Решение:

1. График зависимости переменных X и Y строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака X, а по оси ординат — результативного признака Y. Учитывая небольшое число пар значений переменных, по каждой из них выделим пять интервалов, используя формулу:

$h = (x_{max} - x_{min})/ k$,

где h — длина интервала,

x_{max} - наибольшее значение признака, x_{min} - наименьшее значение признака, k — число интервалов.

Для переменной X:

$h = (112 - 33)/ 5 = 15{,}8$.

Длина интервала округляется в сторону увеличения до удобного значения, h = 16.

Получим следующие границы интервалов:

33 + 16 = 49, 49 + 16 = 65, 65 + 16 = 81, 81 + 16 = 97, 97 + 16 = 113.

Аналогично для переменной Y:

h = (37,9 — 13,8)/ 5 = 4,82, h = 5.

Границы интервалов составят: 13; 18; 23; 28; 33; 38.

На график наносятся точки, координаты которых соответствуют значениям

X и Y.

Характер расположения точек на графике показывает, что связь между переменными может выражаться линейным уравнением регрессии

$$y' = b_0 + b_1 x \qquad (2)$$

На график наносятся точки, координаты которых соответствуют значениям X и Y .

Характер расположения точек на графике показывает, что связь между переменными может выражаться линейным уравнением регрессии

$$y' = b_0 + b_1x \qquad (2)$$

2. Параметры уравнения регрессии находим МНК, путем составления и решения системы нормальных уравнений регрессии.

Для проведения всех расчетов строится вспомогательная таблица:

№ п/п	x	y	x^2	y^2	xy	y'	$y - y'$	$(y - y')^2$	$A=\left\lvert\frac{y-y'}{y}\right\rvert$
1	33	13.8							
2	40	13.8							
3	36	14							
4	60	22.5							
5	55	24							
6	80	28							
7	95	32							
8	70	20.9							
9	48	22							
10	53	21.5							
11	95	32							
12	75	35							
13	63	24							
14	112	37.9							
15	70	27.5							
Итого	985	368.9							
Среднее значение	65.667	24.593							

В таблице все средние находятся по формуле средней арифметической простой:

$$\bar{x} = \sum \frac{x}{n} \qquad (3)$$

Подставим полученные суммы в систему уравнений, учитывая, что
n = 15:

$$68.9 = 15 \cdot b_0 + 985 \cdot b_1$$
$$26466.7 = 985 \cdot b_0 + 72111 \cdot b_1.$$

Решив систему, получим $b_0 = 4.7743$, $b_1 = 0.3018$.

Параметры уравнения регрессии также можно найти по формулам, которые получаются из системы нормальных уравнений.

$$b_1 = \frac{\overline{XY} - \overline{X}\overline{Y}}{\overline{X}^2 - (\overline{X})^2} = \frac{1764.447 - 65.667 \cdot 24.593}{4807.4 - (65.667)^2} = 0.30118 \qquad (4)$$

$$b_0 = \overline{Y} - b_1 \cdot \overline{X} = 24.593 - 0.3018 \cdot 65.667 = 4.7743 \qquad (5)$$

Таким образом, уравнение регрессии имеет вид

$$y' = 4.7743 + 0.3018 \cdot x \qquad (6)$$

Коэффициент регрессии показывает, что при увеличении общей площади квартиры на 1 $м^2$ стоимость квартиры в среднем увеличивается на 0,3018 тыс. у.е., или на 301,8 у.е.

Если в уравнение регрессии подставить фактические значения переменной X, то определяются возможные (теоретические) значения переменной y, которые наносятся на график в виде уравнения прямой.

Задание 3. Оценить качество уравнения регрессии.

Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации

$$\overline{A} = \frac{1}{n} \times \sum_{i}^{n} \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%$$

$$\overline{A} = \frac{135,465}{15} = 9,031\ \%$$

2.1. Вычисление коэффициента эластичности

При линейной форме связи средний коэффициент эластичности находится по формуле

$$Э = b_1 \times \frac{\bar{X}}{\bar{Y}}$$, где

$\bar{X}$ и $\bar{Y}$ – средние значения признаков.

$$Э = 0.318 \times \frac{65.667}{24.593} = 0.806.$$

Коэффициента эластичности показывает, что при увеличении общей площади квартиры на 1 % ее стоимость в среднем возрастает на 0.806 % .

