



Институт интеллектуальных кибернетических систем

КАФЕДРА КИБЕРНЕТИКИ

БДЗ

по курсу "Математическая статистика"

студента группы Б21-504

Мандрова А.П.

Вариант № 16

Оценка: _____

Подпись: _____

2023 г.

ОТЧЕТ № 1

по теме «Проверка статистических гипотез»

Вариант №16

ФИО студента: Мандров А.П. группа: Б21-504

Оценка: _____ Подпись: _____

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение ($\alpha = 0.1$)	Вывод
4.1	$H_0: F_X(x) = N$	Хи-квадрат	H_0 отклоняется	$F_X(x) \neq N$
4.2	$H_0: F_X(x) = N$	Харке-Бера	H_0 отклоняется	$F_X(x) \neq N$
5.1	$H_0: F_1(x) = F_2(x)$	знаков	H_0 отклоняется	$F_1(x) \neq F_2(x)$
5.2	$H_0: F_1(x) = F_2(x)$	Хи-квадрат	H_0 отклоняется	$F_1(x) \neq F_2(x)$

Выводы:

В результате проведённого в п.4 статистического анализа обнаружено, что гипотеза о нормальном распределении квадратных футов жилой площади отклоняется

В результате проведённого в п.5 статистического анализа обнаружено, что гипотеза о равенстве распределений оценочной цены 1 и оценочный цены 2:

- при вероятности ошибки первого рода равной 0.01 – принимается
- при вероятности ошибки первого рода равной 0.1 – отклоняется

ОТЧЕТ № 2

по теме «Анализ статистических взаимосвязей»

Вариант №16

ФИО студента: Мандров А.П. группа: Б21-504

Оценка: _____ Подпись: _____

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение ($\alpha = 0.1$)	Вывод
6	$H_0: F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$	χ^2 -квaдpат	H_0 принимается	$F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$
7	$H_0: m_{Y X=x_1} = m_{Y X=x_2} = m_{Y X=x_3}$	ANOVA	H_0 принимается	$F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$

Выводы:

В результате проведённого в п.6 статистического анализа обнаружено, что нет оснований отвергать гипотезу о независимости распределения углового расположения домов от сектора, в котором эти дома расположены

В результате проведённого в п.7 статистического анализа обнаружено, что гипотеза о зависимости распределения концентрации оксидов азота от расположения в разных секторах города отвергается

ОТЧЕТ № 3

по теме «Основы регрессионного анализа»

Вариант №16

ФИО студента: Мандров А.П. группа: Б21-504

Оценка: _____ Подпись: _____

Сводная таблица свойств различных регрессионных моделей:

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	Нет	Нет	Нет
Значимость	При $\alpha = 0.1$	При $\alpha = 0.1$	При $\alpha = 0.1$
Адекватность	Нет	Нет	Нет
Степень тесноты связи	Слабая	Слабая	Слабая

Выводы:

В результате проведенного в п.8 статистического анализа обнаружено, что гипотеза о монотонной зависимости количества функций в доме от концентрации оксидов азота и оценочной стоимости дома 1 отвергается

В результате проведенного в п.9 статистического анализа обнаружено, что:

- 1) линейная корреляционная связь между оценочной стоимостью 1 и концентрацией оксида азота слаба
- 2) корреляционная связь оценочной стоимости 1 связанная с концентрацией оксида азота и квадратом концентрации слаба
- 3) корреляционная связь оценочной стоимости 1 связанная с концентрацией оксида азота и взвешенным расстоянием до пяти городских центров занятости слаба

1. Описательные статистики

1.1. Выборочные характеристики

Анализируемый признак 1 – Appraisal price² (\$hundreds) (D2)

Анализируемый признак 2 – Appraisal price³ (\$hundreds) (D3)

Анализируемый признак 3 – Square feet of living space (D4)

а) Привести формулы расчёта выборочных характеристик

Выборочная хар-ка	Формула расчета
Объём выборки	177
Среднее	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Выборочная дисперсия	$D_X^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Выборочное среднеквадратическое отклонение	$\sigma_X^* = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Выборочный коэффициент асимметрии	$\gamma_X^* = \frac{\mu_{3X}^*}{(\sigma_X^*)^3}$
Выборочный эксцесс	$\varepsilon_X^* = \frac{\mu_{4X}^*}{(\sigma_X^*)^4} - 3$

б) Рассчитать выборочные характеристики

Выборочная хар-ка	Признак 1	Признак 2	Признак 3
Среднее	1166.461	1260.444	1653.854
Выборочная дисперсия	145636.59	149682.503	271941.252
Выборочное среднеквадратическое отклонение	381.624	386.888	521.48
Выборочный коэффициент асимметрии	1.381	1.366	1.172
Выборочный эксцесс	1.451	1.45	1.514

1.2. Группировка и гистограммы частот

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117

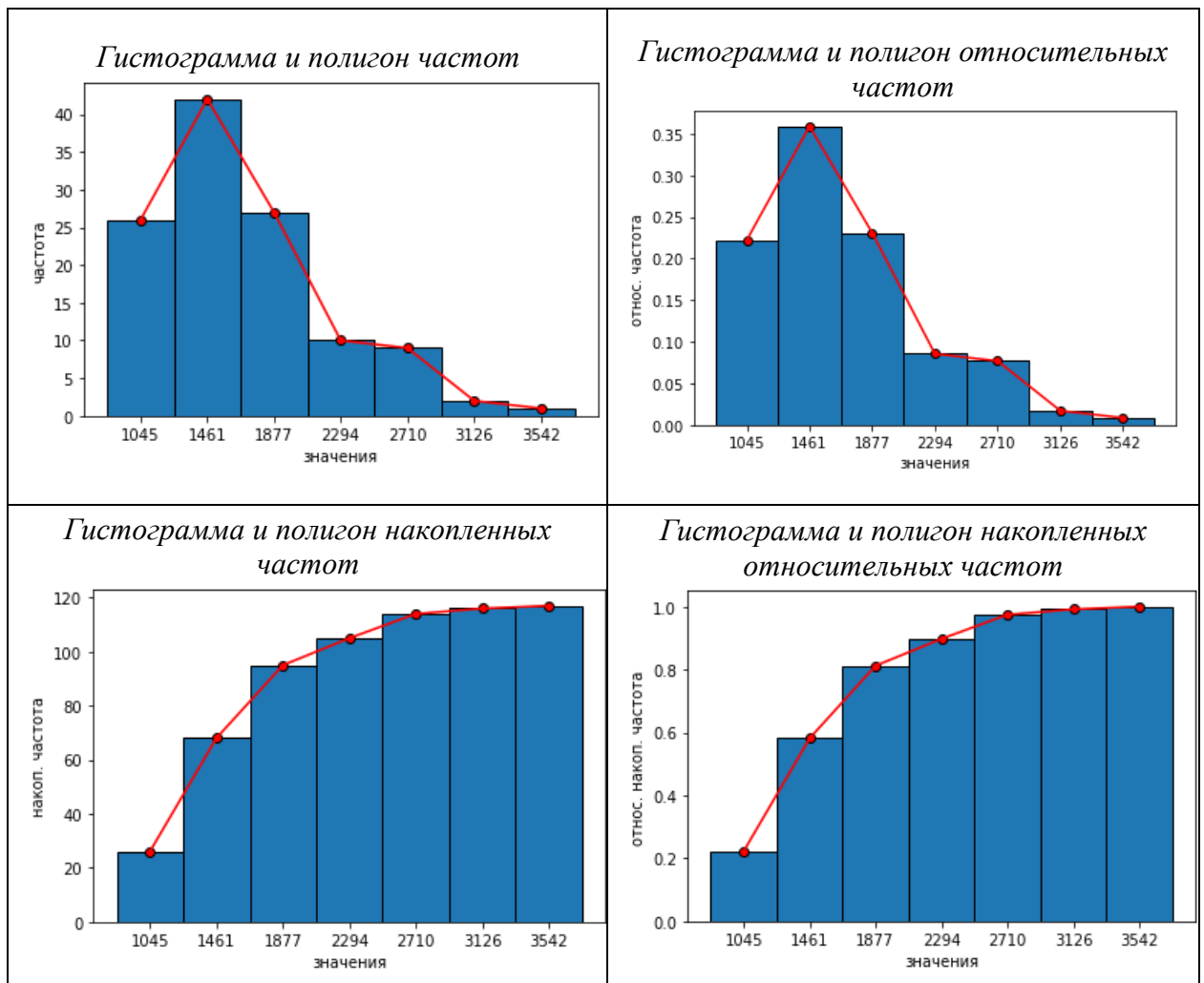
а) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
7	Формула Стёрджеса: $k \approx [1 + \log_2 n]$	416.142

