

ОТЧЕТ № 1

по теме «Проверка статистических гипотез»

Вариант № 9

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение ($\alpha = 0.1$)	Вывод
4.1	$H_0 : F(x) \square N$	Хи-квадрат	H_0 отклоняется	Распределение А7 не является нормальным
4.2	$H_0 : F(x) \square N$	Харке-Бера	H_0 отклоняется	Распределение А7 не является нормальным
5.1	$H_0: F_X(\xi) = F_Y(\xi)$	знаков	H_0 отклоняется	А7 и А8 имеют различные распределения
5.2	$H_0: F_X(\xi) = F_Y(\xi)$	Хи-квадрат	H_0 принимается	А7 и А8 имеют одно распределение

Выводы:

В результате проведённого в п.4 статистического анализа обнаружено, что средняя заработная плата доцентов, не является нормально распределённой случайной величиной.

В результате проведённого в п.5 статистического анализа обнаружено, что средняя заработная плата доцентов и средняя зарплата всех должностей в ВУЗе являются случайными величинами, имеющими одинаковый закон распределения по критерию Хи-квадрат, но не являются случайными величинами, имеющими одинаковый закон распределения по критерию знаков.

ОТЧЕТ № 2

по теме «Анализ статистических взаимосвязей»

Вариант № 9

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение ($\alpha = 0.1$)	Вывод
6	$H_0: F_Y(y X = x_1) = \dots = F_Y(y X = x_n) = F_Y(y)$ $H': \neg H_0$	Хи-квадрат	H_0 отклоняется	Статистическая связь присутствует
7	$H_0: F_{x_1}(x) = \dots = F_{x_K}(x) = F_X(x)$ ($H_0: m_1 = \dots = m_K$) $H': \neg H_0$	ANOVA	H_0 отклоняется	Статистическая связь присутствует

Выводы:

В результате проведённого в п.6 статистического анализа обнаружено, что тип университета влияет на среднюю зарплату сотрудников университета

В результате проведённого в п.7 статистического анализа обнаружено, что тип университета влияет на среднюю зарплату доцентов университета

ОТЧЕТ № 3

по теме «Основы регрессионного анализа»

Вариант №9

Сводная таблица свойств различных регрессионных моделей:

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	47,5	53,9	97,2
Значимость	да	да	да
Адекватность	да	да	да
Степень тесноты связи	Умеренная	Заметная	Сильная

Выводы:

В результате проведённого в п.8 статистического анализа обнаружено, что между средними зарплатами доцентов университета и всех сотрудников университета есть зависимость. А также, что статистическая связь между средними зарплатами всех профессоров университета, доцентов университета и всех сотрудников университета присутствует.

В результате проведённого в п.9 статистического анализа обнаружено, что количество доцентов, умеренно-заметно влияет на среднюю компенсацию доцентам в университете. Однако средняя зарплата доцентов в университете оказывает сильное воздействие на среднюю компенсацию доцентам в университете.

1. Описательные статистики

1.1. Выборочные характеристики

Анализируемый признак 1 – А5

Анализируемый признак 2 – А7

Анализируемый признак 3 – А8

а) Привести формулы расчёта выборочных характеристик

Выборочная хар-ка	Формула расчета
Объём выборки	$1073 = n$
Среднее	$\frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$
Выборочная дисперсия	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Выборочное среднеквадратическое отклонение	$\sigma_X^* = \sqrt{d_X^*}$
Выборочный коэффициент асимметрии	$\gamma_X^* = \frac{\mu_3^*}{(\sigma_X^*)^3}$
Выборочный эксцесс	$\epsilon_X^* = \frac{\mu_4^*}{(\sigma_X^*)^4} - 3$

б) Рассчитать выборочные характеристики

Выборочная хар-ка	Признак 1	Признак 2	Признак 3
Среднее	526.4836905871389	355.08853681267476	428.0326188257223
Выборочная дисперсия	13868.856310247456	2904.2897285473846	8217.62113651222
Выборочное среднеквадратическое отклонение	117,7661085	53,89146248	90,65109562
Выборочный коэффициент асимметрии	0.6786486256734204	0.44737126211025235	0.8191411695756438
Выборочный эксцесс	0.5304432832288937	0.4319131904766307	0.9757478446505425

1.2. Группировка и гистограммы частот

Анализируемый признак – А7

Объём выборки – 1073

а) Выбрать число групп

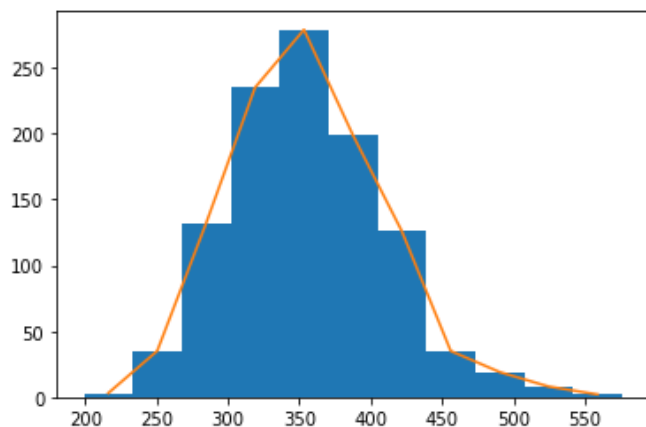
Число групп	Обоснование выбора числа групп	Ширина интервалов
11	По формуле Стерджесса: $k = \lceil 1 + \log_2 n \rceil$	34

