

РГПУ им. А.И. Герцена

К работе допущены _____
Работа выполнена _____
Отчёт сдан _____

Отчет по лабораторной работе №8(часть1)

«Регрессионный анализ: модели и методы»

Работу выполнили: Леонтьева А.В.,
Шадрин А.В.,
Тихонова Э.К.,
Столяренко К.А.

Факультет: ИИТиТО
Группа: 2об-ИВТ-1/20

Санкт-Петербург, 2022 г

Лабораторная работа №8.1

Регрессионный анализ: модели и методы

Выполнил: студент 2 курса РГПУ им. Герцена, ИВТ 1/2, Шадрин А.В.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Построить уравнение регрессии. Оценить его качество.

Имеются следующие выборочные данные о стоимости квартир и из общей площади в некотором городе.

y	13.8	13.8	14	22.5	24	28	32	20.9	22	21.5	32	35	24	37.9	27.5
x	33	40	36	60	55	80	95	70	48	53	95	75	63	112	70

x – общая площадь квартиры в кв. м;

y – рыночная стоимость квартиры в тыс. у.е.

Решение:

График зависимости переменных X и Y строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака X, 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:

$$h = \frac{(112 - 33)}{5} = 15,8$$

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

Границы интегралов для X составят: 33; 49; 65; 81; 97; 113.

Аналогично для переменной Y:

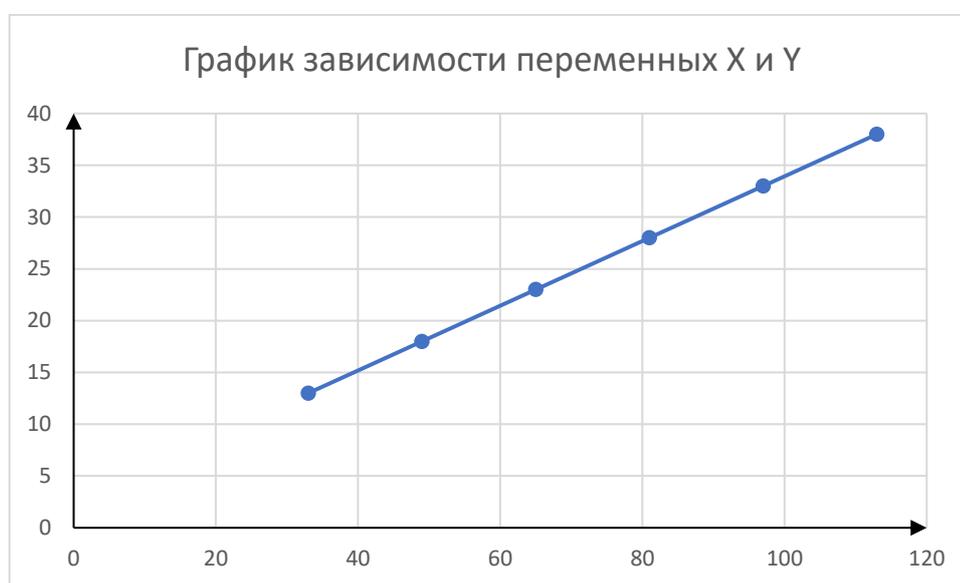
$$h = \frac{(37,9 - 13,8)}{5} = 4,82$$

Длина интервала округляется в сторону увеличения до удобного значения h=5

Границы интегралов для Y составят: 13; 18; 23; 28; 33; 38.

