Полное исследование интервального ряда

ЗАДАНИЕ. Для исследования доходов населения города, составляющего 20000 чел. по схеме бесповторной выборки было отобрано некоторое количество жителей. Получено следующее распределение жителей по месячному доходу (см. таблицу вариантов).
Построить гистограмму, полигон и кумуляту относительных частот.

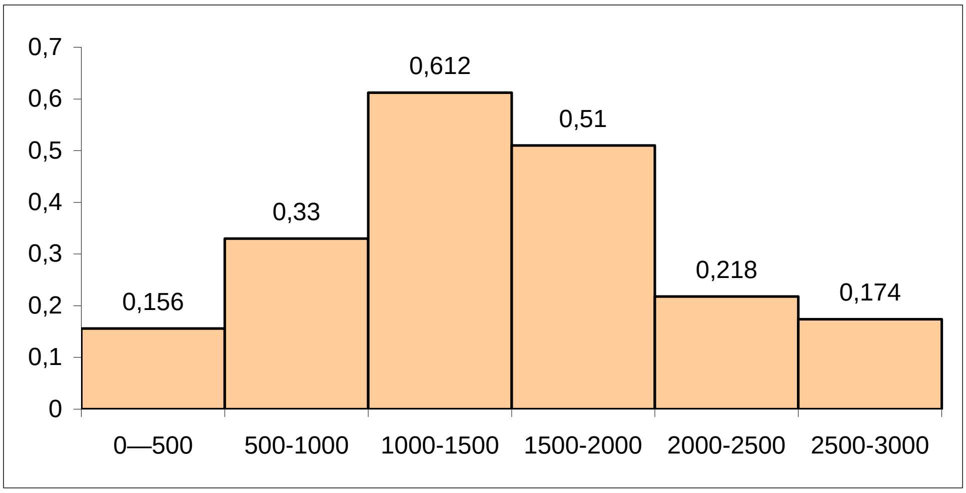
Найти вероятность того, что истинный средний доход отличается от среднего дохода по выборке не более, чем на 45 у.е. (по абсолютной величине);
определить границы, в которых заключен доход с вероятностью 0,99.
Найти объем выборки, при котором, гарантируется вероятность тех же границ, равная 0,9973.

№	Интервал x_i	Частота n_i
1	0—500	78
2	500-1000	165
3	1000-1500	306
4	1500-2000	255
5	2000-2500	109
6	2500-3000	87

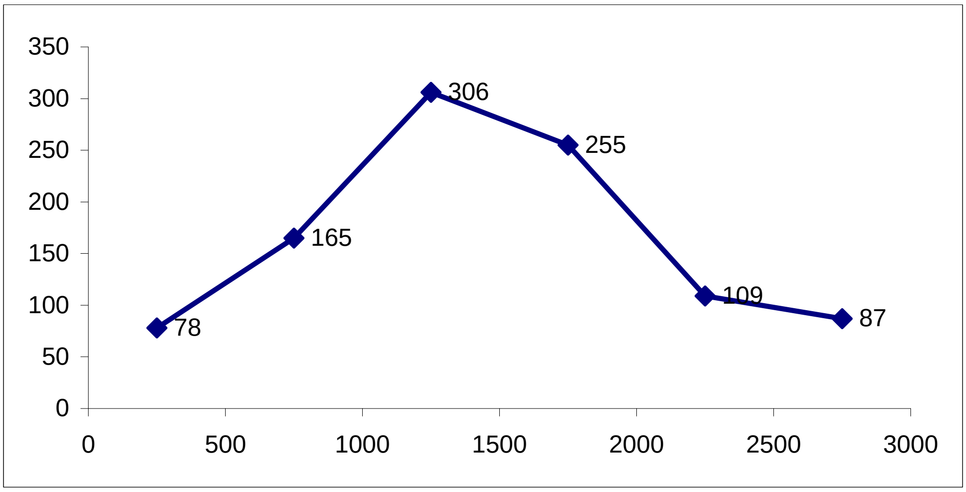
РЕШЕНИЕ.
Построим гистограмму, полигон и кумуляту относительных частот. Для этого найдем середины интервалов, вычислим относительные частоты $w_i = \frac{n_i}{n}$ и накопленные относительные частоты. Вычислим плотности частот $f_i = \frac{n_i}{h} = \frac{n_i}{500}$ Результаты внесем в таблицу:

начало	конец	x_i	n_i	w_i	накопл. отн. частоты	f_i
0	500	250	78	0,078	0,078	0,156
500	1000	750	165	0,165	0,243	0,33
1000	1500	1250	306	0,306	0,549	0,612
1500	2000	1750	255	0,255	0,804	0,51
2000	2500	2250	109	0,109	0,913	0,218
2500	3000	2750	87	0,087	1	0,174