б) Построить таблицу частот

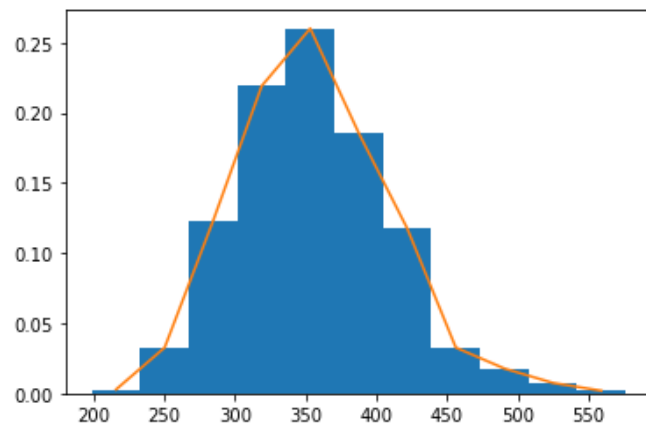
Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Накопл. частота	Относит. накопл. частота
1	199	233.27272727	3	0.0027959	3	0.0027959
2	233.27272727	267.54545455	35	0.03261883	38	0.03541473
3	267.54545455	301.81818182	132	0.12301957	170	0.1584343
4	301.81818182	336.09090909	235	0.21901212	405	0.37744641
5	336.09090909	370.36363636	279	0.26001864	684	0.63746505
6	370.36363636	404.63636364	199	0.18546132	883	0.82292637
7	404.63636364	438.90909091	126	0.11742777	1009	0.94035415
8	438.90909091	473.18181818	35	0.03261883	1044	0.97297297
9	473.18181818	507.45454545	19	0.01770736	1063	0.99068034
10	507.45454545	541.72727273	8	0.00745573	1071	0.99813607
11	541.72727273	576	2	0.00186393	1073	1

в) Построить гистограммы частот и полигоны частот

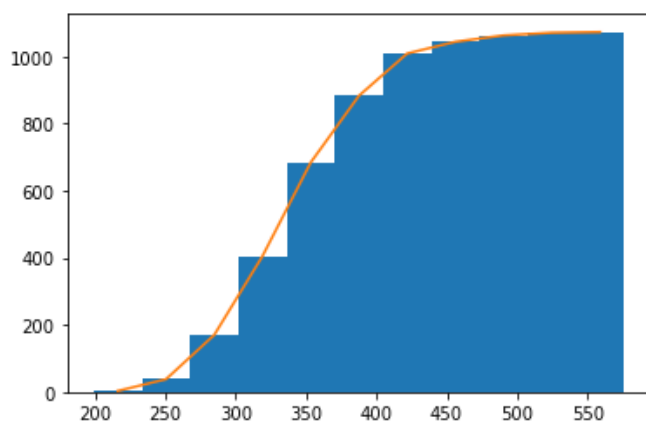
Гистограмма и полигон частот



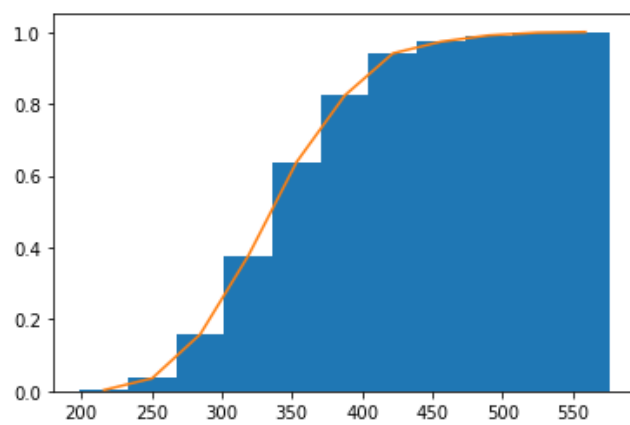
Гистограмма и полигон относительных частот