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

Полученный график:



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

$$y' = b_0 + b_{1x}$$

Параметры уравнения регрессии находятся методом наименьших квадратов, 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= (y-y')/y
1	33	13,8	1089	190,44	455,4	14,73412	-0,93412	0,872581	0,067689899
2	40	13,8	1600	190,44	552	16,84681	-3,04681	9,283045	0,220783264
3	36	14	1296	196	504	15,63956	-1,63956	2,688152	0,117111322
4	60	22,5	3600	506,25	1350	22,88306	-0,38306	0,146736	0,017024966
5	55	24	3025	576	1320	21,374	2,626001	6,895884	0,109416726
6	80	28	6400	784	2240	28,91931	-0,91931	0,845139	0,032832658
7	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
8	70	20,9	4900	436,81	1463	25,90119	-5,00119	25,01188	0,239291296

9	48	22	2304	484	1056	19,26131	2,73869	7,500422	0,124485904
10	53	21,5	2809	462,25	1139,5	20,77037	0,729627	0,532355	0,033936126
11	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
12	75	35	5625	1225	2625	27,41025	7,589749	57,60429	0,216849964
13	63	24	3969	576	1512	23,7885	0,2115	0,044732	0,008812515
14	112	37,9	12544	1436,41	4244,8	38,57732	-0,67732	0,458761	0,017871207
15	70	27,5	4900	756,25	1925	25,90119	1,598812	2,5562	0,058138615
Итого	985	368,9	72111	9867,85	26466,7	368,9	4,62E-14	118,6249	1,354650959
Среднее значение	65,666667	24,59333	4807,4	657,8567	1764,447	24,59333	3,08E-15	7,908328	0,090310064

В таблице все средние находятся по формуле средней арифметической простой:

$$\bar{x} = \sum \frac{x}{n}$$

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 $b_0 = 4.7743$, $b_1 = 0.3018$.

$$b_1 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{1764.447 - 65.667 \cdot 24.593}{4807.4 - (65.667)^2} = 0.3018$$

$$b_0 = \bar{Y} - b_1 \cdot \bar{X} = 24.593 - 0.3018 \cdot 65.667 = 4.7743$$

Таким образом уравнение имеет вид:

$$y' = 4.7743 + 0.3018 \cdot x$$

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.

Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \times \sum_i^n \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%$$

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

Значение средней ошибки аппроксимации находится в пределе допустимых значений.

Вывод: в ходе лабораторной работы был построен график зависимости между переменными, по которому была подобрана линейная модель регрессии, было построено уравнение регрессии, которое имеет вид $y' = 4.7743 + 0.3018 \cdot x$. Также было оценено качество уравнения регрессии с помощью средней ошибки аппроксимации, она составила $\bar{A} = 9,031\%$, её значение находится в пределе допустимых значений.

Лабораторная работа №8.1

Регрессионный анализ: модели и методы

Выполнил: студент 2 курса РГПУ им. Герцена, ИВТ 1/2, Столяренко К.А.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Построить уравнение регрессии. Оценить его качество.

Имеются следующие выборочные данные о стоимости квартир и из общей площади в некотором городе.

y	13.8	13.8	14	22.5	24	28	32	20.9	22	21.5	32	35	24	37.9	27.5
x	33	40	36	60	55	80	95	70	48	53	95	75	63	112	70

x – общая площадь квартиры в кв. м;

y – рыночная стоимость квартиры в тыс. у.е.

Решение:

График зависимости переменных X и Y строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака X, 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:

$$h = \frac{(112 - 33)}{5} = 15,8$$

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

Границы интегралов для X составят: 33; 49; 65; 81; 97; 113.

Аналогично для переменной Y:

