Лабораторная работа по статистике: Исследование выборки туристических фирм

Имеются следующие данные о распределении турфирм города по размеру затрат на рекламу:

Затраты на рекламу, у.е.	*Число турфирм*
6 – 8	*3*
8 – 10	*9*
10 – 12	*5*
12 – 14	*4*

Задание:

1. Постройте график вариационного ряда (гистограмму и полигон).

2. Вычислите:

– среднее значение варьирующего признака;

– моду и медиану;

– показатели вариации: размах, среднее линейное, среднее квадратическое отклонение, дисперсию, коэффициент вариации;

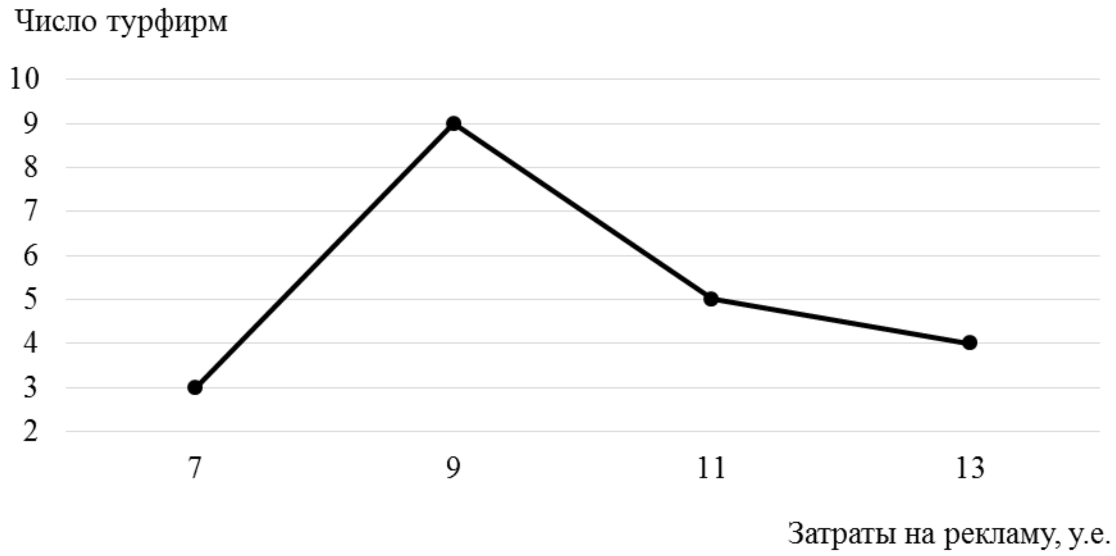
– коэффициент асимметрии.

3. Сделайте письменный вывод по каждому показателю, рассчитанному в п.2.

Решение

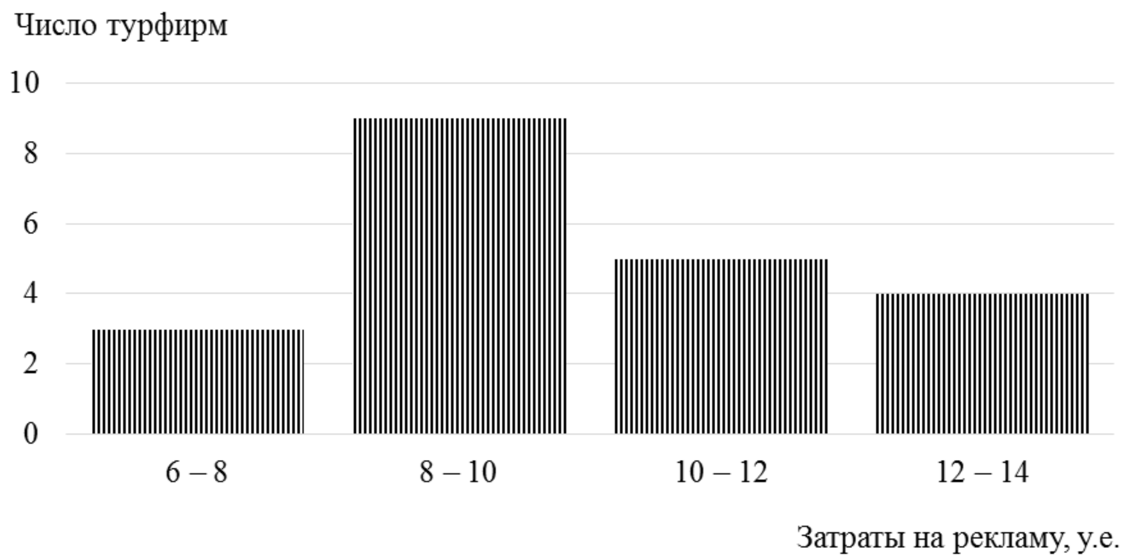
Построим график вариационного ряда (гистограмму и полигон).