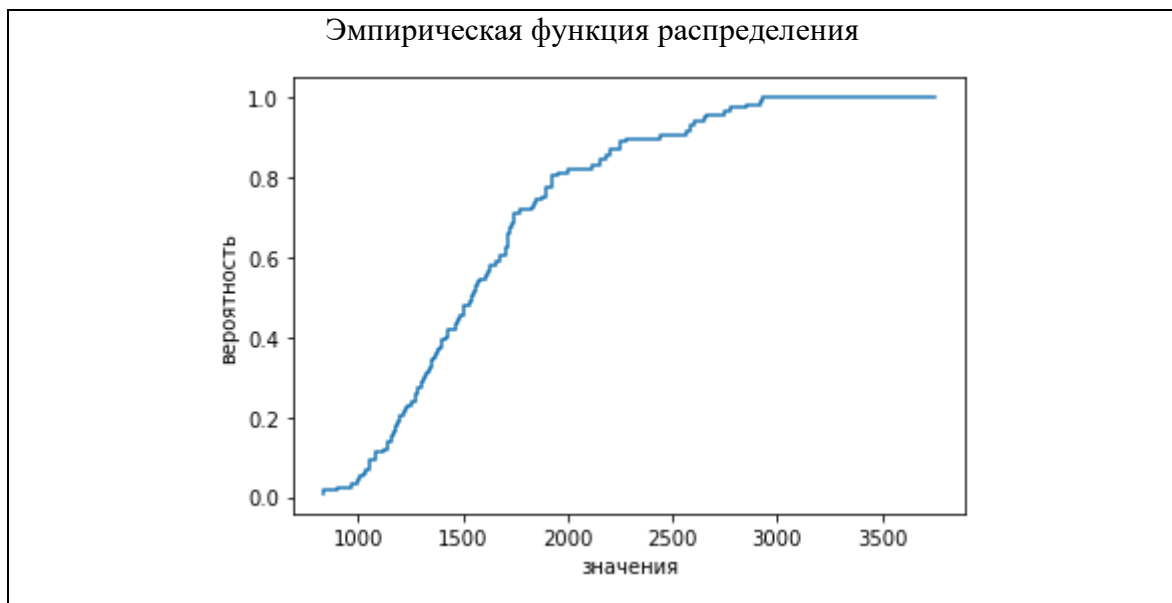
б) Построить таблицу частот

Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Накопл. частота	Относит. накопл. частота
1	837	1253	26	0.222	26	0.222
2	1253	1669	42	0.359	68	0.581
3	1669	2085	27	0.231	95	0.812
4	2085	2501	10	0.085	105	0.897
5	2501	2917	9	0.077	114	0.974
6	2917	3333	2	0.017	116	0.991
7	3333	3750	1	0.009	117	1

в) Построить гистограммы частот и полигоны частот



г) Построить график эмпирической функции распределения



2. Интервальные оценки

2.1. Доверительные интервалы для мат. ожидания

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117

Оцениваемый параметр – m (при неизвестной σ)

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$\bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)$
Верхняя граница	$\bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	1527.053	1557.956	1573.572
Верхняя граница	1780.656	1749.753	1734.136

2.2. Доверительные интервалы для дисперсии

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117

Оцениваемый параметр – σ^2 (при неизвестном m)

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}$
Верхняя граница	$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	200136	215416	223846
Верхняя граница	395136	361214	345336

2.3. Доверительные интервалы для разности мат. ожиданий

Анализируемый признак 1 – Appraisal price1 (\$hundreds) (D1)

Анализируемый признак 2 – Appraisal price2 (\$hundreds) (D2)

Объёмы выборок – 117

Оцениваемый параметр – $(m_1 - m_2)$ (при неизвестных σ_1^2 и σ_2^2)

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$(\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} t_{1-\frac{\alpha}{2}}(k)$
Верхняя граница	$(\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} t_{1-\frac{\alpha}{2}}(k)$

Где:

$$k = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	-4601	-3814	-3414
Верхняя граница	4394	3606	3206

2.4. Доверительные интервалы для отношения дисперсий

Анализируемый признак 1 – Appraisal price1 (\$hundreds) (D1)

Анализируемый признак 2 – Appraisal price2 (\$hundreds) (D2)

Объёмы выборок – 117

Оцениваемый параметр – σ_1^2 / σ_2^2 (при неизвестном m_1 и m_2)

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$\frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2}}(n_2 - 1, n_1 - 1)$
Верхняя граница	$\frac{S_1^2}{S_2^2} f_{1 - \frac{\alpha}{2}}(n_2 - 1, n_1 - 1)$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	0.608	0.683	0.725
Верхняя граница	1.594	1.42	1.338

3. Проверка статистических гипотез о математических ожиданиях и дисперсиях

3.1. Проверка статистических гипотез о математических ожиданиях

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117

Статистическая гипотеза – $H_0: m = m_0$
 $H': m \neq m_0$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение
Формула расчета статистики критерия	$Z = \frac{\bar{X} - m_0}{\frac{S}{\sqrt{n}}}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim T(n - 1)$
Формулы расчета критических точек	$z' = z_{\frac{\alpha}{2}}$ $z'' = -z_{\frac{\alpha}{2}}$
Формула расчета p -value	$p = 2 * \min (F(Z < z_{\text{выб}} _{H_0}), 1 - F(Z < z_{\text{выб}} _{H_0}))$

б) Выбрать произвольные значения m_0 и проверить статистические гипотезы

m_0	Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
1653	0.1	0.018	0.986	H_0 принимается	$m = 1653$
1660	0.1	-0.127	0.899	H_0 принимается	$m = 1660$
1800	0.1	-3.018	0.003	H_0 отклоняется	$m \neq 1800$

3.2. Проверка статистических гипотез о дисперсиях

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117

Статистическая гипотеза – $H_0: \sigma = \sigma_0$
 $H': \sigma \neq \sigma_0$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение
Формула расчета статистики критерия	$Z = \frac{(n-1)S^2}{\sigma_0^2}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(n-1)$
Формулы расчета критических точек	$z' = z_{\frac{\alpha}{2}}$ $z'' = z_{1-\frac{\alpha}{2}}$
Формула расчета p -value	$p = 2 * \min (F(Z < z_{\text{выб}} _{H_0}), 1 - F(Z < z_{\text{выб}} _{H_0}))$