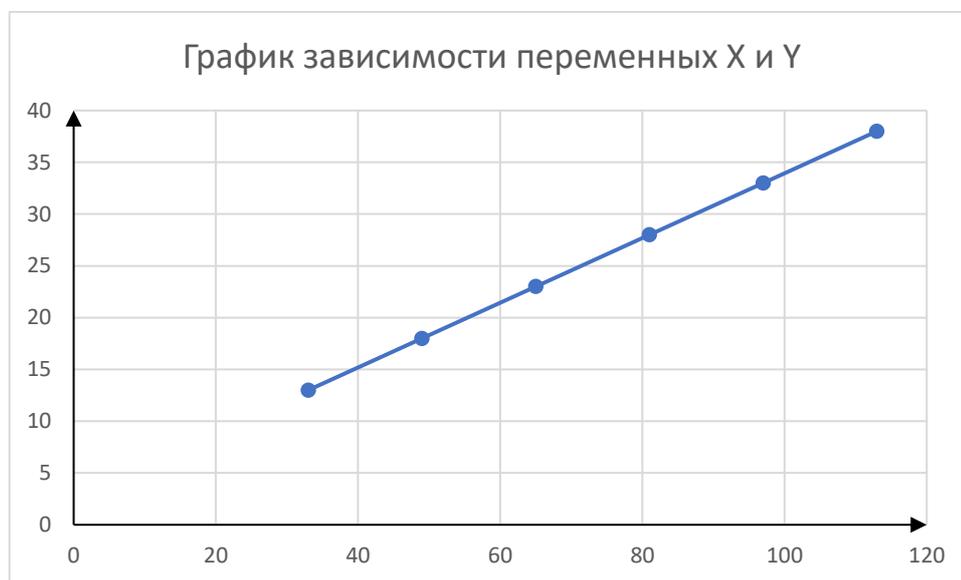
$$h = \frac{(37,9 - 13,8)}{5} = 4,82$$

Длина интервала округляется в сторону увеличения до удобного значения h=5

Границы интегралов для Y составят: 13; 18; 23; 28; 33; 38.

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

Полученный график:



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

$$y' = b_0 + b_{1x}$$

Параметры уравнения регрессии находятся методом наименьших квадратов, 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= (y-y')/y
1	33	13,8	1089	190,44	455,4	14,73412	-0,93412	0,872581	0,067689899
2	40	13,8	1600	190,44	552	16,84681	-3,04681	9,283045	0,220783264
3	36	14	1296	196	504	15,63956	-1,63956	2,688152	0,117111322
4	60	22,5	3600	506,25	1350	22,88306	-0,38306	0,146736	0,017024966
5	55	24	3025	576	1320	21,374	2,626001	6,895884	0,109416726
6	80	28	6400	784	2240	28,91931	-0,91931	0,845139	0,032832658
7	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
8	70	20,9	4900	436,81	1463	25,90119	-5,00119	25,01188	0,239291296

9	48	22	2304	484	1056	19,26131	2,73869	7,500422	0,124485904
10	53	21,5	2809	462,25	1139,5	20,77037	0,729627	0,532355	0,033936126
11	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
12	75	35	5625	1225	2625	27,41025	7,589749	57,60429	0,216849964
13	63	24	3969	576	1512	23,7885	0,2115	0,044732	0,008812515
14	112	37,9	12544	1436,41	4244,8	38,57732	-0,67732	0,458761	0,017871207
15	70	27,5	4900	756,25	1925	25,90119	1,598812	2,5562	0,058138615
Итого	985	368,9	72111	9867,85	26466,7	368,9	4,62E-14	118,6249	1,354650959
Среднее значение	65,666667	24,59333	4807,4	657,8567	1764,447	24,59333	3,08E-15	7,908328	0,090310064

В таблице все средние находятся по формуле средней арифметической простой:

$$\bar{x} = \sum \frac{x}{n}$$

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 $b_0 = 4.7743$, $b_1 = 0.3018$.

$$b_1 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{1764.447 - 65.667 \cdot 24.593}{4807.4 - (65.667)^2} = 0.3018$$

$$b_0 = \bar{Y} - b_1 \cdot \bar{X} = 24.593 - 0.3018 \cdot 65.667 = 4.7743$$

Таким образом уравнение имеет вид:

$$y' = 4.7743 + 0.3018 \cdot x$$

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.

Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \times \sum_i^n \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%$$

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

Значение средней ошибки аппроксимации находится в пределе допустимых значений.

Вывод: в ходе лабораторной работы был построен график зависимости между переменными, по которому была подобрана линейная модель регрессии, было построено уравнение регрессии, которое имеет вид $y' = 4.7743 + 0.3018 \cdot x$. Также было оценено качество уравнения регрессии с помощью средней ошибки аппроксимации, она составила $\bar{A} = 9,031\%$, её значение находится в пределе допустимых значений.