При построении полигона на горизонтальной оси (ось абсцисс) откладывают значения варьирующего признака, а на вертикальной оси (ось ординат) – частоты или частости. Если полигон строят по данным интервального ряда, то в качестве абсцисс точек берут середины соответствующих интервалов.



Гистограмма – это диаграмма, на которой данные выделены с помощью столбиков.

Для построения гистограммы по оси абсцисс указывают значения границ интервалов и на их основании строят прямоугольники, высота которых пропорциональна частотам (или частостям).



Вычислим среднее значение варьирующего признака; моду и медиану; показатели вариации: размах, среднее линейное, среднее квадратическое отклонение, дисперсию, коэффициент вариации; коэффициент асимметрии.

Рассмотрим формулы для расчета показателей.

Отметим, что данный ряд распределения представляет собой группировку, поэтому формулы будем использовать взвешенные.

Для сгруппированных данных используют формулу средней арифметической взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}.$$

Чтобы посчитать среднее значение показателя в интервальном ряде распределения, надо определить середину интервала и использовать ее как значения признака x_i.

Затраты на рекламу, у.е.	Середина интервала, x_i	Число турфирм, f_i	$x_i f_i$	Кумулятивная частота, S_i
6 – 8	7	3	21	3
8 – 10	9	9	81	12
10 – 12	11	5	55	17
12 – 14	13	4	52	21
Итого	–	21	209	–

Средние затраты на рекламу:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{209}{21} = 9{,}95 \text{ у.е.}$$

Мода – величина признака, которая чаще всего встречается, то есть вариант, который в ряде распределения имеет самую большую частоту.

В интервальном ряде по самой большой частоте определяем модальный интервал.

Тогда:

$$M_o = x_0 + h\frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})},$$

где x_0 и h – нижняя граница и ширина модального интервала; $f_{mo}, f_{mo-1}, f_{mo+1}$ – частоты модального, передмодального и послемодального интервала.

По самой большой частоте – 9 турфирм, определяем модальный интервал – (8 – 10).

Тогда:

$$M_o = 8 + (10-8)\cdot\frac{9-3}{(9-3)+(9-5)} = 9{,}20 \text{ у.е.}$$

Медиана – вариант, размещенный в центре упорядоченного ряда распределения. Она делит ряд на две равные части таким образом, что по обе стороны от нее находится одинаковое количество единиц совокупности. При этом значения варьирующего признака в одной половины единиц совокупности не меньше медианы, а во второй – не больше.

В интервальном ряде распределения определяем медианный интервал – интервал, кумулятивная частота которого равняется или превышает половину объема совокупности.

Кумулятивные частоты образовываются последовательным суммированием абсолютных частот:

$$S_1 = f_1, S_2 = f_1 + f_2, ..., S_n = f_1 + \ldots + f_n.$$

Медиану вычисляем по формуле:

$$Me = x_0 + h\frac{0{,}5\sum f - S_{f_{me-1}}}{f_{me}},$$

где x_0 и h – нижняя граница и ширина медианного интервала; f_{me} – частота медианного интервала; $S_{f_{me}-1}$ – кумулятивная частота передмедианного интервала.