Гистограмма и полигон накопленных частот

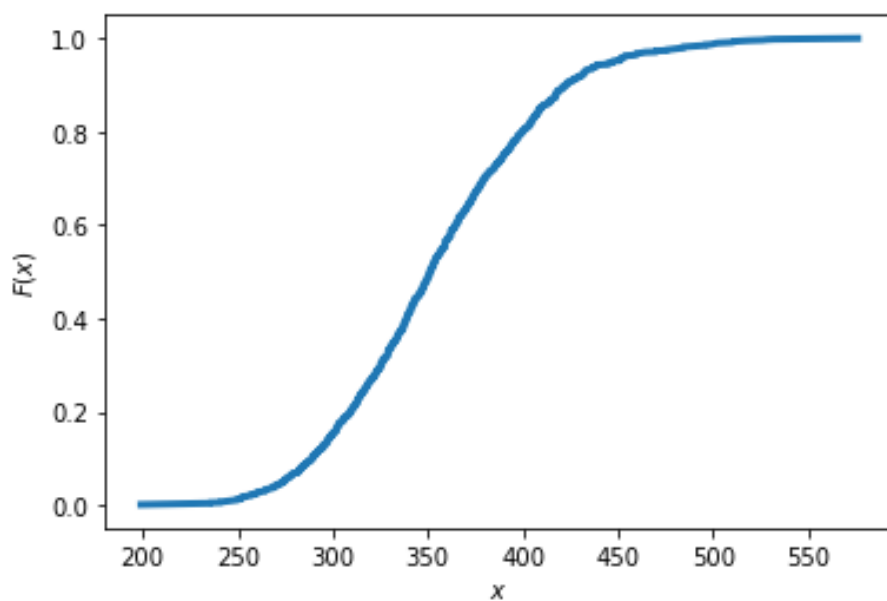


Гистограмма и полигон накопленных относительных частот



г) Построить график эмпирической функции распределения

Эмпирическая функция распределения



2. Интервальные оценки

2.1. Доверительные интервалы для мат. ожидания

Анализируемый признак – А7

Объём выборки – 1073

Оцениваемый параметр – математическое ожидание

а) Привести формулы расчёта доверительных интервалов

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

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	350.8451892973075	351.8618534618906	352.3813366740001
Верхняя граница	359.331884328042	358.31522016345895	357.7957369513494

2.2. Доверительные интервалы для дисперсии

Анализируемый признак – А7

Объём выборки – 1073

Оцениваемый параметр – дисперсия

а) Привести формулы расчёта доверительных интервалов

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

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	2602.8759858554313	2670.7979435901934	2706.469323945736
Верхняя граница	3251.996008997353	3163.787823996446	3119.8827579119193

2.3. Доверительные интервалы для разности мат. ожиданий

Анализируемый признак 1 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

Оцениваемый параметр – математическое ожидание

а) Привести формулы расчёта доверительных интервалов

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

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	-81.24051183093	-79.25482812572	-78.2395253447
Верхняя граница	-64.64765219515	-66.63333590036	-67.6486386813

2.4. Доверительные интервалы для отношения дисперсий

Анализируемый признак 1 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

Оцениваемый параметр – дисперсия

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$\frac{S_1^2}{S_2^2} f_{\alpha/2}(n_2 - 1, n_1 - 1)$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
Верхняя граница	$\frac{S_1^2}{S_2^2} f_{1-\alpha/2}(n_2 - 1, n_1 - 1)$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	0.30193194744336704	0.3135239713241961	0.3196233807857685
Верхняя граница	0.4136933883736934	0.3983977680193247	0.39079509793382433

3. Проверка статистических гипотез о математических ожиданиях и дисперсиях

3.1. Проверка статистических гипотез о математических ожиданиях

Анализируемый признак – А7

Объём выборки – 1073

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	$Z = \frac{\bar{X} - m_0}{S/\sqrt{n}}$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim T(n-1)$
Формулы расчета критических точек	$\pm t_{1-\alpha/2}(n-1)$
Формула расчета p -value	$2 \min(F_Z(z H_0), 1 - F_Z(z H_0))$

б) Выбрать произвольные значения m_0 и проверить статистические гипотезы

m_0	Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
355	0.1	0.053815053313224996	0.957092542957	H_0 принимается	$m = 355$
77	0.1	169.02968361133475	0.0	H_0 отклоняется	$m \neq 77$
390	0.1	-21.22012523276862	1.03637924e-83	H_0 отклоняется	$m \neq 390$

3.2. Проверка статистических гипотез о дисперсиях

Анализируемый признак – А7

Объём выборки – 1073

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	$Z = \frac{(n-1)S^2}{\sigma_0^2}$ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(n-1)$
Формулы расчета критических точек	$\chi^2_{\alpha/2}(n-1), \chi^2_{1-\alpha/2}(n-1)$
Формула расчета p -value	$2\min(F_Z(z H_0), 1 - F_Z(z H_0))$

б) Выбрать произвольные значения σ_0 и проверить статистические гипотезы

σ_0	Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
54	0.1	1067.694989369957	0.93727479152404	H_0 принимается	$\sigma = 54$
340	0.1	26.932513745698916	0.0	H_0 отклоняется	$\sigma \neq 340$
70	0.1	635.3874671434274	9.5075382099e-29	H_0 отклоняется	$\sigma \neq 70$

3.3. Проверка статистических гипотез о равенстве математических ожиданий

Анализируемый признак 1 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

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	<p>Для более точной проверки статистической гипотезы, поставленной в этой задаче, необходимо сначала проверить гипотезу пункта 3.4:</p> <p>1) Если $\sigma_1 = \sigma_2$:</p>

	$Z = \frac{\bar{X}_1 - \bar{X}_2}{S / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ <p>2) Если $\sigma_1 \neq \sigma_2$:</p> $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$
Закон распределения статистики критерия при условии истинности основной гипотезы	<p>1)</p> $Z _{H_0} \sim T(n_1 + n_2 - 2)$ <p>2)</p> $Z _{H_0} \sim T([1 / k])$ $k = \frac{\left(\frac{S_1^2/n_1}{S_1^2/n_1 + S_2^2/n_2}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2/n_2}{S_1^2/n_1 + S_2^2/n_2}\right)^2}{n_2 - 1}$
Формулы расчета критических точек	<p>1) $\pm t_{1-\alpha}(n_1 + n_2 - 2)$</p> <p>2) $\pm t_{1-\alpha}([1 / k])$</p>
Формула расчета p -value	$2 \min(F_Z(z H_0), 1 - F_Z(z H_0))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	-22.65688597773376	7.72831783303963e-100	H_0 отклоняется	$m_1 \neq m_2$
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 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