Лабораторная работа №8.1

Регрессионный анализ: модели и методы

Выполнил: студентка 2 курса РГПУ им. Герцена, ИВТ 1/2, Леонтьева А.В.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Построить уравнение регрессии. Оценить его качество.

Имеются следующие выборочные данные о стоимости квартир и из общей площади в некотором городе.

y	13.8	13.8	14	22.5	24	28	32	20.9	22	21.5	32	35	24	37.9	27.5
x	33	40	36	60	55	80	95	70	48	53	95	75	63	112	70

x – общая площадь квартиры в кв. м;

y – рыночная стоимость квартиры в тыс. у.е.

Решение:

График зависимости переменных X и Y строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака X, 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:

$$h = \frac{(112 - 33)}{5} = 15,8$$

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

Границы интегралов для X составят: 33; 49; 65; 81; 97; 113.

Аналогично для переменной Y:

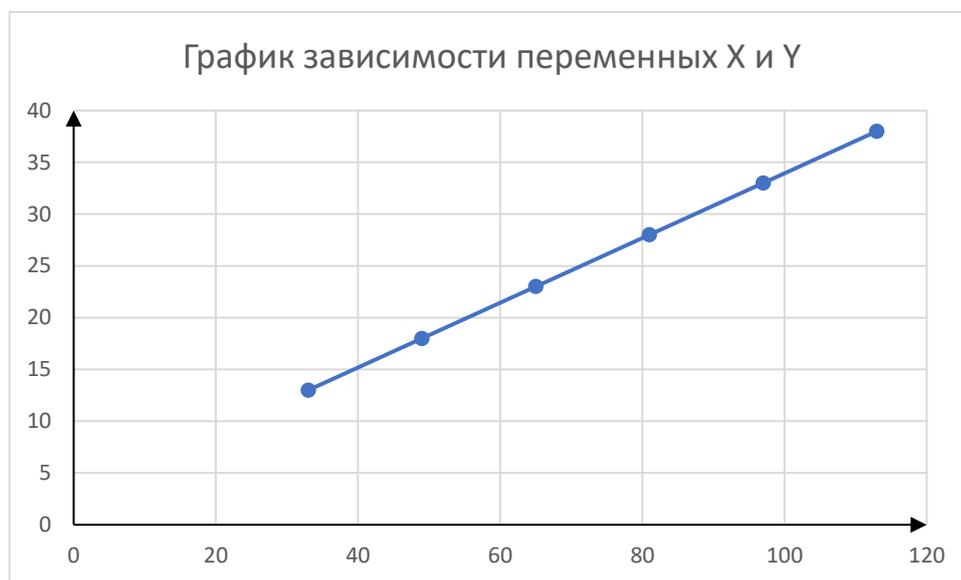
$$h = \frac{(37,9 - 13,8)}{5} = 4,82$$

Длина интервала округляется в сторону увеличения до удобного значения h=5

Границы интегралов для Y составят: 13; 18; 23; 28; 33; 38.

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

Полученный график:



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

$$y' = b_0 + b_{1x}$$

Параметры уравнения регрессии находятся методом наименьших квадратов, 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= (y-y')/y
1	33	13,8	1089	190,44	455,4	14,73412	-0,93412	0,872581	0,067689899
2	40	13,8	1600	190,44	552	16,84681	-3,04681	9,283045	0,220783264
3	36	14	1296	196	504	15,63956	-1,63956	2,688152	0,117111322
4	60	22,5	3600	506,25	1350	22,88306	-0,38306	0,146736	0,017024966
5	55	24	3025	576	1320	21,374	2,626001	6,895884	0,109416726
6	80	28	6400	784	2240	28,91931	-0,91931	0,845139	0,032832658
7	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
8	70	20,9	4900	436,81	1463	25,90119	-5,00119	25,01188	0,239291296

9	48	22	2304	484	1056	19,26131	2,73869	7,500422	0,124485904
10	53	21,5	2809	462,25	1139,5	20,77037	0,729627	0,532355	0,033936126
11	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
12	75	35	5625	1225	2625	27,41025	7,589749	57,60429	0,216849964
13	63	24	3969	576	1512	23,7885	0,2115	0,044732	0,008812515
14	112	37,9	12544	1436,41	4244,8	38,57732	-0,67732	0,458761	0,017871207
15	70	27,5	4900	756,25	1925	25,90119	1,598812	2,5562	0,058138615
Итого	985	368,9	72111	9867,85	26466,7	368,9	4,62E-14	118,6249	1,354650959
Среднее значение	65,666667	24,59333	4807,4	657,8567	1764,447	24,59333	3,08E-15	7,908328	0,090310064

В таблице все средние находятся по формуле средней арифметической простой:

$$\bar{x} = \sum \frac{x}{n}$$

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 $b_0 = 4.7743$, $b_1 = 0.3018$.

$$b_1 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{1764.447 - 65.667 \cdot 24.593}{4807.4 - (65.667)^2} = 0.3018$$

$$b_0 = \bar{Y} - b_1 \cdot \bar{X} = 24.593 - 0.3018 \cdot 65.667 = 4.7743$$

Таким образом уравнение имеет вид:

$$y' = 4.7743 + 0.3018 \cdot x$$

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.

Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \times \sum_i^n \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%$$

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

Значение средней ошибки аппроксимации находится в пределе допустимых значений.

Вывод: в ходе лабораторной работы был построен график зависимости между переменными, по которому была подобрана линейная модель регрессии, было построено уравнение регрессии, которое имеет вид $y' = 4.7743 + 0.3018 \cdot x$. Также было оценено качество уравнения регрессии с помощью средней ошибки аппроксимации, она составила $\bar{A} = 9,031\%$, её значение находится в пределе допустимых значений.

Лабораторная работа №8.1

Регрессионный анализ: модели и методы

Выполнил: студентка 2 курса РГПУ им. Герцена, ИВТ 1/2, Тихонова Э.К.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Построить уравнение регрессии. Оценить его качество.

Имеются следующие выборочные данные о стоимости квартир и из общей площади в некотором городе.

y	13.8	13.8	14	22.5	24	28	32	20.9	22	21.5	32	35	24	37.9	27.5
x	33	40	36	60	55	80	95	70	48	53	95	75	63	112	70

x – общая площадь квартиры в кв. м;

y – рыночная стоимость квартиры в тыс. у.е.

Решение:

График зависимости переменных X и Y строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака X, 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:

$$h = \frac{(112 - 33)}{5} = 15,8$$

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

Границы интегралов для X составят: 33; 49; 65; 81; 97; 113.

Аналогично для переменной Y:

