

Пример 2.5. Требуется определить, одинаковы ли функции распределения листьев по размерам у двух ярусов (частот n по выборкам X и Y) кроны акации. Данные эксперимента приведены в таблице справа. Использовать уровень значимости $\alpha = 0,05$.

Результаты эксперимента

диапазон размеров		эксперимент	
		$n(X)$	$n(Y)$
8,2	9,8	1	0
9,8	11,4	5	7
11,4	13,0	12	10
13,0	14,6	1	10
14,6	16,2	1	4
16,2	17,8	0	7

Непараметрический критерий Манна – Уитни позволяют обнаруживать лишь различия в центральных тенденциях распределений двух случайных величин. Если важно обнаружить любые расхождения в форме распределений, то можно использовать двухвыборочный критерий однородности Смирнова (Колмогорова-Смирнова).

Статистика критерия Смирнова $D_{m,n}$ определяется как максимум модуля разности между эмпирической функцией $F_1(x)$, построенной по выборке x_1, x_2, \dots, x_n , и эмпирической функцией $F_2(x)$, построенной по выборке y_1, y_2, \dots, y_m

$$D_{m,n} = \max_x \left| F_1(x) - F_2(x) \right|.$$

При справедливости гипотезы H_0 статистика $\lambda = D_{m,n} \sqrt{\frac{mn}{m+n}} < \lambda_{\text{крит}}$ имеет асимптотическое распределение Колмогорова, а $\lambda_{\text{крит}}$ определяется из $P\{\lambda > \lambda_\alpha\} = \alpha$, где α – уровень значимости.

Алгоритм решения предусматривает следующую последовательность операций.

1. Вводятся данные по диапазону размеров (ячейки A6:B11). В диапазон C6:C11 заносятся исходные данные по выборке X, в диапазон D6:D11 заносятся исходные данные по выборке Y.
2. В требуемые ячейки (рис. А) заносятся поясняющие данные.
3. В ячейку H15 заносится величина уровня значимости, в ячейку F15 – величина $\lambda_{\text{крит}}$, соответствующая этому уровню. Последняя выбирается из сокращенной (см. ниже) или полной (таблица $\lambda_{\text{крит}}$) таблиц критических значений статистики.

$P\{\lambda > \lambda_{\alpha}\} = \alpha$			
α	0,1	0,05	0,01
λ_{α}	1,22	1,36	1,63

Критические значения
статистики Колмогорова-Смирнова
(сокращенная таблица)

4. В ячейку C12 заносится формула =СУММ(C6:C11) {=SUM(C6:C11)} подсчета суммарных частот и впоследствии частностей; формула тиражируется горизонтально на диапазон D12:F12.
5. Подсчитываются частоты по выборке X – в E6 заносится формула =C6/C\$12 и тиражируется на диапазон E7:E11. Аналогично для выборки Y – в D6 заносится =D6/D\$12 и тиражируется на D7:D11.

	A	B	C	D	E	F	G	H	I
1	Нулевая гипотеза: функции распределения X и Y								
2	случайных величин не различаются								
3									
4			эксперимент	частоты		функция расп			
5	диапазон	$n(X)$	$n(Y)$	$f(X)$	$f(Y)$	$F(X)$	$F(Y)$	D	
6	8,2	9,8	1	0	0,05	0,00	0,05	0,00	0,05
7	9,8	11,4	5	7	0,25	0,18	0,30	0,18	0,12
8	11,4	13	12	10	0,60	0,26	0,90	0,45	0,45
9	13	14,6	1	10	0,05	0,26	0,95	0,71	0,24
10	14,6	16,2	1	4	0,05	0,11	1,00	0,82	0,18
11	16,2	17,8	0	7	0,00	0,18	1,00	1,00	0,00
12		сумма:	20	38	1,00	1,00			
13									
14		$D_{max} =$	0,45		$n =$	3,62			
15		$\lambda_{\text{эксп}} =$	1,64		$\lambda_{\text{крит}} =$	1,36	при $\alpha = 0,05$		
16									
17	Поскольку $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$ то нулевая гипотеза отвергается								
18	на выбранном уровне значимости $\alpha = 0,05$								
19	Различия между выборками признаются статистически								
20	значимыми								

Рис. А. Скриншот расчетного листа MS Excel для примера 2.5.

6. Строится интегральная функция распределения для выборки X – в ячейку G6 заносится формула =E6; в G7 заносится формула =G6+E7; последняя тиражируется на диапазон G8:G11.
7. Строится интегральная функция распределения для выборки Y – в ячейку H6 заносится формула =F6; в H7 заносится формула =H6+F7 и тиражируется на диапазон H8:H11.
8. Определяется максимум модуля разности между эмпирическими функциями распределений: в ячейку I6 заносится формула =ABS(G6-H6), формула тиражируется на диапазон I7:I11.
9. В ячейку C14 заносится формула =МАКС(I6:I11) {=MAX(I6:I11)}, выявляющая максимальное значение статистики D .
10. Формулой =C14* КОРЕНЬ (C12*D12 / (C12+D12)) {=C14* SQRT (C12* D12 /(C12+D12))} в ячейке C15 определяется значение $\lambda_{\text{экс}}$.
11. Сравнивая $\lambda_{\text{экс}} > \lambda_{\text{крит}}$ делается заключение: на заданном уровне значимости нулевая гипотеза отвергается, т.е.



функции распределения листьев по размерам у двух ярусов различны.