б) Выбрать произвольные значения σ_0 и проверить статистические гипотезы

σ_0	Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
271941	0.1	117	0.913	H_0 принимается	$\sigma = 271941$
275000	0.1	115.699	0.981	H_0 принимается	$\sigma = 275000$
200000	0.1	159.086	0.010	H_0 отклоняется	$\sigma = 200000$

3.3. Проверка статистических гипотез о равенстве математических ожиданий

Анализируемый признак 1 – Appraisal price1 (\$hundreds) (D1)

Анализируемый признак 2 – Appraisal price2 (\$hundreds) (D2)

Объёмы выборок – 117

Статистическая гипотеза – $H_0: m_1 = m_2$
 $H': m_1 \neq m_2$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение
Формула расчета статистики критерия	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim T\left(\frac{1}{k}\right)$ $k = \frac{\left(\frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)^2}{n_1 - 1} + \frac{\left(\frac{\frac{S_2^2}{n_2}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)^2}{n_2 - 1}$

Формулы расчета критических точек	$z' = z_{\frac{\alpha}{2}}$ $z'' = z_{1-\frac{\alpha}{2}}$
Формула расчета p -value	$p = 2 * \min (F(Z < z_{\text{выб}} _{H_0}), 1 - F(Z < z_{\text{выб}} _{H_0}))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	-2.073	0.039	H_0 принимается	$m_1 = m_2$
0.05			H_0 отклоняется	$m_1 \neq m_2$
0.1			H_0 отклоняется	$m_1 \neq m_2$

3.4. Проверка статистических гипотез о равенстве дисперсий

Анализируемый признак 1 – Appraisal price1 (\$hundreds) (D1)

Анализируемый признак 2 – Appraisal price2 (\$hundreds) (D2)

Объёмы выборок – 117

Статистическая гипотеза – $H_0: \sigma_1 = \sigma_2$
 $H': \sigma_1 \neq \sigma_2$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение
Формула расчета статистики критерия	$Z = \frac{S_1^2}{S_2^2}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim F(n_1 - 1, n_2 - 1)$
Формулы расчета критических точек	$z' = z_{\frac{\alpha}{2}}$ $z'' = z_{1-\frac{\alpha}{2}}$
Формула расчета p -value	$p = 2 * \min (F(Z < z_{\text{выб}} _{H_0}), 1 - F(Z < z_{\text{выб}} _{H_0}))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение	Вывод
0.01	0.977	0.9	H_0 принимается	$\sigma_1 = \sigma_2$
0.05			H_0 принимается	$\sigma_1 = \sigma_2$
0.1			H_0 принимается	$\sigma_1 = \sigma_2$

4. Критерии согласия

Анализируемый признак – Square feet of living space (D4)

Объём выборки – 117 (но один элемент удален, как выброс)

4.1. Критерий хи-квадрат

Теоретическое распределение – нормальное.

Статистическая гипотеза – $H_0: F_X(x) = N$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$	n_i - число элементов в интервале Δ_i p_i - вероятность того что $X \in \Delta_i$ при $X \sim G(x)$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(k - r - 1)$	k – количество интервалов r – количество неизвестных параметров N
Формула расчета критической точки	$z' = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

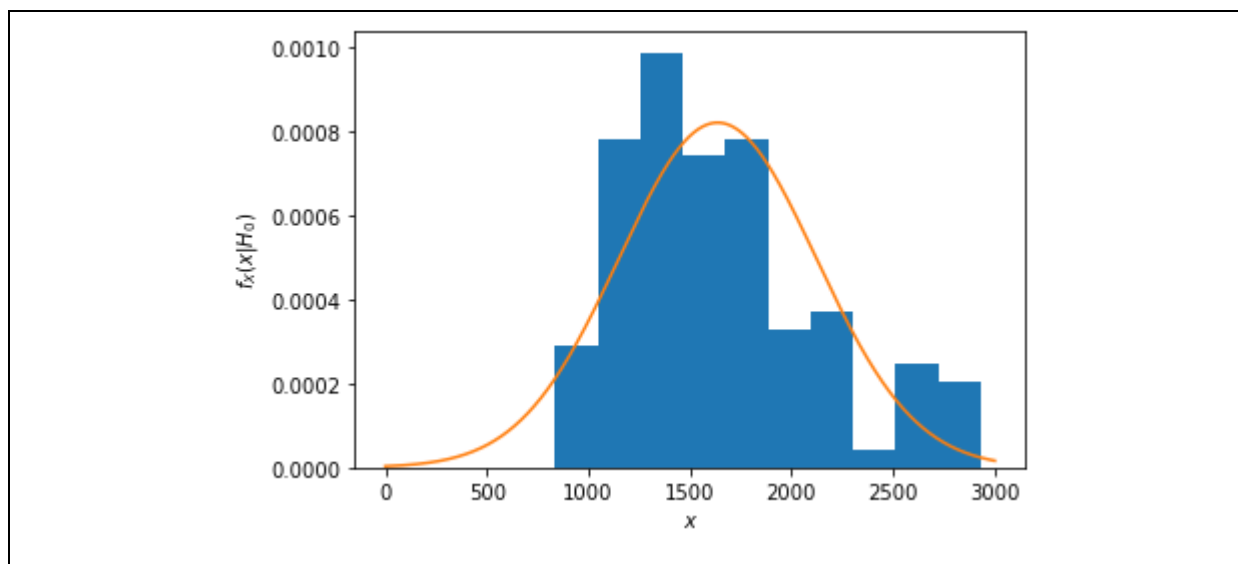
б) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
9 (фактически 8, так как два последних интервала были склеены из-за $np_i < 5$)	Формула Стёрджеса: $k \approx [1 + \log_2 n]$	261.75 (у последнего интервала ширина 523.5)

в) Построить таблицу частот

Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Вероятность попадания в интервал при условии истинности основной гипотезы
1	837	1098.75	12	0.104	0.084
2	1098.75	1360.5	28	0.243	0.151
3	1360.5	1622.25	25	0.217	0.203
4	1622.25	1884	22	0.191	0.206
5	1884	2145.75	9	0.078	0.158
6	2145.75	2407.5	8	0.069	0.091
7	2407.5	2931.0	12	0.104	0.052

г) Построить гистограмму относительных частот и функцию плотности теоретического распределения на одном графике



д) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	18.442	0.001	H_0 отклоняется	$F_X(x) \neq N$
0.05	18.442	0.001	H_0 отклоняется	$F_X(x) \neq N$
0.1	18.442	0.001	H_0 отклоняется	$F_X(x) \neq N$

4.2. Проверка гипотезы о нормальности на основе коэффициента асимметрии и эксцесса (критерий Харке-Бера)

Статистическая гипотеза – $H_0: F_X(x) = N$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \left(\frac{\gamma_X^*}{\sqrt{\frac{6}{n}}} \right)^2 + \left(\frac{\varepsilon_X^*}{\sqrt{\frac{24}{n}}} \right)^2$	γ_X^* - выборочный коэффициент асимметрии ε_X^* - выборочный коэффициент эксцесса n – объем выборки
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(2)$	
Формула расчета критической точки	$z' = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	37.968	0	H_0 отклоняется	$F_X(x) \neq N$
0.05			H_0 отклоняется	$F_X(x) \neq N$
0.1			H_0 отклоняется	$F_X(x) \neq N$

