Домашнее задание 3, Новиков Герман, 277

Все вычисления вместе с кодом и комментариями находятся в файле на **github**

Задача 1. 8-го января 2003 года в New York Times были сообщены следующие данные из штата Мэриленд: в случае, если происходило убийство афроамериканца, было вынесено 14 смертных приговоров для преступника, а в 641 случае смертных приговор не последовало. В случае, если происходило убийство белого, то в 62 случаях был вынесен смертный приговор и в 594 случаях не был. Проанализируйте эти данные, используя статистические техники, и интерпретируйте результаты.

Решение: Проверим, является ли значимым цвет кожи убитого, то есть посмотрим - есть ли зависимость между количеством осужденных за убийство человека того или иного цвета кожи от самого цвета кожи убитого. Выдвинем гипотезу, о том что зависимости нет и воспользуемся критерием χ^2:

Ответ: Получаемое значение $p_{value} < 10^{-8}$, соотвественно отвегаем гипотезу о независимости.

Задача 2. (Задача номер 59). Построить критерий для проверки гипотезы $H_1 : p = \frac{1}{2}$ при альтернативной гипотезе $H_2 : p \neq \frac{1}{2}$ по результатам восьми испытаний, подчиняющихся схеме Бернулли. Вероятность ошибки первого рода α положить равной 0,05.

Решение: ~~Предноложим, что верная гипотеза $H_1 : p = \frac{1}{2}$. В этом предноложении случайная величина~~

$$\cancel{\theta = 2\sqrt{n}(\frac{n-k}{k} - \frac{1}{2}),}$$

~~где $n = 8$, k — количество 1, имеет в силу ЦПТ распределение к близкое к $N(0, 1)$ (сходимость по вероятности).И, таким образом, можно установить пару квантилей, соответствующих $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$ (так как распределение $N(0, 1)$ является известным) и принимать или отвергать гипотезу в соответствии с ними.~~

Задача 3. (Задача номер 3) Пусть $X_1, ..., X_n$ — простая выборка, полученная из абсолютно непрерывного распределения с плотностью f. Найти:

- Функцию плотности совместного распределения вариационного ряда $X_{(1)}, ..., X_{(n)}$.

- Совместное распределение $X_{(1)}$ и $X_{(n)}$ при условии, что X_i имеет равномерное распределение на отрезке $[a, b]$. Вычислить также их математические ожидания, дисперсии и корреляцию.

Решение:

- Функцию плотности рассмотрим для $n = 2$, для $n \geqslant 2$ ее построение ничем не отличается: Пусть $x_1 < x_2$ и будем рассматривать интервалы, по которым могут распределиться все наши элементы ряда: ровно $i_1 - 1$ должны попасть в интервал до x_1, один из элементов в $[x_1, x_1 + dx]$, дальше ещё $i_2 - 1 - i_1$ элемент ровно между $x_1 + dx$ и x_2, один элемент в $[x_2, x_2 + dx]$ и все оставшиеся дальше $x_2 + dx$. Таким образом, получим вероятность такого события:

$$C_n^{i_1-1}F^{i_1-1}(x_1)(n-i_1+1)f(x_1)dxC_{n-i_1}^{i_2-1-i_1}(F(x_2)- \\ F(x_1+dx))^{i_2-1-i_1}(n-i_2+1)f(x_2)dx(1-F(x_2))^{n-i_2}$$

  Просто берем предел при $dx \to 0$ и получаем функцию распределения для $X(i_1)$ и $X_{(i_2)}$. Дальше аналогично для произвольного n.

- Теперь рассмотрим совместное распределение для двух элементов $i_1 = 1, i_2 = n$, по формуле, выведенной выше и при условии X_i - равномерна на $[a, b]$

$$C_n^0F^0(x_1)(n)f(x_1)C_{n-1}^{n-2}(F(x_2)- \\ F(x_1))^{n-2}(n)f(x_2)dx(1-F(x_2))^0$$

  Где F и f - ФР и ФП для равномерного на $[a, b]$

  Математическое ожидание для

Задача 4.

Задача 5. (Критерий инверсий)

Решение:

- Введем случайную величину n_i, которая будет показывать количество инверсий для X_i, то есть число стоящих левее элементов таких, что $X_i < X_j$. Тогда получаем, что случайные величины $n_1, ...n_{n-1}$ независимы.

  Найдем производящую функцию для n_i:

  Т.к. она может с равными вероятностями $\frac{1}{n-i+1}$ принимать значения $0, ..., n-i$, то её производящая функция:

$$\phi_i(z) = \sum_{r=0}^{n-i} \mathsf{P}\, n_i = rz^r = \frac{1}{n-i+1}(1 + z + ... + z^{n-i})$$

  Тогда для статистики T_n, являющейся количесвтом всех инверсий: $T_n = \sum n_i$ получем:

  $\mathsf{E}T_n = \sum\limits_{i=1}^{n-1} \mathsf{E}n_i = \frac{n(n-1)}{4}$

  $DT_n = \sum\limits_{i=1}^{n-1} Dn_i = \frac{n(n-1)(2n+5}{72}$

Задача 6. Проанализируйте данные о возрасте и доходах по ссылке: http://lib.stat.cmu.edu/DASL/Datafiles/montanadat.html

Решение: Воспользуемся аналогично задаче 1 гипотезой независимости и используем критерий χ^2: