Лабораторная работа №1

Первичная обработка результатов наблюдений двух измеримых признаков и статистические точечные оценки генеральных параметров

Цель работы. Ознакомиться с методикой первичной обработки двумерной выборки, составить интервальные и дискретные вариационные ряды распределения признаков X и Y, эмпирическую функцию распределения одного из признаков, построить полигон, гистограмму и график эмпирической функции распределения. Оценить генеральные параметры.

Порядок выполнения работы.

1. Найти минимальные ($x_{\min}$, $y_{\min}$) и максимальные ($x_{\max}$, $y_{\max}$) значения выборочной совокупности X и Y. Установить границы интервалов α_i для признака X: $\alpha_i = \alpha_{i-1} + h_x$, $i = 1 \div l$, где α_0 — начало первого интервала, h_x — длина интервалов, l — число интервалов ($\alpha_0 < x_{\min} < \alpha_1 < ... < x_{\max} < x_l$). Для признака Y : $\beta_j = \beta_{j-1} + h_y$, $j = 1 \div m$, где β_0 — начало первого интервала, h_y — длина интервалов, m — число интервалов ($\beta_0 < y_{\min} < \beta_1 < ... < y_{\max} < \beta_m$).

2. По данным выборки заполнить корреляционную таблицу абсолютных частот.

3. Найти середины интервалов x_i, y_j и составить вариационный ряд каждого признака. Вычислить относительные $n_{отн}(x) = \frac{n_{x_i}}{n}$, $n_{отн}(y) = \frac{n_{y_j}}{n}$ и накопленные $n_{накоп}(x) = \sum_{x_i < x} \frac{n_{x_i}}{n}$, $n_{накоп}(y) = \sum_{y_j < y} \frac{n_{y_j}}{n}$ частоты.

4. Построить полигон частот и гистограмму относительных частот. Найти эмпирическую функцию распределения признака X по формуле: $F^*(x) = \sum_{x_i < x} \frac{n_{x_i}}{n}$ /

5. Найти выборочное среднее, выборочную дисперсию, а также несмещенные оценки дисперсии и среднеквадратического отклонения генеральной совокупности по формулам:

$$\bar{x} = \bar{x}_в = \frac{1}{n}\sum_{i=1}^{l} x_i n_{x_i}, \qquad D_{xв} = \frac{1}{n}\sum_{i=1}^{l} (x_i)^2 n_{x_i} - (\bar{x}_в)^2,$$

$$\bar{y} = \bar{y}_в = \frac{1}{n}\sum_{j=1}^{m} y_j n_{y_j} \qquad D_{yв} = \frac{1}{n}\sum_{j=1}^{m} (y_j)^2 n_{y_j} - (\bar{y}_в)^2,$$

$$S_x^{\ 2} = \frac{n}{n-1} D_{xв}, \quad s_x = \sqrt{S_x^{\ 2}},$$

$$S_y^{\ 2} = \frac{n}{n-1} D_{yв}, \quad s_y = \sqrt{S_y^{\ 2}}.$$

Рассмотрим на примере данной выборки как выполнить первичную обработку данных и оценить параметры генеральной совокупности. Система количественных признаков (X, Y) задана таблицей. Объем выборки, начало первого интервала и рекомендуемая длина интервала для каждого признака указаны.

Таблица 1

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
73	577	63	495	74	578	70	558	79	627	63	496
69	548	80	635	68	537	61	480	77	607	63	495
72	575	71	564	69	550	62	486	78	618	67	530
72	573	74	583	71	559	63	500	66	521	68	534
65	519	68	534	60	479	71	560	63	497	55	437
67	530	65	518	56	446	65	517	69	546	56	446
56	443	73	574	71	565	70	558	74	582	58	460
70	555	57	454	68	538	70	559	68	538	70	551
63	502	71	566	66	520	63	496	62	495	59	465
64	506	66	522	60	478	73	580	70	554	68	543
70	554	76	599	70	554	68	542	70	552	69	550
67	535	70	554	69	542	59	462	65	511	63	497
60	478	68	542	72		64	504	70	555	70	558
				566							

	X	Y
Объем выборки $n = 78$		
Начало первого интервала	53	420
Длина интервала	5	34

1. Для данной выборки минимальные и максимальные значения X и Y равны:

$$x_{\min} = 55, \qquad x_{\max} = 80,$$
$$y_{\min} = 437, \qquad y_{\max} = 635.$$

Найдем границы интервалов α_i и β_j:

$\alpha_0 = 53, \quad \alpha_1 = 58, \quad \alpha_2 = 63, \quad \alpha_3 = 68, \quad \alpha_4 = 73, \quad \alpha_5 = 78;$

$\beta_0 = 420, \quad \beta_1 = 454, \quad \beta_2 = 488, \quad \beta_3 = 522, \quad \beta_4 = 556, \quad \beta_5 = 590, \quad \beta_6 = 624.$

2. Составим корреляционную таблицу.

