Решение задач к семинару по лекции 1

Задача 1

Дана выборка:

$$9, 10, 10, 11, 8, 8, 8, 11, 8, 7, 7, 8, 8, 7, 10, 11, 8, 10, 7, 10, 8, 10, 9, 8, 10$$

Объём выборки: $n = 25$.

а) Вариационный ряд

Упорядочим выборку по возрастанию:

$$7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11$$

б) Статистическое распределение

Составим таблицу частот и относительных частот:

x_i	7	8	9	10	11
n_i	4	9	2	7	3
$w_i = n_i/n$	0.16	0.36	0.08	0.28	0.12

в) Полигоны частот и относительных частот

Полигон частот строится по точкам (x_i, n_i), соединённым ломаной линией.

Полигон относительных частот строится по точкам (x_i, w_i), соединённым ломаной линией.

г) Эмпирическая функция распределения

Эмпирическая функция распределения $F_n(x)$ определяется формулой:

$$F_n(x) = \frac{\text{число элементов выборки} < x}{n}.$$

Получаем:

$$F_n(x) = \begin{cases} 0, & x \leq 7 \\ 0.16, & 7 < x \leq 8 \\ 0.52, & 8 < x \leq 9 \\ 0.60, & 9 < x \leq 10 \\ 0.88, & 10 < x \leq 11 \\ 1, & x > 11 \end{cases}$$

График $F_n(x)$ — ступенчатая функция со скачками в точках 7, 8, 9, 10, 11.

д) Эмпирические моменты

Начальный момент порядка k: $a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$.

Центральный момент порядка k: $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - a_1)^k$.

Вычислим a_1 и a_2:

$$a_1 = \frac{1}{25}(7 \cdot 4 + 8 \cdot 9 + 9 \cdot 2 + 10 \cdot 7 + 11 \cdot 3) = \frac{221}{25} = 8.84.$$

$$a_2 = \frac{1}{25}(7^2 \cdot 4 + 8^2 \cdot 9 + 9^2 \cdot 2 + 10^2 \cdot 7 + 11^2 \cdot 3) = \frac{2585}{25} = 103.4.$$

Тогда:

$$m_1 = 0 \quad \text{(по определению)},$$
$$m_2 = a_2 - a_1^2 = 103.4 - (8.84)^2 = 103.4 - 78.1456 = 25.2544.$$

Задача 2

Дано распределение двумерной случайной величины (X, Y):

	$X = 2$	$X = 3$	$X = 4$	$X = 5$
$Y = 3$	1	0	1	0
$Y = 4$	2	4	4	2
$Y = 5$	0	1	0	1

Объём выборки: $n = 16$.

Найдём эмпирический коэффициент корреляции r_{xy}.

1. Найдём маргинальные распределения

Для X:

x_i	2	3	4	5
n_i	3	5	5	3

Выборочное среднее X:

$$\bar{x} = \frac{2 \cdot 3 + 3 \cdot 5 + 4 \cdot 5 + 5 \cdot 3}{16} = \frac{56}{16} = 3.5.$$

Выборочная дисперсия X:

$$s_x^2 = \frac{2^2 \cdot 3 + 3^2 \cdot 5 + 4^2 \cdot 5 + 5^2 \cdot 3}{16} - \bar{x}^2 = \frac{212}{16} - 12.25 = 13.25 - 12.25 = 1.$$

Стандартное отклонение: $s_x = 1$.

Для Y:

y_j	3	4	5
n_j	2	12	2

Выборочное среднее Y:

$$\bar{y} = \frac{3 \cdot 2 + 4 \cdot 12 + 5 \cdot 2}{16} = \frac{64}{16} = 4.$$

Выборочная дисперсия Y:

$$s_y^2 = \frac{3^2 \cdot 2 + 4^2 \cdot 12 + 5^2 \cdot 2}{16} - \bar{y}^2 = \frac{260}{16} - 16 = 16.25 - 16 = 0.25.$$

Стандартное отклонение: $s_y = 0.5$.

2. Выборочная ковариация

Вычислим $\sum_{i,j} x_i y_j n_{ij}$:

$$\sum x_i y_j n_{ij} = 2 \cdot 3 \cdot 1 + 2 \cdot 4 \cdot 2 + 2 \cdot 5 \cdot 0 + 3 \cdot 3 \cdot 0 + 3 \cdot 4 \cdot 4 + 3 \cdot 5 \cdot 1$$
$$+ 4 \cdot 3 \cdot 1 + 4 \cdot 4 \cdot 4 + 4 \cdot 5 \cdot 0 + 5 \cdot 3 \cdot 0 + 5 \cdot 4 \cdot 2 + 5 \cdot 5 \cdot 1 = 226.$$

Тогда выборочная ковариация:

$$\text{cov}_{xy} = \frac{226}{16} - \bar{x} \cdot \bar{y} = 14.125 - 14 = 0.125.$$

3. Эмпирический коэффициент корреляции

$$r_{xy} = \frac{\mathrm{cov}_{xy}}{s_x \cdot s_y} = \frac{0.125}{1 \cdot 0.5} = 0.25.$$

Задача 3

Дано распределение роста 1000 мужчин. Предположим, что интервалы идут последовательно с шагом 3 см, начиная с 143 см. Тогда получаем следующую таблицу:

Интервал роста (см)	Количество мужчин
143–146	1
146–149	2
149–152	8
152–155	26
155–158	65
158–161	120
161–164	181
164–167	201
167–170	170
170–173	120
173–176	64
176–179	28
179–182	10
182–185	3
185–188	1

Объём выборки: $n = 1000$.

а) Гистограмма относительных частот при $h = 3$ см

Так как все интервалы имеют длину 3 см, высота каждого столбца гистограммы равна плотности относительной частоты: $h_i = \frac{w_i}{3}$, где $w_i = \frac{n_i}{1000}$.

Интервал	n_i	w_i	$h_i = w_i/3$
143–146	1	0.001	0.000333
146–149	2	0.002	0.000667
149–152	8	0.008	0.002667
152–155	26	0.026	0.008667
155–158	65	0.065	0.021667
158–161	120	0.120	0.040000
161–164	181	0.181	0.060333
164–167	201	0.201	0.067000
167–170	170	0.170	0.056667
170–173	120	0.120	0.040000
173–176	64	0.064	0.021333
176–179	28	0.028	0.009333
179–182	10	0.010	0.003333
182–185	3	0.003	0.001000
185–188	1	0.001	0.000333

Гистограмма строится по этим данным: на каждом интервале строится прямоугольник высотой h_i.

б) Гистограмма относительных частот при $h = 6$ см

Объединим соседние интервалы длиной 3 см в интервалы длиной 6 см.

Интервал	n_i	w_i	$h_i = w_i/6$
143–149	3	0.003	0.000500
149–155	34	0.034	0.005667
155–161	185	0.185	0.030833
161–167	382	0.382	0.063667
167–173	290	0.290	0.048333
173–179	92	0.092	0.015333
179–185	13	0.013	0.002167
185–191	1	0.001	0.000167

Гистограмма строится аналогично: на каждом интервале длиной 6 см строится прямоугольник высотой h_i.