Половина совокупности – это $10{,}5$ наблюдения. Медианный интервал (8 – 10), поскольку его частота 12 и это первое значение кумулятивной частоты (предыдущее – 3), превышающее половину совокупности.

Найдем медиану:

$$Me = 8 + (10-8)\cdot\frac{0{,}5\cdot 21 - 3}{9} = 9{,}67 \text{ у.е.}$$

Размах вариации – это разница между наибольшим и наименьшим значениям признака:

$$R = x_{\max} - x_{\min}.$$

Он характеризует границы, в которых варьируют значение признака. Используют для предварительной оценки вариации.

В интервальном ряду распределения R определяют, как разность между верхней границей последнего интервала и нижней границей первого или же разность между средними значениями этих интервалов. Поскольку в изучаемом ряде распределения все интервалы закрытые, то:

$$R = x_{\max} - x_{\min} = 14 - 6 = 8 \text{ у. е.}$$

Среднее линейное отклонение – среднее значение из модулей отклонений каждого варианта от среднего значения.

Затраты на рекламу, у.е.	Середина интервала, x_i	Число турфирм, f_i	$x_i f_i$	$\lvert x_i - \bar{x}\rvert f_i$	$(x_i - \bar{x})^2 f$
6 – 8	7	3	21	8,86	26,15
8 – 10	9	9	81	8,57	8,16
10 – 12	11	5	55	5,24	5,49
12 – 14	13	4	52	12,19	37,15
Итого	–	21	209	34,86	76,95

$$\bar{\ell} = \frac{\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert\cdot f_i}{\sum_{i=1}^{n} f_i} = \frac{34{,}86}{21} = 1{,}66 \text{ у. е.}$$

Еще один способ усреднения отклонений вариантов от средней арифметической, что позволяет обойти сложность, обусловленную равенством нулю их алгебраической суммы – расчет квадратов отклонений вариант от средней со следующим их усреднением. Этот показатель называется дисперсией:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{n} f_i} = \frac{76{,}95}{21} = 3{,}66.$$

Среднее квадратическое отклонение – характеризует абсолютное колебание значений варьирующего признака и выражается в тех же единицах, что и варианты.

$$\sigma = \sqrt{\sigma^2}.$$

Среднеквадратическое отклонение затрат на рекламу турфирм:

$$\sigma = \sqrt{3{,}66} = 1{,}91 \text{ у. е.}$$

В статистической практике часто возникает необходимость сравнения вариации разных признаков. Для этого вышеупомянутые показатели нельзя использовать. Для такого рода сравнений статистика использует коэффициент вариации – выраженное в процентах отношения среднего квадратичного отклонения к средней арифметической:

$$V = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{1{,}91}{9{,}95} \cdot 100 = 19{,}23\%.$$

Выводы.

Средние затраты на рекламу турфирмы составляли 9,95 у.е.

Для большего количества турфирм затраты на рекламу составляли 9,20 у.е.

Для половины всех турфирм затраты на рекламу не превышали 9,67 у.е., для второй половины – были не меньше 9,67 у.е.

Размах вариации затраты на рекламу турфирм составлял 8 у.е.

В среднем затраты на рекламу турфирм отклоняются от среднего их значения на 1,66 у. е.

Дисперсия затрат на рекламу турфирм составляет 3,66.

Среднеквадратическое отклонение затрат на рекламу турфирм $\sigma = 1{,}91$ у. е.

Коэффициент вариации затрат на рекламу турфирм $V = 19{,}23\%$.

Для нормальных и близких к нормальному распределений показатель V служит индикатором однородности совокупности: принято считать, что при выполнимости неравенства

$$V \le 33\%,$$

совокупность является количественно однородной по данному признаку.

Так как коэффициент вариации затрат на рекламу турфирм $V = 19{,}23\%$ не превышает 33%, то можно считать совокупность турфирм по величине затрат на рекламу достаточно однородной.