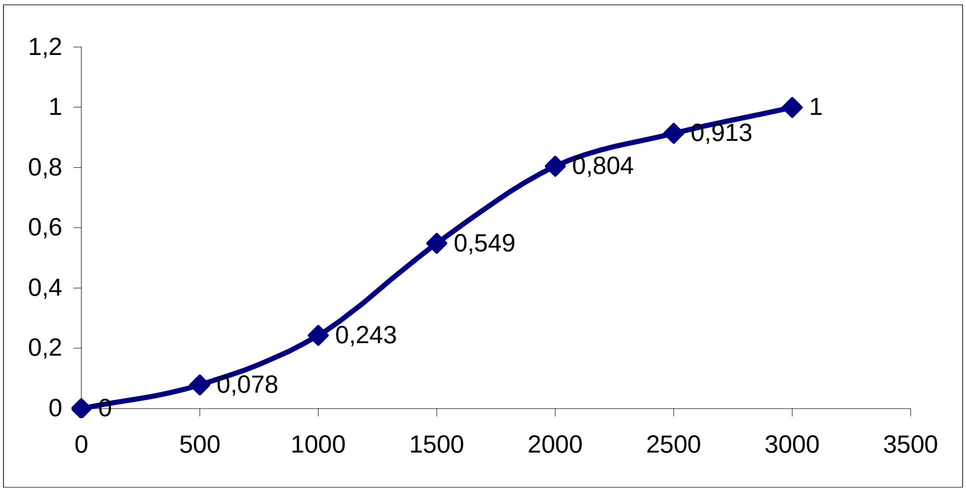
Строим гистограмму, используя значения плотностей частот.



Строим полигон частот:



Строим кумуляту относительных частот:



Для выполнения второй части задания, нужно вычислить параметры выборки.

Выборочное среднее значение $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i = \frac{1}{1000}1456500 = 1456,5.$

Исправленная (несмещенная) выборочная дисперсия

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 n_i = \frac{1}{999}445107750 \approx 445553,3.$$

Исправленное выборочное среднее квадратическое отклонение $S = \sqrt{S^2} \approx 667,498.$

Промежуточные вычисления приведены в таблице:

x_i	n_i	$x_i n_i$	$(x_i - \overline{x})^2 n_i$
250	78	19500	113540096
750	165	123750	82358471
1250	306	382500	13048529
1750	255	446250	21966274
2250	109	245250	68631005
2750	87	239250	145563376
Сумма	**1000**	**1456500**	**445107750**

Найдем вероятность того, что истинный средний доход отличается от среднего дохода по выборке не более, чем на 45 у.е. (по абсолютной величине), то есть что предельная ошибка выборки $\Delta_x = 45$.

Используем формулу для предельной ошибки:

$\Delta_x = t \cdot \mu_x = t \cdot \frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$, где $N = 20000$, $n = 1000$. Получаем:

$$\Delta_x = t \cdot \frac{667,498}{\sqrt{1000}}\sqrt{1-\frac{1000}{20000}} = 45,$$

$$t \cdot 20,574 = 45,$$

$$t \approx 2,19,$$

$$\gamma = 2\Phi(2,19) = 2 \cdot 0,4858 \approx 0,972$$

Вероятность 97,2%.

Определим границы, в которых заключен доход с вероятностью 0,99. Используем формулу:

$$\overline{x} - t\frac{S}{\sqrt{n}}\sqrt{1-\frac{n}{N}} < a < \overline{x} + t\frac{S}{\sqrt{n}}\sqrt{1-\frac{n}{N}},$$

где t определяется по доверительной вероятности из Лапласа $t(0,99) = \Phi^{-1}(0,99/2) = \Phi^{-1}(0,495) = 2,58$. Получаем:

$$1456,5-2,58\frac{667,498}{\sqrt{1000}}\sqrt{1-\frac{1000}{20000}}<a<1456,5+2,58\frac{667,498}{\sqrt{1000}}\sqrt{1-\frac{1000}{20000}}$$

$1403,42<a<1509,58$.

Найдем объем выборки, при котором гарантируется вероятность тех же границ, равная 0,9973. То есть предельная ошибка равна $\Delta_x=2,58\frac{667,498}{\sqrt{1000}}\sqrt{1-\frac{1000}{20000}}=53,08$, а $t=\Phi^{-1}(\gamma/2)=\Phi^{-1}(0,9973/2)=\Phi^{-1}(0,49865)=3$. Объем выборки тогда можно вычислить по формуле:

$$n=\frac{t^2NS^2}{\Delta_x^2N+t^2S^2}=\frac{3^2\cdot 20000\cdot 667,498^2}{53,08^2\cdot 20000+3^2\cdot 667,498^2}\approx 1329.$$