Вывод (в терминах предметной области)

В результате проведённого в п.4 статистического анализа обнаружено, что гипотеза о нормальном распределении квадратных футов жилой площади отклоняется

5. Проверка однородности выборок

Анализируемый признак 1 – Appraisal price1 (\$hundreds) (D1)

Анализируемый признак 2 – Appraisal price2 (\$hundreds) (D2)

Объёмы выборок – 117

5.1 Критерий знаков

Статистическая гипотеза – $H_0 : F_1(x) = F_2(x)$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{K_+ - \frac{n}{2}}{\frac{\sqrt{n}}{2}}$ $K_+ = \sum_{i=1}^n 1[\text{sign}(x_i - y_i) > 0]$	Z – статистика критерия n – объем выборки
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim N(0, 1)$	
Формула расчета критической точки	$z' = z_{\frac{\alpha}{2}}$ $z'' = z_{1-\frac{\alpha}{2}}$	α – уровень значимости, критическая область двусторонняя
Формула расчета p -value	$p = 2 * \min(F(Z < z_{\text{выб}} _{H_0}), 1 - F(Z < z_{\text{выб}} _{H_0}))$	$z_{\text{выб}}$ – реализация статистики критерия Z

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	-58.5	0	H_0 отклоняется	$F_1(x) \neq F_2(x)$
0.05			H_0 отклоняется	$F_1(x) \neq F_2(x)$
0.1			H_0 отклоняется	$F_1(x) \neq F_2(x)$

5.2. Критерий хи-квадрат

Статистическая гипотеза – $H_0 : F_1(x) = F_2(x)$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = n_X n_Y \sum_{i=1}^k \frac{1}{n_i^{(X)} + n_i^{(Y)}} \left(\frac{n_i^{(X)}}{n_X} - \frac{n_i^{(Y)}}{n_Y} \right)^2$	$n_i^{(X)}$ - число наблюдений СВ X в интервале Δ_i n_X - число наблюдений СВ X $n_i^{(Y)}$ - число наблюдений СВ Y в интервале Δ_i n_Y - число наблюдений СВ Y
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(k-1)$	k – число интервалов
Формула расчета критической точки	$z^* = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета <i>p-value</i>	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

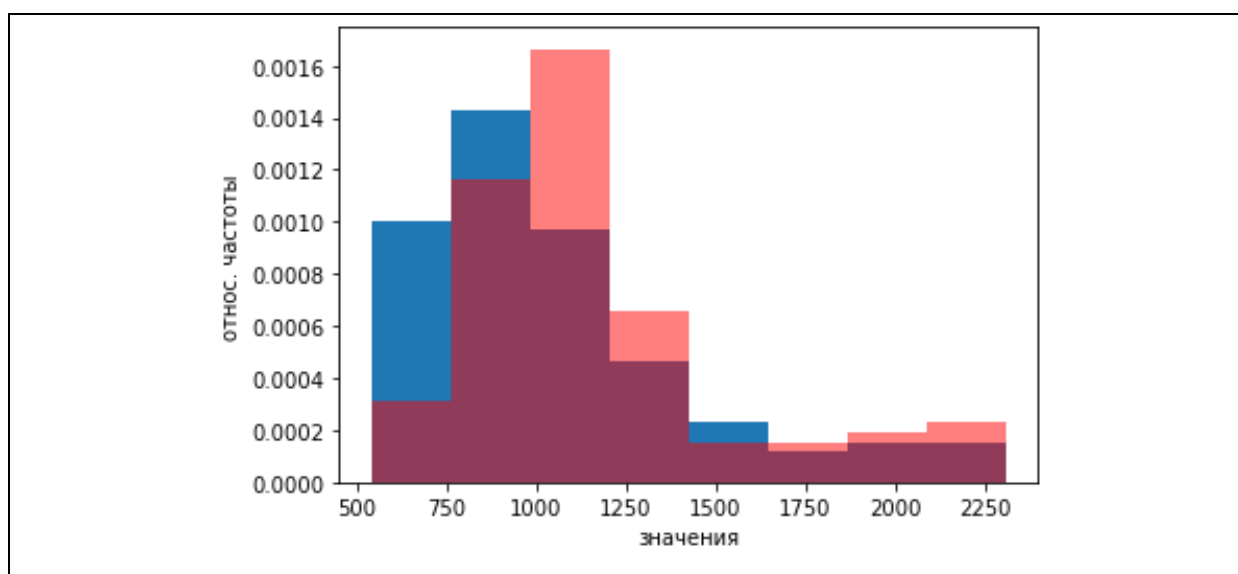
б) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
8	Формула Стёрджеса: $k \approx [1 + \log_2 n]$	220.875

в) Построить таблицу частот

Номер интервала	Нижняя граница	Верхняя граница	Частота признака 1	Частота признака 2	Относит. частота признака 1	Относит. частота признака 2
1	540	760.875	26	8	0.222	0.068
2	760.875	981.75	38	30	0.316	0.256
3	981.75	1202.625	25	43	0.213	0.368
4	1202.625	1423.5	12	17	0.103	0.145
5	1423.5	1644.375	6	4	0.051	0.034
6	1644.375	1865.25	3	4	0.025	0.034
7	1865.25	2086.125	4	5	0.034	0.0427
8	2086.125	2307	4	5	0.034	0.0427

г) Построить гистограммы относительных частот на одном графике



д) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	16.653	0.02	H_0 принимается	$F_1(x) = F_2(x)$
0.05			H_0 отклоняется	$F_1(x) \neq F_2(x)$
0.1			H_0 отклоняется	$F_1(x) \neq F_2(x)$

Вывод (в терминах предметной области)

В результате проведенного в п.5 статистического анализа обнаружено, что гипотеза о равенстве распределений оценочной цены 1 и оценочной цены 2:

- при вероятности ошибки первого рода равной 0.01 – принимается
- при вероятности ошибки первого рода равной 0.1 – отклоняется

6. Таблицы сопряжённости

Факторный признак x – Location in sector of city (D7)

Результативный признак y – Corner location (D8)

Объёмы выборок – 117

Статистическая гипотеза – $H_0: F_Y(y|x_1) = F_Y(y|x_2) = F_Y(y|x_3)$

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$	n_{ij} - число наблюдений при $X = x^{(i)}$, где $x^{(i)}$ - i-ый вариант СВ X и при $Y = y^{(j)}$, где $y^{(j)}$ - j-ый вариант СВ Y m_{ij} - n_{ij} при верном H_0 k – число вариантов СВ X l – число вариантов СВ Y
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2((k-1)(l-1))$	
Формула расчета критической точки	$z^* = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

б) Построить эмпирическую таблицу сопряжённости

$x \backslash y$	Y	N	Σ
north	12	29	41
south	10	35	45
other	5	26	31
Σ	27	90	117

в) Построить теоретическую таблицу сопряжённости

$x \backslash y$	Y	N	Σ
north	9.462	31.538	41
south	10.385	34.615	45
other	7.154	23.846	31
Σ	27	90	117

г) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	1.747	0.418	H_0 принимается	$F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$
0.05			H_0 принимается	$F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$
0.1			H_0 принимается	$F_Y(y x_1) = F_Y(y x_2) = F_Y(y x_3)$

Вывод (в терминах предметной области)

В результате проведённого в п.6 статистического анализа обнаружено, что нет оснований отвергать гипотезу о независимости распределения углового расположения домов от сектора, в котором эти дома расположены