2.2. Оценка значимости коэффициентов корреляции и регрессии по критерию t-Стьюдента

Оценка тесноты связи между признаками
(по t-критерию Стьюдента)

При линейной зависимости, степень тесноты связи между переменными X и Y определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$$

где σ_x и σ_y - средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - \bar{X}^2} = \sqrt{4807.4 - 65.667^2} = 22.254;$$

$$\sigma_y = \sqrt{\overline{Y^2} - \bar{Y}^2} = \sqrt{657.857 - 24.593^2} = 7.283;$$

$$r = \frac{1764.447 - 65.667 * 24.593}{22.254 * 7.283} = 0.922.$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0.922^2 = 0.850$ показывает, что 85 % различий в стоимости квартир объясняется вариацией их общей площади, а 15 % = другими, неучтенными факторами (местоположение квартир, благоустроенность территории и другими).

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции.

Выдвигаем нулевую гипотезу H_0 : коэффициента корреляции в генеральной совокупности равен нулю, и изучаемый фактор не оказывает существенного влияния на результативный признак.

$H_0 : r_s = 0$, при $H_1 : r_s \neq 0$.

Для проверки нулевой гипотезы применим t-критерий Стьюдента.
Найдем расчетное значение t-критерий:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.922}{\sqrt{\frac{1-0.922^2}{15-2}}} = 8.58.$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне значимости $\alpha = 0.05$ и числе степеней свободы $k = n - 2 = 15 - 2 = 13$.
Для двусторонней критической области $t_{кр} = 2.16$.

Используйте необходимую таблицу.

Сравним $t_{расч}$ с $t_{кр.}$ Так как $t_{расч} > t_{кр}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием t-критерия Стьюдента.

Находится расчетное значение критерия:

$$t_{\text{расч}} = \frac{b_1}{m_{b_1}};$$

$$m_{b_1} = \sqrt{\frac{\sum(y-y')^2}{(n-2)*\sum(x-\bar{X})^2}} = \sqrt{\frac{\sum(y-y')^2}{(n-2)*\sigma_x^2*n}}$$

$$= \sqrt{\frac{118.625}{(15-2)*22.254^2*15}};$$

$$t_{\text{расч}} = \frac{0.3018}{0.035} = 8.62.$$

Критическое значение t также равно 2.16. Так как $t_{расч} > t_{кр}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

2.3. Статистическая надежность результатов регрессионного анализа с использованием критерия F-Фишера

Статистическая надежность уравнения регрессии с использованием критерия F-Фишера

Статистическая надежность уравнения регрессии проверяется с использованием критерия F-Фишера

Расчетное (фактическое) значение F-критерия находится по формуле:

$$F_{расч} = \frac{\sum(y' - \bar{y})^2/k}{\sum(y' - \bar{y})^2/(n-k-1)};$$

где k – число параметров при переменных X.

Если применяется линейное уравнение регрессии, то расчетное $F_{рачс}$ упрощается.

$$F_{расч} = \frac{r^2}{1-r^2}\,(n-2) = \frac{0.85}{1-0.85} * 13 = 73.67.$$

При уровне значимости $\alpha = 0.05$ и числе степеней свободы $k_1 = k = 1$, $k_2 = n - k - 1 = 15 - 1 - 1 = 15 - 2 = 13$ по таблице находится критическое значение F –критерия. $F_{кр} = F_{0.05;1,13} = 4.67$.

Используйте необходимую таблицу.

Так как $F_{расч} > F_{кр}$, то уравнение регрессии статистически значимое или надежное.
При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты, так как $t_{b_1}^2 = t_{r}^2 = F_{\alpha,k_1,k_2}$.

2.4. Вычисление прогнозного значения результативного признака

Прогнозное значение результативного признака

Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p) .

По условию $x_p = \bar{x} * 1.2 = 65.667 = 78.8$.

Тогда прогнозное значение стоимости квартиры составит

$y'_p = b_0 + b_1 * x_p = 4.7743 + 0.3018*78.8 = 28.56$.

Значит при общей площади квартиры в 78.8 м^2 возможная стоимость квартиры составляет 25.56 тыс. у.е.