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	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ $Z = \frac{S_1^2}{S_2^2}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim F(n_1 - 1, n_2 - 1)$
Формулы расчета критических точек	$f_{\alpha/2}(n_1-1, n_2-1); f_{1-\alpha/2}(n_1-1, n_2-1)$
Формула расчета p -value	$2 \min(F_Z(z H_0), 1 - F_Z(z H_0))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	0.3534221985050097	2.5081268589815e-62	H_0 отклоняется	$\sigma_1 \neq \sigma_2$
0.05			H_0 отклоняется	$\sigma_1 \neq \sigma_2$
0.1			H_0 отклоняется	$\sigma_1 \neq \sigma_2$

4. Критерии согласия

Анализируемый признак – А7

Объём выборки – 1073

4.1. Критерий хи-квадрат

Теоретическое распределение – нормальное.

Статистическая гипотеза – $H_0 : F(x) \square N$

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = n \sum_{i=1}^k \frac{(\tilde{p}_i - p_i)^2}{p_i} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$	k – число интервалов; n _i – число элементов в i-м интервале; p _i – вероятности попадания в i-й интервал при условии истинности H ₀ ; n – объём выборки; r – число неизвестных параметров распределения.
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(k - r - 1)$	
Формула расчета критической точки	$\chi^2_{1-\alpha}(k - r - 1)$	
Формула расчета p-value	$p\text{-value} = 1 - F_z(z H_0).$	

б) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
11	По формуле Стерджесса: $k = [1 + \log_2 n]$	34