7. Дисперсионный анализ

Факторный признак x – Location in sector of city (D7)

Результативный признак y – Nitric oxides concentration (D9)

Число вариантов факторного признака – 3

Объёмы выборок – 117

Статистическая гипотеза – $H_0: m_{Y|X=x_1} = m_{Y|X=x_2} = m_{Y|X=x_3}$

а) Рассчитать групповые выборочные характеристики

№ п/п	Вариант факторного признака	Объём выборки	Групповые средние	Групповые дисперсии
1	north	41	6.117	0.186
2	south	45	6.178	0.145
3	other	31	6.453	0.285

б) Привести формулы расчёта показателей вариации, используемых в дисперсионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	$D_b^* = \frac{1}{n} \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$	$k - 1$	$\frac{n}{k - 1} D_b^*$
Остаточные признаки	$D_\omega^* = \frac{1}{n} \sum_{i=1}^k n_i \tilde{\sigma}_i^2$ $\tilde{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n - k$	$\frac{n}{n - k} D_\omega^*$
Все признаки	$D_X^* = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n - 1$	$\frac{n}{n - 1} D_X^*$

в) Рассчитать показатели вариации, используемые в дисперсионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	0.0187	2	1.093
Остаточные признаки	0.1966	114	0.202
Все признаки	0.215	116	0.217

г) Проверить правило сложения дисперсий

Показатель	$D_{\text{межгр}}$	$D_{\text{внутригр}}$	$D_{\text{общ}}$	$D_{\text{межгр}} + D_{\text{внутригр}}$
Значение	0.0187	0.1966	0.215	0.2153

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Эмпирический коэффициент детерминации	$\tilde{\eta}^2 = \frac{D_b^*}{D_x^*}$	0.087
Эмпирическое корреляционное отношение	$\tilde{\eta} = \sqrt{\frac{D_b^*}{D_x^*}}$	0.295

е) Охарактеризовать тип связи между факторным и результативным признаками

Связь между результативным и факторным признаком слабая

ж) Указать формулы расчёта показателей, используемых при проверке статистической гипотезы дисперсионного анализа

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{\frac{D_b^*}{(k-1)}}{\frac{D_w^*}{(n-k)}}$	D_b^* - выборочная межгрупповая дисперсия D_w^* - выборочная внутригрупповая дисперсия k – количество групп n – объем выборки

Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim F(k-1, n-k)$	
Формула расчета критической точки	$z' = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

з) Проверить статистическую гипотезу дисперсионного анализа

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	5.413	0.006	H_0 отклоняется	$F_Y(y x_1) \neq F_Y(y x_2) \neq F_Y(y x_3)$
0.05			H_0 отклоняется	$F_Y(y x_1) \neq F_Y(y x_2) \neq F_Y(y x_3)$
0.1			H_0 отклоняется	$F_Y(y x_1) \neq F_Y(y x_2) \neq F_Y(y x_3)$

Вывод (в терминах предметной области)

В результате проведенного в п.7 статистического анализа обнаружено, что гипотеза о зависимости распределения концентрации оксидов азота от расположения в разных секторах города отвергается

8. Корреляционный анализ

8.1. Расчёт парных коэффициентов корреляции

Анализируемый признак 1 – Appraisal price2 (\$hundreds) (D2)

Анализируемый признак 2 – Nitric oxides concentration (D9)

Объёмы выборок – 117

а) Рассчитать точечные оценки коэффициентов корреляции

	Формула расчета	Значение
Линейный коэффициент корреляции	$\rho_{XY}^* = \frac{cov^*(X, Y)}{\sigma_X^* \sigma_Y^*}$	0.18
Ранговый коэффициент корреляции по Спирмену	$\tilde{\rho}_{XY}^{(сп)} = \frac{6 \sum_{i=1}^n (r_i - s_i)^2}{n(n^2 - 1)}$	0.029
Ранговый коэффициент корреляции по Кендаллу	$\tilde{\tau}_{XY} = \frac{N^+ + N^-}{\frac{1}{2}n(n-1)}$	0.015

б) Привести формулы расчёта доверительного интервала для линейного коэффициента корреляции

Граница доверительного интервала	Формула расчета
Нижняя граница	$\rho_{XY}^* + \frac{\rho_{XY}^*(1 - (\rho_{XY}^*)^2)}{2n} - u_{1-\frac{\alpha}{2}} \frac{1 - (\rho_{XY}^*)^2}{\sqrt{n}}$
Верхняя граница	$\rho_{XY}^* + \frac{\rho_{XY}^*(1 - (\rho_{XY}^*)^2)}{2n} + u_{1-\frac{\alpha}{2}} \frac{1 - (\rho_{XY}^*)^2}{\sqrt{n}}$

в) Рассчитать доверительные интервалы для линейного коэффициента корреляции

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	-0.049	0.006	0.034
Верхняя граница	0.411	0.356	0.328

г) Указать формулы расчёта показателей, используемых при проверке значимости коэффициентов корреляции

Статистическая гипотеза	Формула расчета статистики критерия	Закон распределения статистики критерия при условии истинности основной гипотезы
$H_0: \rho = 0$ $H': \rho \neq 0$	$Z = \frac{\rho_{XY}^* \sqrt{n-2}}{\sqrt{1 - (\rho_{XY}^*)^2}}$	$T(n-2)$

$H_0 : r^{(cn)} = 0$ $H' : r^{(cn)} \neq 0$	$Z = \frac{\tilde{\rho}_{xy}^{(cn)} \sqrt{n-2}}{\sqrt{1 - (\tilde{\rho}_{xy}^{(cn)})^2}}$	$T(n-2)$
$H_0 : r^{(кен)} = 0$ $H' : r^{(кен)} \neq 0$	$Z = \tilde{r}_{xy} \sqrt{\frac{9n(n+1)}{2(2n+5)}}$	$N(0,1)$

д) Проверить значимость коэффициентов корреляции

Статистическая гипотеза	Уровень значимости	Выборочное значение статистики критерия	p-value	Статистическое решение	Вывод
$H_0 : \rho = 0$ $H' : \rho \neq 0$	0.1	1.966	0.052	Н ₀ отклоняется	$\rho \neq 0$
$H_0 : r^{(cn)} = 0$ $H' : r^{(cn)} \neq 0$	0.1	0.315	0.753	Н ₀ принимается	$r^{(cn)} = 0$
$H_0 : r^{(кен)} = 0$ $H' : r^{(кен)} \neq 0$	0.1	0.238	0.814	Н ₀ принимается	$r^{(кен)} = 0$

8.2. Расчёт множественных коэффициентов корреляции

Анализируемый признак 1 – Appraisal price2 (\$hundreds) (D2)

Анализируемый признак 2 – Nitric oxides concentration (D9)

Анализируемый признак 3 – Number out of 11 features (dishwasher, refrigerator, microwave, disposer, washer, intercom, skylight(s), compactor, dryer, handicap fit, cable TV access) (D6)

Объёмы выборок – 117

а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу

Признак \ Признак	D2	D9	D6
D2	1	0.015	0.326
D9	0.015	1	0.001
D6	0.326	0.001	1

б) Рассчитать матрицу значений p-value для ранговых коэффициентов корреляции по Кендаллу (статистическая гипотеза $H_0 : r^{(кен)} = 0$, $H' : r^{(кен)} \neq 0$)