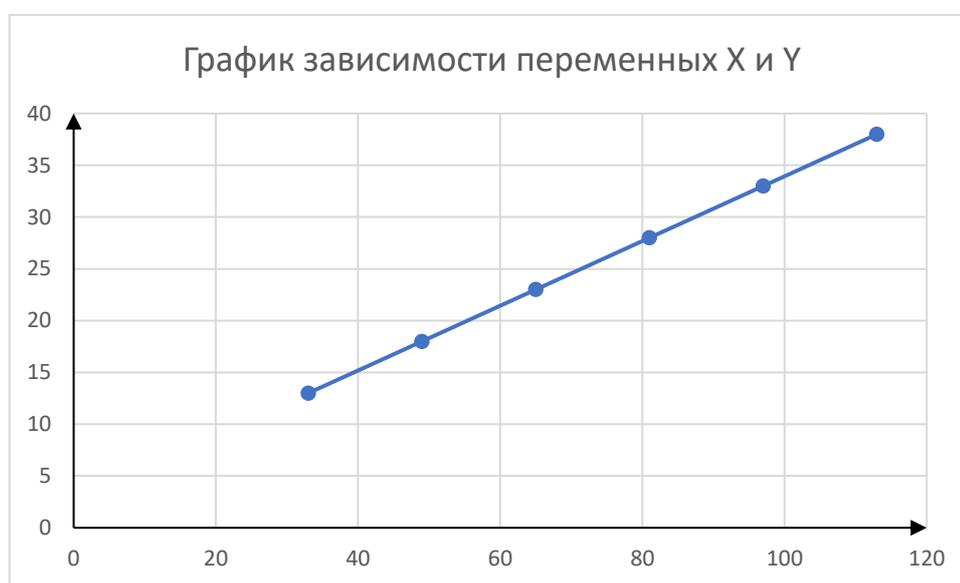
$$h = \frac{(37,9 - 13,8)}{5} = 4,82$$

Длина интервала округляется в сторону увеличения до удобного значения h=5

Границы интегралов для Y составят: 13; 18; 23; 28; 33; 38.

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

Полученный график:



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

$$y' = b_0 + b_{1x}$$

Параметры уравнения регрессии находятся методом наименьших квадратов, 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= (y-y')/y
1	33	13,8	1089	190,44	455,4	14,73412	-0,93412	0,872581	0,067689899
2	40	13,8	1600	190,44	552	16,84681	-3,04681	9,283045	0,220783264
3	36	14	1296	196	504	15,63956	-1,63956	2,688152	0,117111322
4	60	22,5	3600	506,25	1350	22,88306	-0,38306	0,146736	0,017024966
5	55	24	3025	576	1320	21,374	2,626001	6,895884	0,109416726
6	80	28	6400	784	2240	28,91931	-0,91931	0,845139	0,032832658
7	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
8	70	20,9	4900	436,81	1463	25,90119	-5,00119	25,01188	0,239291296

9	48	22	2304	484	1056	19,26131	2,73869	7,500422	0,124485904
10	53	21,5	2809	462,25	1139,5	20,77037	0,729627	0,532355	0,033936126
11	95	32	9025	1024	3040	33,4465	-1,4465	2,092374	0,045203248
12	75	35	5625	1225	2625	27,41025	7,589749	57,60429	0,216849964
13	63	24	3969	576	1512	23,7885	0,2115	0,044732	0,008812515
14	112	37,9	12544	1436,41	4244,8	38,57732	-0,67732	0,458761	0,017871207
15	70	27,5	4900	756,25	1925	25,90119	1,598812	2,5562	0,058138615
Итого	985	368,9	72111	9867,85	26466,7	368,9	4,62E-14	118,6249	1,354650959
Среднее значение	65,666667	24,59333	4807,4	657,8567	1764,447	24,59333	3,08E-15	7,908328	0,090310064

В таблице все средние находятся по формуле средней арифметической простой:

$$\bar{x} = \sum \frac{x}{n}$$

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 $b_0 = 4.7743$, $b_1 = 0.3018$.

$$b_1 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{1764.447 - 65.667 \cdot 24.593}{4807.4 - (65.667)^2} = 0.3018$$

$$b_0 = \bar{Y} - b_1 \cdot \bar{X} = 24.593 - 0.3018 \cdot 65.667 = 4.7743$$

Таким образом уравнение имеет вид:

$$y' = 4.7743 + 0.3018 \cdot x$$

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.

Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \times \sum_i^n \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%$$

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

Значение средней ошибки аппроксимации находится в пределе допустимых значений.

Вывод: в ходе лабораторной работы был построен график зависимости между переменными, по которому была подобрана линейная модель регрессии, было построено уравнение регрессии, которое имеет вид $y' = 4.7743 + 0.3018 \cdot x$. Также было оценено качество уравнения регрессии с помощью средней ошибки аппроксимации, она составила $\bar{A} = 9,031\%$, её значение находится в пределе допустимых значений.