Математическая статистика для психологов

Пример решения: проверка гипотезы о законе распределения

В исследовании порогов социального атома студентов – психологов просили определить, с какой частотой встречаются в записной книжке их мобильного телефона мужские и женские имена. Определите, отличается ли распределение, полученное по Вашей записной книжке, от равномерного распределения.

Решение.

Количество женских имен – 20, мужских – 25.

Для того чтобы проверить гипотезу о равномерном распределении X, т.е. по закону: f(x) = 1/(b-a) в интервале (a,b) надо:

Оценить параметры a и b - концы интервала, в котором наблюдались возможные значения X, по формулам (через знак * обозначены оценки параметров):

$a^* = \overline{x} - \sqrt{3}\sigma$, $b^* = \overline{x} + \sqrt{3}\sigma$

2. Найти плотность вероятности предполагаемого распределения $f(x) = 1/(b^* - a^*)$

3. Найти теоретические частоты:

$n_1 = nP_1 = n[f(x)*(x_1 - a^*)] = n*1/(b^* - a^*)*(x_1 - a^*)$

$n_2 = n_3 = ... = n_{s-1} = n*1/(b^* - a^*)*(x_i - x_{i-1})$

$n_s = n*1/(b^* - a^*)*(b^* - x_{s-1})$

4. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы k = s-3, где s - число первоначальных интервалов выборки; если же было произведено объединение малочисленных частот, следовательно, и самих интервалов, то s - число интервалов, оставшихся после объединения.

1. Найдем оценки параметров a^* и b^* равномерного распределения по формулам:

$a^* = \overline{x} - \sqrt{3}\sigma$, $b^* = \overline{x} + \sqrt{3}\sigma$

$a^* = 0,56 - \sqrt{3} \cdot 0,5 = -0,31$, $b^* = 0,56 + \sqrt{3} \cdot 0,5 = 1,42$.

2. Найдем плотность предполагаемого равномерного распределения:

$f(x) = 1/(b^* - a^*) = 1/(1,42 - (-0,31)) = 0,581$.

3. Найдем теоретические частоты:

$n_1 = n \cdot f(x)(x_1 - a^*) = 45 \cdot 0,581(0-(-0,31)) = 7,98$.

$n_2 = n \cdot f(x)(b^* - x_1) = 45 \cdot 0,581(1,42-1) = 10,88$.

Остальные n_s будут равны:

$n_s = n \cdot f(x)(x_i - x_{i-1})$

i	n_i	n^*_i	$n_i - n^*_i$	$(n_i - n^*_i)^2$	$(n_i - n^*_i)^2/n^*_i$
1	20	7,98	12,02	144,57	18,12
2	25	10,88	14,12	199,34	18,32
Итого	45				36,45

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение $K_{набл}$, тем сильнее довод против основной гипотезы.

Поэтому критическая область для этой статистики всегда правосторонняя: $[K_{kp};+\infty)$.

Её границу $K_{kp} = \chi^2(k\text{-}r\text{-}1;\alpha)$ находим по таблицам распределения χ^2 и заданным значениям s, k (число интервалов), r=2 (параметры a и b).

Kkp = 233,9943; Kнабл = 36,45.

Наблюдаемое значение статистики Пирсона не попадает в критическую область: Kнабл < Kkp, поэтому нет оснований отвергать основную гипотезу. Справедливо предположение о том, что данные выборки имеют равномерный закон.