Признак \ Признак	D2	D9	D6
D2	–	0.814	0
D9	0.814	–	0.992
D6	0	0.992	–

в) Рассчитать точечную оценку коэффициента конкордации

	Формула расчета	Значение
Коэффициент конкордации	$\bar{W}_{x_0 x_1 \dots x_k} = \frac{12 \sum_{i=1}^n (\sum_{j=1}^k R_{ij} - \bar{R})^2}{(k+1)^2 (n^3 - n)}$	0.236

г) Указать формулы расчёта показателей, используемых при проверке значимости коэффициента конкордации

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{12 \sum_{i=1}^n (\sum_{j=1}^k R_{ij} - \bar{R})^2}{(k+1)(n^2 - n)}$	R_{ij} - ранг элемента x_{ij} \bar{R} - среднеарифметическое из всех рангов k - количество факторов n - объем выборки
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(n-1)$	
Формула расчета критической точки	$z' = z_{1-\alpha}$	α - уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

д) Проверить значимость коэффициента конкордации

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	111.5	0.6	H_0 принимается	Между признаками отсутствует монотонная связь
0.05			H_0 принимается	Между признаками отсутствует монотонная связь

0.1			H_0 принимается	Между признаками отсутствует монотонная связь
-----	--	--	-------------------	---

Вывод (в терминах предметной области)

В результате проведённого в п.8 статистического анализа обнаружено, что гипотеза о монотонной зависимости количества функций в доме от концентрации оксидов азота и оценочной стоимости дома 1 отвергается

9. Регрессионный анализ

9.1 Простейшая линейная регрессионная модель

Факторный признак x – Nitric oxides concentration (D9)

Результативный признак y – Appraisal price1 (\$hundreds) (D1)

Уравнение регрессии – $f(x) = \beta_0 + \beta_1 x$

9.1.1. Точечные оценки линейной регрессионной модели

а) Рассчитать точечные оценки параметров линейной регрессионной модели

Параметр	Формула расчета	Значение
β_0	$\tilde{\beta}_0 = \bar{y} - \rho_{xy}^* \frac{\sigma_Y^*}{\sigma_X^*} \bar{x}$	113.178
β_1	$\tilde{\beta}_1 = \rho_{xy}^* \frac{\sigma_Y^*}{\sigma_X^*}$	152.42

б) Записать точечную оценку уравнения регрессии

$$f(x) = 113.178 + 152.42x$$

в) Привести формулы расчёта показателей вариации, используемых в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	$D_{Y X}^* = \frac{1}{n} \sum_{i=1}^n (f(x_i, \tilde{\beta}_0, \tilde{\beta}_1) - \bar{y})^2$	$k - 1$	$\frac{n}{k - 1} D_{Y X}^*$
Остаточные признаки	$D_{resY}^* = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i, \tilde{\beta}_0, \tilde{\beta}_1))^2$	$n - k$	$\frac{n}{n - k} D_{resY}^*$
Все признаки	$D_Y^* = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$	$\frac{n}{n - 1} D_Y^*$

г) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	5002.097	1	585245.332

Остаточные признаки	138493.175	115	140901.752
Все признаки	143495.272	116	144732.3

д) Проверить правило сложения дисперсий

Показатель	$D_{рег}$	$D_{ост}$	$D_{общ}$	$D_{рег} + D_{ост}$
Значение	5002.097	138493.175	143495.272	143495.272

е) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R^2_{Y X} = \frac{D_{Y X}^*}{D_Y^*}$	0.035
Корреляционное отношение	$R^*_{Y X} = \sqrt{\frac{D_{Y X}^*}{D_Y^*}}$	0.187

ж) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

Наблюдается слабая линейная связь результативного признака от факторного

9.1.2. Интервальные оценки линейной регрессионной модели

а) Привести формулы расчёта доверительных интервалов для параметров линейной регрессионной модели

Параметр	Границы доверительного интервала	Формула расчета
β_0	Нижняя граница	$\bar{\beta}_0 - t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\bar{D}_{resY}^{несмещ}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n^2 D_X^*}}$
	Верхняя граница	$\bar{\beta}_0 + t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\bar{D}_{resY}^{несмещ}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n^2 D_X^*}}$
β_1	Нижняя граница	$\bar{\beta}_1 - t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\bar{D}_{resY}^{несмещ}} \sqrt{\frac{1}{n D_X^*}}$
	Верхняя граница	$\bar{\beta}_1 + t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\bar{D}_{resY}^{несмещ}} \sqrt{\frac{1}{n D_X^*}}$

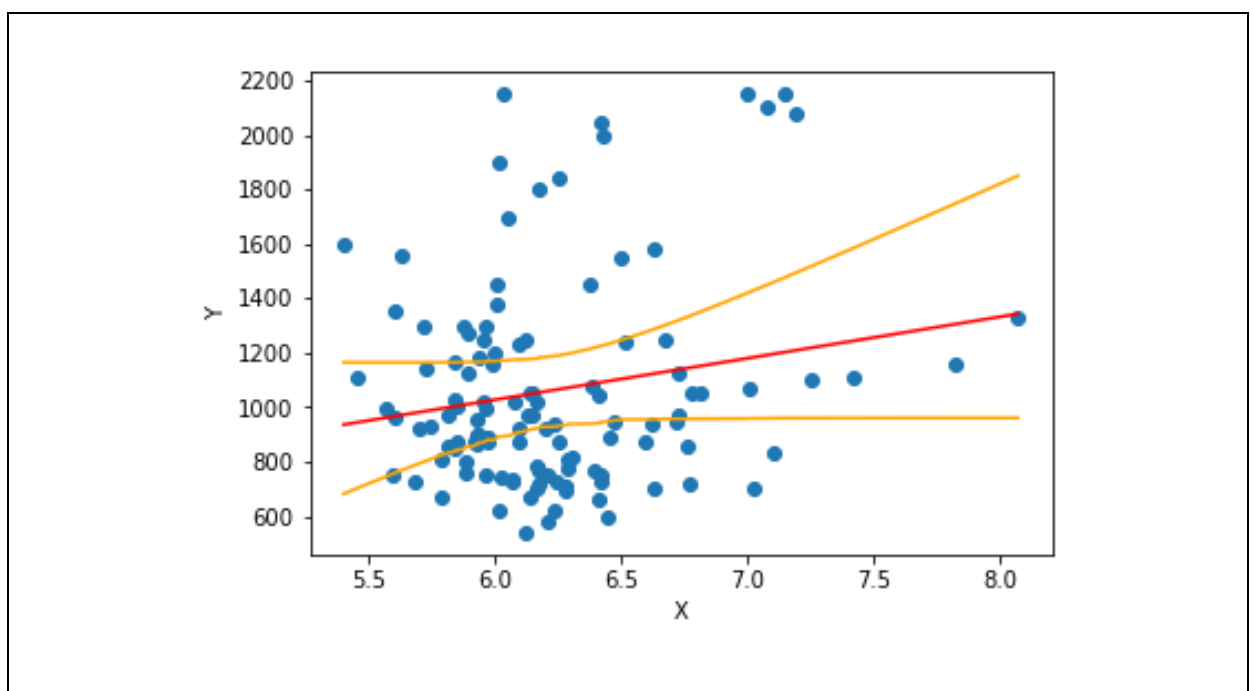
б) Рассчитать доверительные интервалы для параметров линейной регрессионной модели

Параметр	Границы доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
β_0	Нижняя граница	-1110.564	-812.274	-661.554
	Верхняя граница	1336.92	1038.630	887.91
β_1	Нижняя граница	-43.469	4.28	28.406
	Верхняя граница	348.309	300.561	276.434