Таблица 2

Y\X	[53, 58)	[58, 63)	[63, 68)	[68, 73)	[73, 78)	[78, 83)	n_{y_j}
[420, 454)	4						4
[454, 488)	1	8	16				9
[488, 522)		1	4	21			17
[522, 556)				12	6		25
[556, 590)					2	1	18
[590, 624)						2	3
[624, 658)							2
n_{x_i}	5	9	20	33	8	3	78

3. Найдем середины интервалов и вычислим относительные и накопленные частоты. Составим дискретные и интервальные вариационные ряды для признаков X и Y. Полученные данные занесем в таблицы (табл.3, табл.4).

Таблица 3

№	$[\alpha_{i-1},\ \alpha_i)$	x_i	n_{x_i}	$n_{отн}$	$n_{накоп}$	$\frac{n_{x_i}}{n \cdot h_x}$
1	[53, 58)	55,5	5	0,064	0	0.013
2	[58, 63)	60,5	9	0,115	0,064	0,023
3	[63, 68)	65,5	20	0,257	0,179	0,051
4	[68, 73)	70,5	33	0,423	0,436	0,085
5	[73, 78)	75,5	8	0,103	0,850	0,021
6	[78, 83)	80,5	3	0,038	0,962	0,008
					1,000	

$n = 78, \quad h_x = 5.$

Таблица 4

№	$[\beta_{j-1},\ \beta_j)$	y_j	n_j	$n_{отн}$	$n_{накоп}$	$\frac{n_{y_j}}{n \cdot h_y}$
1	[420, 454)	437	4	0,051	0	0.0015
2	[454, 488)	471	9	0,115	0.051	0,0034
3	[488, 522)	505	17	0,218	0,166	0,0064
4	[522, 556)	539	25	0,321	0,384	0,0094
5	[556, 590)	573	18	0,231	0,705	0,0068
6	[590, 624)	607	3	0,038	0,936	0,0011
7	[624, 658)	641	2	0,026	0,974	0,0008
					1,000	

$n = 78, \quad h_y = 34.$

4. Построим гистограмму относительных частот и полигон частот для признака X. Гистограмма — ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной $h_x = 5$, а высоты равны плотности относительной частоты $\frac{n_{отн}}{h_x} = \frac{n_{x_i}}{n \cdot h_x}$. Гистограмма изображена на рис.1. Полигон частот — ломаная линия, соединяющая точки $(\bar{x}_i, n_{x_i})$, изображен на рис 2.