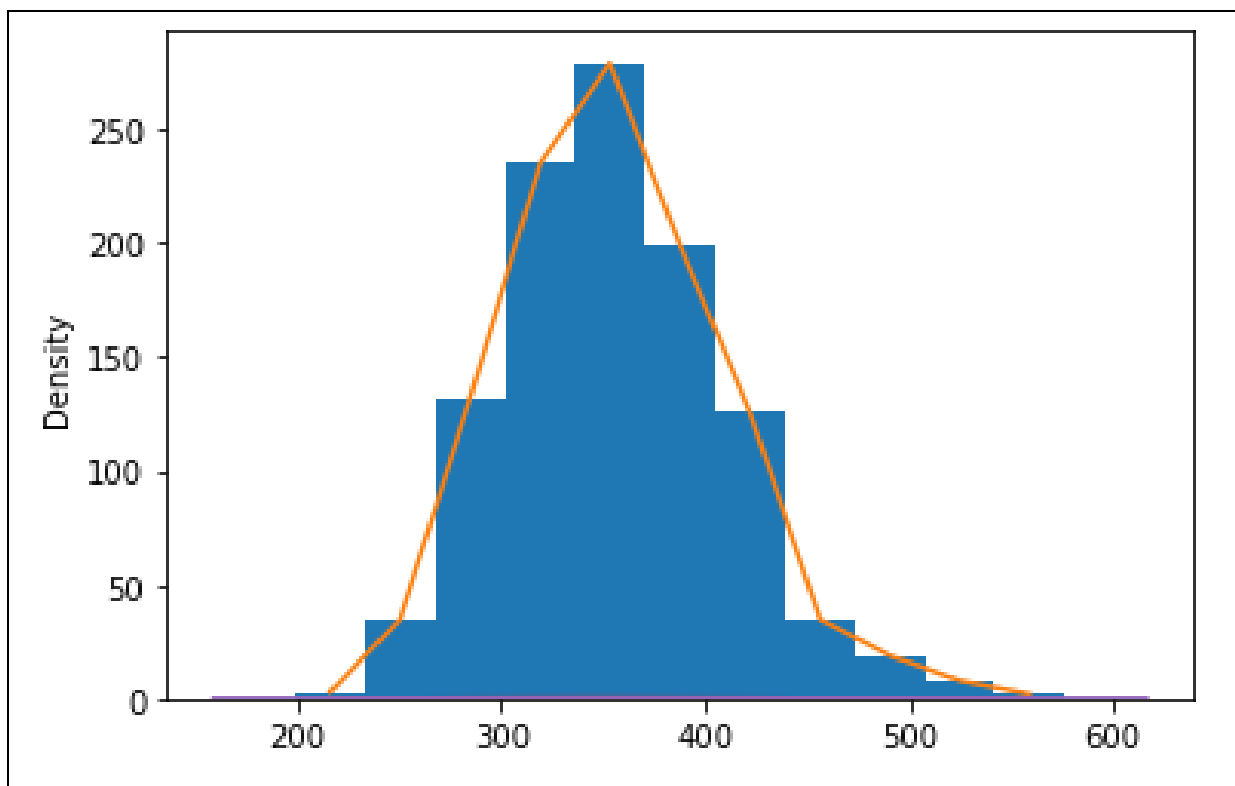
в) Построить таблицу частот

Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Вероятность попадания в интервал при условии истинности основной гипотезы
1	199	233.27272727	3	0.0027959	0.009986290973813804
2	233.27272727	267.54545455	35	0.03261883	0.04019512683994798
3	267.54545455	301.81818182	132	0.12301957	0.10928631759450905
4	301.81818182	336.09090909	235	0.21901212	0.2008165790928036
5	336.09090909	370.36363636	279	0.26001864	0.24946774164480978
6	370.36363636	404.63636364	199	0.18546132	0.20953734111919553
7	404.63636364	438.90909091	126	0.11742777	0.11898699212071995
8	438.90909091	473.18181818	35	0.03261883	0.04566654164424344

9	473.18181818	507.45454545	19	0.01770736	0.01183987292244637
10	507.45454545	541.72727273	8	0.00745573	0.0020723848098110897
11	541.72727273	576	2	0.00186393	0.0002447056549627069

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	47.70428870305429	1.12524314113663e-07	H_0 отклоняется	Распределение не является нормальным
0.05			H_0 отклоняется	Распределение не является нормальным
0.1			H_0 отклоняется	Распределение не является нормальным

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

Статистическая гипотеза – $H_0 : F(x) \square N$

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \left(\gamma_{\text{ст}}^*\right)^2 + \left(\varepsilon_{\text{ст}}^*\right)^2 = \frac{n}{6} \left((\gamma^*)^2 + \frac{(\varepsilon^*)^2}{4} \right),$	<p>n – объем выборки;</p> <p>$\gamma = \mu_3 / \sigma^3$ – коэффициент асимметрии;</p> <p>$\varepsilon = \mu_4 / \sigma^4 - 3$ – коэффициент эксцесса;</p> <p>μ_i = i-й центральный момент;</p> <p>σ – среднее квадратичное отклонение (корень из второго центрального момента).</p>
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(2)$	
Формула расчета критической точки	$\chi^2_{1-\alpha}(2)$	
Формула расчета <i>p-value</i>	$p\text{-value} = 1 - F_Z(z H_0).$	

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	<i>p-value</i>	Статистическое решение	Вывод
0.01	44.13218548064099	2.61106469e-10	H_0 отклоняется	Распределение не является нормальным
0.05			H_0 отклоняется	Распределение не является нормальным
0.1			H_0 отклоняется	Распределение не является нормальным

Вывод (в терминах предметной области)

В результате проведенного в п.4 статистического анализа обнаружено, что средняя заработная плата доцентов, не является нормально распределённой случайной величиной.

5. Проверка однородности выборок

Анализируемый признак 1 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

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	$Z = \frac{H - 1/2}{\sqrt{\frac{1}{4n}}} = 2\sqrt{n}(H - 1/2)$	$H = K/n$ – частота успеха; K – число знаков «+» в последовательности знаков разностей $X_i - Y_i$; n – объём выборок.
Закон распределения статистики критерия при условии истинности основной гипотезы	$f_Z(z H_0) \sim N(0,1)$	
Формула расчета критической точки	$\pm N_{1-\alpha/2}(0,1)$	
Формула расчета p -value	$2\min(F_Z(z H_0), 1 - F_Z(z H_0))$	

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	-533.5	4.069063355e-315	H_0 отклоняется	$F_X(\xi) \neq F_Y(\xi)$
0.05			H_0 отклоняется	$F_X(\xi) \neq F_Y(\xi)$
0.1			H_0 отклоняется	$F_X(\xi) \neq F_Y(\xi)$

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	$Z_{n_X, n_Y} = n_X n_Y \sum_{i=1}^k \frac{1}{m_i^{(X)} + m_i^{(Y)}} \left(\frac{m_i^{(X)}}{n_X} - \frac{m_i^{(Y)}}{n_Y} \right)^2$	k – число интервалов; n_X, n_Y – число элементов в выборках; $m_i^{(X)}, m_i^{(Y)}$ – частота 1 и 2 выборок в i -й группе.
Закон распределения статистики критерия при условии истинности основной гипотезы	$Z _{H_0} \sim \chi^2(k-1)$	
Формула расчета критической точки	$\chi^2_{1-\alpha}(k-1)$	
Формула расчета p -value	$p\text{-value} = 1 - F_Z(z H_0).$	