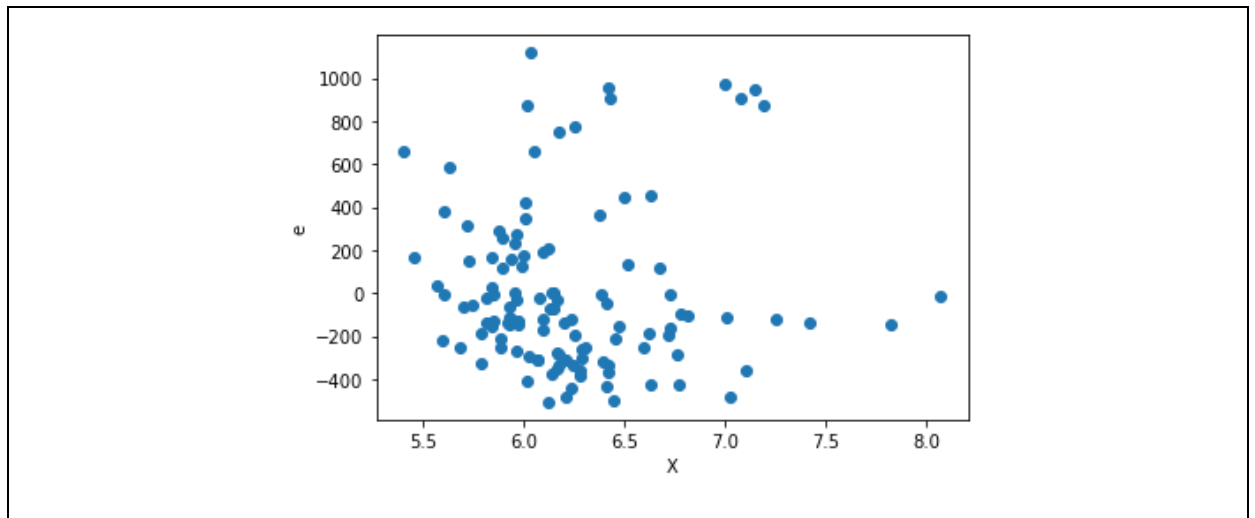
в) Привести формулы расчёта доверительного интервала для значений регрессии $f(x)$

Границы доверительного интервала	Формула расчета
Нижняя граница $f_{low}(x)$	$\tilde{f}(x) - t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\tilde{D}_{resY}^{несмещ} \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{nD_X^*} \right)}$
Верхняя граница $f_{high}(x)$	$\tilde{f}(x) + t_{1-\frac{\alpha}{2}}(n-2) \sqrt{\tilde{D}_{resY}^{несмещ} \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{nD_X^*} \right)}$

г) Построить диаграмму рассеяния признаков x и y . Нанести на диаграмму функцию регрессии $f(x)$, а также нижние и верхние границы линии регрессии $f_{low}(x)$ и $f_{high}(x)$ на уровне значимости $\alpha = 0.1$



д) Построить график остатков $\varepsilon(x) = y - f(x)$



9.1.3. Проверка значимости линейной регрессионной модели

Статистическая гипотеза – $H_0: \beta_1 = 0$
 $H': \beta_1 \neq 0$

а) Указать формулы расчёта показателей, используемых при проверке значимости линейной регрессионной модели

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{D_{\text{регр}Y X}^*}{\frac{D_{\text{рес}Y}^*}{(n-2)}}$	$D_{\text{регр}Y X}^*$ - дисперсия обусловленная регрессией $D_{\text{рес}Y}^*$ - остаточная дисперсия n – объем выборки
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim F(1, n-2)$	
Формула расчета критической точки	$z' = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	p-value	Статистическое решение	Вывод
0.01	4.154	0.04	H ₀ принимается	$\beta_1 = 0$
0.05			H ₀ отклоняется	$\beta_1 \neq 0$
0.1			H ₀ отклоняется	$\beta_1 \neq 0$

9.2 Линейная регрессионная модель общего вида

Факторный признак x – Nitric oxides concentration (D9)

Результативный признак y – Appraisal price1 (\$hundreds) (D1)

Уравнение регрессии – квадратичное по x : $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

9.2.1. Точечные оценки линейной регрессионной модели

а) Рассчитать точечные оценки параметров линейной регрессионной модели

Параметр	Формула расчета	Значение
β_0	$\tilde{\beta} = (F^T F)^{(-1)} F^T y$ $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)^T$ $y = (y_1, \dots, y_n)^T$	5818.8649
β_1		-1610.466
β_2		135.212

б) Записать точечную оценку уравнения регрессии

$$f(x) = 5818.8649 - 1610.466x + 135.212x^2$$

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	7180.065	2	840067.564
Остаточные признаки	136315.185	114	139902.427
Все признаки	143495.272	116	144732.3

г) Проверить правило сложения дисперсий

Показатель	$D_{регр}$	$D_{ост}$	$D_{общ}$	$D_{регр} + D_{ост}$
Значение	7180.065	136315.185	143495.272	143495.25

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R^2_{Y X} = \frac{D_{Y X}^*}{D_Y^*}$	0.05
Корреляционное отношение	$R^*_{Y X} = \sqrt{\frac{D_{Y X}^*}{D_Y^*}}$	0.224

е) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

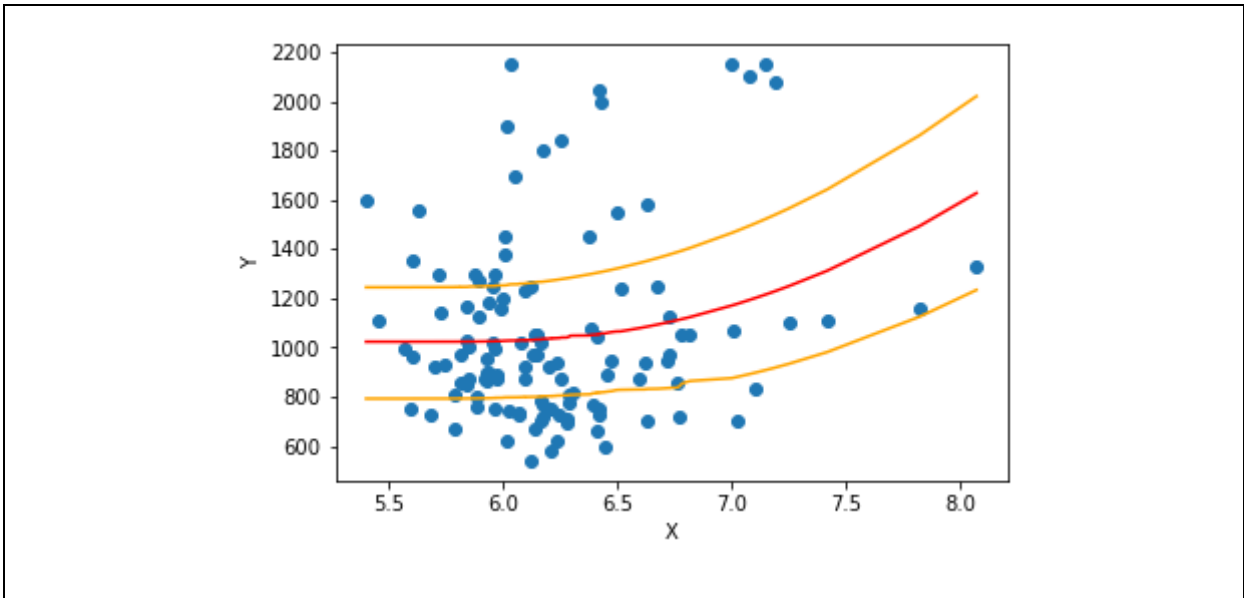
Наблюдается слабая связь результативного признака от факторного

