Домашнее задание 2, Новиков Герман, 277

Все вычисления вместе с кодом и комментариями находятся в файле на [github](#)

Задача 1. (Задача номер 9). Цифры $0, 1, 2, \dots, 9$ среди 800 первых десятичных знаков числа π появляются 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно.

Проверить гипотезу о согласии данных с законом равномерного распределения.

Решение: Для проверки гипотезы о равномерности распределении данных воспользуемся критерием χ^2 :

Получаемое значение T_{χ^2} отвечает нашему предположению для $\alpha = 0.05$, соответственно, принимаем гипотезу о равномерном распределении.

Задача 2. $data = [50numbers]$

Используя Пирсоновский хи-квадрат тест, проверить гипотезу о том, что эти числа распределены согласно равномерному распределению на отрезке $[0, 1]$.

Решение: Аналогично предыдущей задаче, однако в этот раз воспользуемся p – value.

Полученное значение $p_{value} = 0.23$ показывает, что с достаточной вероятностью данные распределены равномерно на $[0, 1]$, что подтверждает нашу гипотезу.

Задача 3. (Задача номер 10) При эпидемии гриппа из 200 контролируемых людей однократное заболевание наблюдалось у 181 человека, а дважды болели гриппом 9 человек. Правдоподобна ли гипотеза о том, что в течение эпидемии гриппа число заболеваний отдельного человека представляет собой случайную величину, подчиняющуюся биномиальному распределению с числом испытаний $n = 2$?

Решение: Изначально, мы не знаем распределение вероятностей (то есть не знаем параметр p), строим правдоподобие для него и находим предполагаемое значение. Далее используем критерий χ^2 .

Полученное $p_{value} = 3.18e^{-29}$ есть почти нулевая величина, что означает, что мы получили очень редкое событие и, в соответствии с этим, мы отвергаем нашу гипотезу о нормальном распределении.

Задача 4. Пусть некоторая статистика, построенная по простой выборке, в условиях истинности гипотезы H_0 имеет абсолютно непрерывное

распределение. Рассматривая $p - value$ как случайную величину, найти ее распределение.

Решение: В предположении, что нулевая гипотеза H_0 справедлива, рассматриваемая нами статистика $T(X)$ имеет распределение $F(t)$. Рассматриваем распределение $p - value$ (учитывая, что статистика имеет абсолютно непрерывное распределение), фактически - $p - value$ - случайная величина: $\xi = 1 - F(T(X))$, но тогда:

$$P(\xi < t) = P(1 - F(T(X)) < t) = 1 - P(F(T(X)) > 1 - t) =$$

$$1 - P(T(X) > F^{-1}(1 - t)) = 1 - F(F^{-1}(1 - t)) = 1 - (1 - t) = t$$

То есть мы получаем, что $p - value$ для любого распределения есть равномерно распределенная на $[0, 1]$ случайная величина.

Задача 5. (Задача номер 12). При снятии показаний измерительного прибора десятые доли деления шкалы прибора оцениваются "на глаз" наблюдателем. Количества цифр 0, 1, 2, ..., 9, записанных наблюдателем в качестве десятых долей 1 при 100 независимых измерениях, равны 5, 8, 6, 12, 14, 18, 11, 6, 13, 7 соответственно.

Проверить гипотезы о согласии данных с законом равномерного распределения и с законом нормального распределения. Для ответа на вопрос можно сравнить значения $p - value$ для обеих гипотез.

Решение: С семинара и в Ивченко-Медведеве выведены формулы для $\hat{\theta}_1$ и $\hat{\theta}_2$, остается посчитать их численно и подставить в T_{chi^2} , для которого у нас будет $N - 1 - 2$ степеней свободы в силу независимости θ_1 и θ_2 .