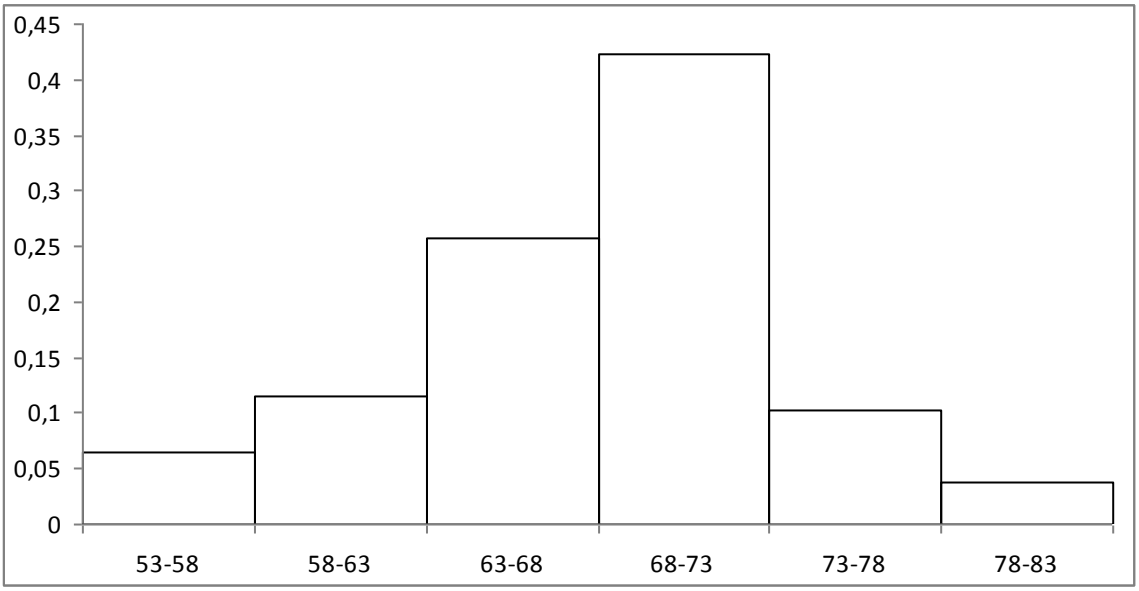

Рис.1. Гистограмма относительных частот

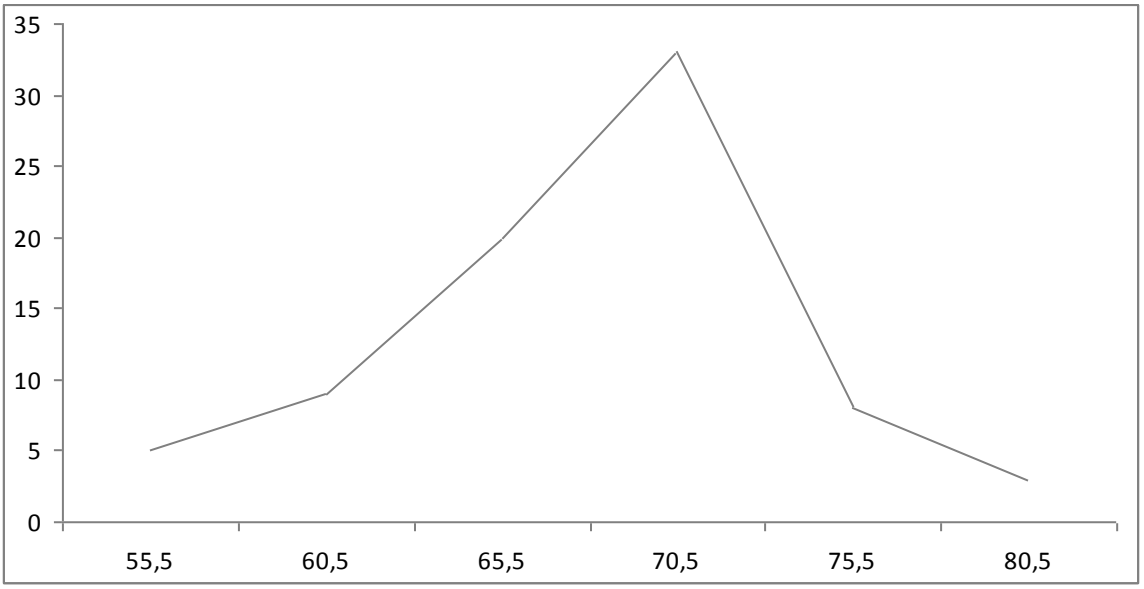

Рис. 2 Полигон частот

Запишем эмпирическую функцию распределения признака X, значения которой уже вычислены в колонке $n_{накоп}$ таблицы 3.

$$F^*(x)=\begin{cases}0, & x\le 55{,}5,\\ 0{,}064, & 55{,}5<x\le 60{,}5,\\ 0{,}179, & 60{,}5<x\le 65{,}5,\\ 0{,}436, & 65{,}5<x\le 70{,}5,\\ 0{,}859, & 70{,}5<x\le 75{,}5,\\ 0{,}962, & 75{,}5<x\le 80{,}5,\\ 1, & x>80{,}5.\end{cases}$$

5. Чтобы найти выборочное среднее и выборочную дисперсию признака X, подставим соответствующие значения x_i и n_{x_i} из таблицы 3 и получим: $\bar{x}_в=68$, $D_{xв}=32{,}45$. Аналогично выборочное среднее и выборочная дисперсия признака Y равны:, $\bar{y}_в=531{,}656$, $D_{yв}=1973{,}67$. Объем выборки равен $n=78$. Подставив это значение в соответствующие формулы для исправленной выборочной дисперсии и исправленного среднеквадратического отклонения, получим:

$$S_x{}^2=32{,}87,\quad s_x=\sqrt{32{,}87}=5{,}73,$$
$$S_y=1999{,}3,\quad s_y=\sqrt{1999{,}3}=44{,}71.$$