9.2.2. Интервальные оценки линейной регрессионной модели

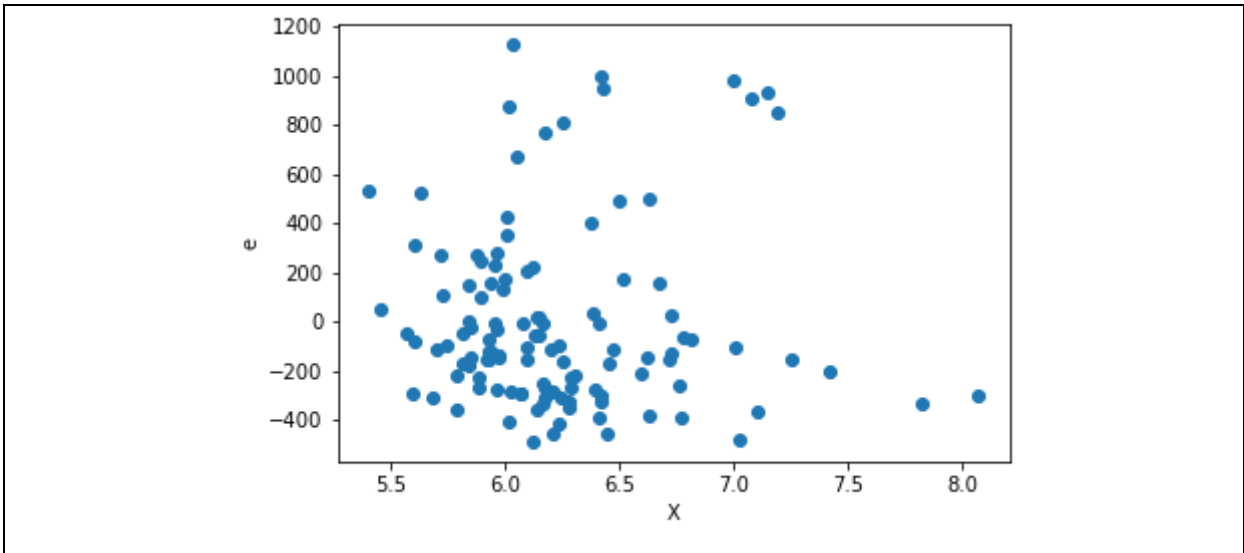
а) Привести формулы расчёта доверительного интервала для значений регрессии $f(x)$

Границы доверительного интервала	Формула расчета
Нижняя граница $f_{low}(x)$	$\tilde{f}(x) - t_{1-\frac{\alpha}{2}}(n-k) \sqrt{\tilde{D}_{resY}} \sqrt{(\varphi^T(x)(F^T F)^{-1} \varphi(x))}$
Верхняя граница $f_{high}(x)$	$\tilde{f}(x) + t_{1-\frac{\alpha}{2}}(n-k) \sqrt{\tilde{D}_{resY}} \sqrt{(\varphi^T(x)(F^T F)^{-1} \varphi(x))}$

б) Построить диаграмму рассеяния признаков x и y . Нанести на диаграмму функцию регрессии $f(x)$, а также нижние и верхние границы линии регрессии $f_{low}(x)$ и $f_{high}(x)$ на уровне значимости $\alpha = 0.1$



в) Построить график остатков $\varepsilon(x) = y - f(x)$



9.2.3. Проверка значимости линейной регрессионной модели

Статистическая гипотеза – $H_0 : \beta_1 = \beta_2 = 0$
 $H' : \text{не } H_0$

а) Указать формулы расчёта показателей, используемых при проверке значимости линейной регрессионной модели

	Выражение	Пояснение использованных обозначений

Формула расчета статистики критерия	$Z = \frac{\frac{D_{\text{регр}Y X}^*}{(k-1)}}{\frac{D_{\text{рес}Y}^*}{(n-k)}}$	$D_{\text{регр}Y X}^*$ - дисперсия обусловленная регрессией $D_{\text{рес}Y}^*$ - остаточная дисперсия n – объем выборки k – количество оцениваемых параметров
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim F(k-1, n-k)$	
Формула расчета критической точки	$z' = z_{1-\alpha}$	α – уровень значимости, критическая область правосторонняя
Формула расчета p -value	$p = 1 - F(Z < z_{\text{выб}} _{H_0})$	$z_{\text{выб}}$ - реализация статистики критерия Z

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	3.002	0.054	H_0 принимается	$\beta_1 = \beta_2 = 0$
0.05			H_0 принимается	$\beta_1 = \beta_2 = 0$
0.1			H_0 отклоняется	Модель значима

9.3 Множественная линейная регрессионная модель

Факторный признак 1 x_1 – Nitric oxides concentration (D9)

Факторный признак 2 x_2 – Index of accessibility to radial highways (D10)

Результативный признак y – Appraisal price1 (\$hundreds) (D1)

Уравнение регрессии – $f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

а) Рассчитать точечные оценки параметров линейной регрессионной модели

Параметр	Формула расчета	Значение
β_0	$\tilde{\beta} = (F^T F)^{(-1)} F^T y$ $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)^T$ $y = (y_1, \dots, y_n)^T$	-8.5389
β_1		148.5189
β_2		30.5576

б) Записать точечную оценку уравнения регрессии

$$f(x) = -8.5389 + 148.5189x_1 + 30.5576x_2$$

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	7434.556	2	869843.063
Остаточные признаки	136060.721	114	139641.266
Все признаки	143495.272	116	144732.3

г) Проверить правило сложения дисперсий

Показатель	$D_{\text{регр}}$	$D_{\text{ост}}$	$D_{\text{общ}}$	$D_{\text{регр}} + D_{\text{ост}}$
Значение	7434.556	136060.721	143495.272	143495.277

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Множественный коэффициент детерминации	$R^2_{Y X} = \frac{D_{Y X}^*}{D_Y^*}$	0.052
Множественное корреляционное отношение	$R^*_{Y X} = \sqrt{\frac{D_{Y X}^*}{D_Y^*}}$	0.228

е) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

Наблюдается слабая связь результативного признака от факторного

9.4. Выводы

а) Сводная таблица показателей вариации для различных регрессионных моделей

Источник вариации	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Факторный признак	5002.097	7180.065	7434.556
Остаточные признаки	138493.175	136315.185	136060.721
Все признаки	143495.272	143495.272	143495.272

б) Сводная таблица свойств различных регрессионных моделей

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	Нет	Нет	нет
Значимость	При $\alpha = 0.05$	При $\alpha = 0.1$	При $\alpha = 0.1$
Адекватность	Нет	Нет	нет
Степень тесноты связи	Слабая	Слабая	Слабая

Вывод (в терминах предметной области)

В результате проведённого в п.9 статистического анализа обнаружено, что:

- 1) линейная корреляционная связь между оценочной стоимостью 1 и концентрацией оксида азота слаба,
- 2) корреляционная связь оценочной стоимости 1 связанная с концентрацией оксида азота и квадратом концентрации слаба
- 3) корреляционная связь оценочной стоимости 1 связанная с концентрацией оксида азота и взвешенным расстоянием до пяти городских центров занятости слаба