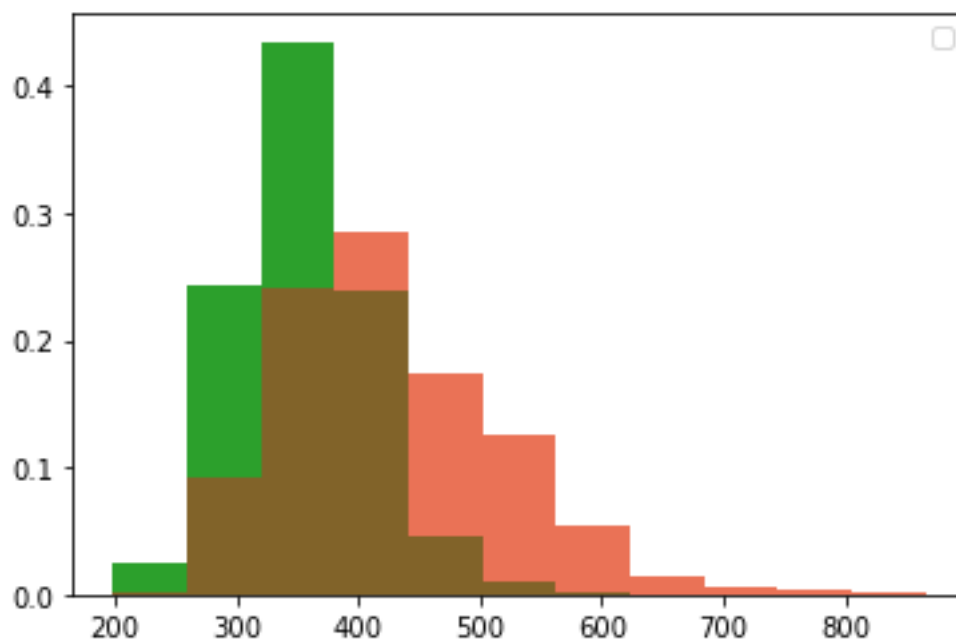
б) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
11	По формуле Стерджесса: $k = [1 + \log 2n]$	34

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	199	259.63636364	27	2	0.02516309	0.00186393
2	259.63636364	320.27272727	262	98	0.24417521	0.09133271
3	320.27272727	380.90909091	467	260	0.43522833	0.24231128
4	380.90909091	441.54545455	256	306	0.23858341	0.28518173
5	441.54545455	502.18181818	50	187	0.04659832	0.17427773
6	502.18181818	562.81818182	10	134	0.00931966	0.1248835
7	562.81818182	623.45454545	1	59	0.00093197	0.05498602
8	623.45454545	684.09090909	0	15	0	0.0139795
9	684.09090909	744.72727273	0	7	0	0.00652377
10	744.72727273	805.36363636	0	3	0	0.0027959
11	805.36363636	866	0	2	0	0.00186393

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	0.9870036925933467	0.9983291620675285	H_0 принимается	$F_X(\xi) = F_Y(\xi)$
0.05			H_0 принимается	$F_X(\xi) = F_Y(\xi)$
0.1			H_0 отклоняется	$F_X(\xi) \neq F_Y(\xi)$

Вывод (в терминах предметной области)

В результате проведённого в п.5 статистического анализа обнаружено, что средняя заработная плата доцентов и средняя зарплата всех должностей в ВУЗе являются случайными величинами, имеющими одинаковый закон распределения по критерию Хи-квадрат, но не являются случайными величинами, имеющими одинаковый закон распределения по критерию знаков.

6. Таблицы сопряжённости

Факторный признак x – А4

Результативный признак y – А20

Объёмы выборок – 1073

Статистическая гипотеза – $H_0: F_Y(y|X = x_1) = \dots = F_Y(y|X = x_n) = F_Y(y)$

$$H': \neg H_0$$

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} – наблюдаемые (эмпирические) частоты, m_{ij} – теоретические частоты, k – число вариантов факторного признака, l – число вариантов результативного признака.
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2((k-1)(l-1))$	
Формула расчета критической точки	$\chi^2_{1-\alpha}((k-1)(l-1))$	
Формула расчета p -value	$p\text{-value} = 1 - F_z(Z H_0)$	

б) Построить эмпирическую таблицу сопряжённости

$x \backslash y$	all-college average (N)	all-college average (Y)	Σ
Type I	6	174	180
IIA	207	152	359
IIB	525	9	534
Σ	738	335	1073

в) Построить теоретическую таблицу сопряжённости

$x \backslash y$	all-college average (N)	all-college average (Y)	Σ
Type I	123.80242311	56.19757689	180
IIA	246.91705499	112.08294501	359
IIB	367.2805219	166.7194781	534
Σ	738	335	1073

г) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	596.6359677097043	1.2398987827376158e-125	H_0 отклоняется	Статистическая связь присутствует
0.05			H_0 отклоняется	Статистическая связь присутствует
0.1			H_0 отклоняется	Статистическая связь присутствует

Вывод (в терминах предметной области)

В результате проведённого в п.6 статистического анализа обнаружено, что тип университета влияет на среднюю зарплату сотрудников университета

7. Дисперсионный анализ

Факторный признак x – А4

Результативный признак y – А7

Число вариантов факторного признака – 3

Объёмы выборок – 1073

Статистическая гипотеза – $H_0: F_{x_1}(x) = \dots = F_{x_K}(x) = F_X(x)$

эквивалентная гипотеза – $H_0: m_1 = \dots = m_K$

$H': \neg H_0$

а) Рассчитать групповые выборочные характеристики

№ п/п	Вариант факторного признака	Объём выборки	Групповые средние	Групповые дисперсии
1	I	180	412.711111111111	1825.4021104903788
2	IIA	359	369.050139275766	1787.9974790308277
3	IIB	534	326.2790262172285	1943.135881976798

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

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	$D_b^* = \frac{1}{n} \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^2$	$K - 1$	$\frac{n}{K-1} D_b^*$
Остаточные признаки	$D_w^* = \frac{1}{n} \sum_{k=1}^K n_k \bar{\sigma}_k^2$	$n - K$	$\frac{n}{n-K} D_w^*$
Все признаки	$D_X^* = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n_k} (x_i^{(k)} - \bar{x})^2$	$n - 1$	$\frac{n}{n-1} D_X^*$

в) Рассчитать показатели вариации, используемые в дисперсионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	1035.2815364756277	2	555428.5443191743
Остаточные признаки	1871.479996119241	1070	1876.7271362952763
Все признаки	2901.583027961599	1072	2904.289728547384

г) Проверить правило сложения дисперсий

Показатель	$D_{\text{межгр}}$	$D_{\text{внутригр}}$	$D_{\text{общ}}$	$D_{\text{межгр}} + D_{\text{внутригр}}$
Значение	1035.281536475627	1871.47999611924	2901.58302796159	2906.7615325948686

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Эмпирический коэффициент детерминации	$\eta^2 = \frac{D_b^*}{D_x^*}$	0.5531886734682816
Эмпирическое корреляционное отношение	$\eta = \sqrt{\frac{D_b^*}{D_x^*}}$	0.7437665450047357

е) Охарактеризовать тип связи между факторным и результативным признаками

По шкале Чеддока значение эмпирического коэффициента корреляции попадает в диапазон 0,5–0,7, характеризующий заметную степень тесноты статистической связи между факторным и результативным признаками.

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$F = \frac{D_b^*/(K-1)}{D_w^*/(n-K)}$	К – количество вариантов номинального группировочного(факторного) признака, n – объем выборки, D_b^* - межгрупповая дисперсия, D_w^* - внутригрупповая дисперсия.
Закон распределения статистики критерия при условии истинности основной гипотезы	$F(K-1, n-K)$	
Формула расчета критической точки	$f_{1-\alpha}(K-1, n-K)$	
Формула расчета p -value	$p\text{-value} = 1 - F_z(z H_0)$	

з) Проверить статистическую гипотезу дисперсионного анализа

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	296.7771416039853	2.921046380968241e-103	H_0 отклоняется	Статистическая связь присутствует
0.05			H_0 отклоняется	Статистическая связь присутствует

0.1			H_0 отклоняется	Статистическая связь присутствует
-----	--	--	-------------------	-----------------------------------

Вывод (в терминах предметной области)

В результате проведённого в п.7 статистического анализа обнаружено, что тип университета влияет на среднюю зарплату доцентов университета

8. Корреляционный анализ

8.1. Расчёт парных коэффициентов корреляции

Анализируемый признак 1 – А7

Анализируемый признак 2 – А8

Объёмы выборок – 1073

а) Рассчитать точечные оценки коэффициентов корреляции

	Формула расчета	Значение
Линейный коэффициент корреляции	$\rho_{XY}^* = \frac{k_{XY}^*}{\sigma_X^* \sigma_Y^*}$ $k_{XY}^* = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$	0.9282016936226042
Ранговый коэффициент корреляции по Спирмену	$\tilde{\rho}_{XY}^{(sp)} = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2 \sum_{i=1}^n (s_i - \bar{s})^2}} = \frac{\mu_{RS}^*}{\sigma_S^* \sigma_R^*}.$ $\bar{r} = \bar{s} = \frac{1}{n} \sum_{i=1}^n i = \frac{n+1}{2}$	0.9376540833607301
Ранговый коэффициент корреляции по Кендаллу	$\tilde{\tau}_{XY} = \frac{4Q}{n(n-1)} - 1,$ $Q = \sum_{i=1}^{n-1} Q_i:$ $Q_i = \sum_{j=i+1}^n [s_j > s_i]$	0.7868780458733714

б) Привести формулы расчёта доверительного интервала для линейного коэффициента корреляции

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

в) Рассчитать доверительные интервалы для линейного коэффициента корреляции

Граница доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Нижняя граница	-2.570055666581273	-1.9141903475724262	-1.5490799899838443
Верхняя граница	2.5900556665812724	2.014190347572426	1.7490799899838445

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
$H_0: \rho = 0$ $H': \rho \neq 0$	$Z = \frac{\rho_{XY}^*}{\sqrt{1 - (\rho_{XY}^*)^2}} \sqrt{n - 2}$	$f_Z(z H_0) \sim T(n - 2)$
$H_0: r^{(cn)} = 0$ $H': r^{(cn)} \neq 0$	$Z = \frac{\tilde{\rho}_{XY}^{(sp)}}{\sqrt{1 - (\tilde{\rho}_{XY}^{(sp)})^2}} \sqrt{n - 2}$	$f_Z(z H_0) \sim T(n - 2)$
$H_0: r^{(кен)} = 0$ $H': r^{(кен)} \neq 0$	$Z = \tilde{\tau}_{XY} \sqrt{\frac{9n(n+1)}{2(2n+5)}}$	$f_Z(z H_0) \sim N(0, 1)$

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	81.64014397824 126	0.0	H_0 отклоняется	Статистическая связь присутствует
$H_0: r^{(cn)} = 0$ $H': r^{(cn)} \neq 0$	0.1	88.28669379903 945	0.0	H_0 отклоняется	Статистическая связь присутствует
$H_0: r^{(кен)} = 0$ $H': r^{(кен)} \neq 0$	0.1	38.63629575753 953	0.0	H_0 отклоняется	Статистическая связь присутствует

8.2. Расчёт множественных коэффициентов корреляции

Анализируемый признак 1 – А5

Анализируемый признак 2 – А7

Анализируемый признак 3 – А8

Объёмы выборок – 1073

а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу

Признак \ Признак	А5	А7	А8
А5	1.000000	0.765362	0.840527
А7	0.765362	1.000000	0.786878
А8	0.840527	0.786878	1.000000

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	A5	A7	A8
A5	–	3.560719e-306	0.0
A7	3.560719e-306	–	0.0
A8	0.000000e+00	0.000000e+00	–

в) Рассчитать точечную оценку коэффициента конкордации

	Формула расчета	Значение
Коэффициент конкордации	$W = \frac{12}{k^2(n^3 - n)} \sum_{i=1}^n \left(\sum_{j=1}^k R_{ij} - \frac{k(n+1)}{2} \right)^2,$ <p>где $R_{ij} \in \{1, \dots, n\}$ - ранг i-го элемента в X_j выборке.</p>	0.9617120846079477

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = n(k-1)W$	<p>n – размер выборки, W – коэффициент конкордации, k – число выборок.</p>
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(n-1)$	
Формула расчета критической точки	$\chi^2_{\alpha}(n-1)$	
Формула расчета p -value	$1 - \chi^2(Z, n-1)$	

д) Проверить значимость коэффициента конкордации

Уровень значимости	Выборочное значение статистики критерия	p -value	Статистическое решение	Вывод
0.01	2063.8341335686555	0.0	H_0 отклоняется	Статистическая связь присутствует
0.05			H_0 отклоняется	Статистическая связь присутствует
0.1			H_0 отклоняется	Статистическая связь присутствует

Вывод (в терминах предметной области)

В результате проведённого в п.8 статистического анализа обнаружено, что между средними зарплатами доцентов университета и всех сотрудников университета есть зависимость. А также, что статистическая связь между средними зарплатами всех профессоров университета, доцентов университета и всех сотрудников университета присутствует.

9. Регрессионный анализ

9.1 Простейшая линейная регрессионная модель

Факторный признак x – A17

Результативный признак y – A12

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	$\tilde{\beta}_0 = \bar{y} - \rho_{XY}^* \frac{\sigma_Y^*}{\sigma_X^*} \bar{x}$	411.4617804233374
β_1	$\tilde{\beta}_1 = \rho_{XY}^* \frac{\sigma_Y^*}{\sigma_X^*}$	0.478340414554699

б) Записать точечную оценку уравнения регрессии

$f(x) = 411.4617804233374 + 0.478340414554699 * x$
--

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

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	$D_{\text{регр } Y X}^* = \frac{1}{n} \sum_{i=1}^n (f(x_i, \beta_0, \dots, \beta_{k-1}) - \bar{y})^2$	$k - 1$	$\frac{n}{k - 1} D_{\text{регр } Y X}^*$
Остаточные признаки	$D_{\text{ост } Y}^* = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i, \beta_0, \dots, \beta_{k-1}))^2$	$n - k$	$\frac{n}{n - k} D_{\text{ост } Y}^*$
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	1246.0556604276064	1	1337017.7236388216
Остаточные признаки	4260.4232730484	1071	4268.37924554709
Все признаки	5506.478933476012	1072	5511.615574272165

д) Проверить правило сложения дисперсий

Показатель	$D_{\text{регр}}$	$D_{\text{ост}}$	$D_{\text{общ}}$	$D_{\text{регр}} + D_{\text{ост}}$
Значение	1246.0556604276064	4260.4232730484	5506.478933476012	5506.478933476006

е) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R_{Y X}^{2*} = \frac{D_{\text{регр } Y X}^*}{D_Y^*} = 1 - \frac{D_{\text{ост } Y}^*}{D_Y^*}$	0.22628900890773465
Корреляционное отношение	$R_{Y X}^* = \sqrt{\frac{D_{\text{регр } Y X}^*}{D_Y^*}} = \sqrt{1 - \frac{D_{\text{ост } Y}^*}{D_Y^*}}$	0.47569844324712124

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

Связь между факторным и результативным признаками является умеренной и не очень сильной.

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	$\tilde{\beta}_0 - t_{1-\alpha/2}(n-2) \sqrt{\tilde{D}_{\text{res}Y}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n^2 D_X^*}}$
	Верхняя граница	$\tilde{\beta}_0 + t_{1-\alpha/2}(n-2) \sqrt{\tilde{D}_{\text{res}Y}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n^2 D_X^*}}$
β_1	Нижняя граница	$\tilde{\beta}_1 - t_{1-\alpha/2}(n-2) \sqrt{\tilde{D}_{\text{res}Y}} \sqrt{\frac{1}{n D_X^*}};$
	Верхняя граница	$\tilde{\beta}_1 + t_{1-\alpha/2}(n-2) \sqrt{\tilde{D}_{\text{res}Y}} \sqrt{\frac{1}{n D_X^*}}$

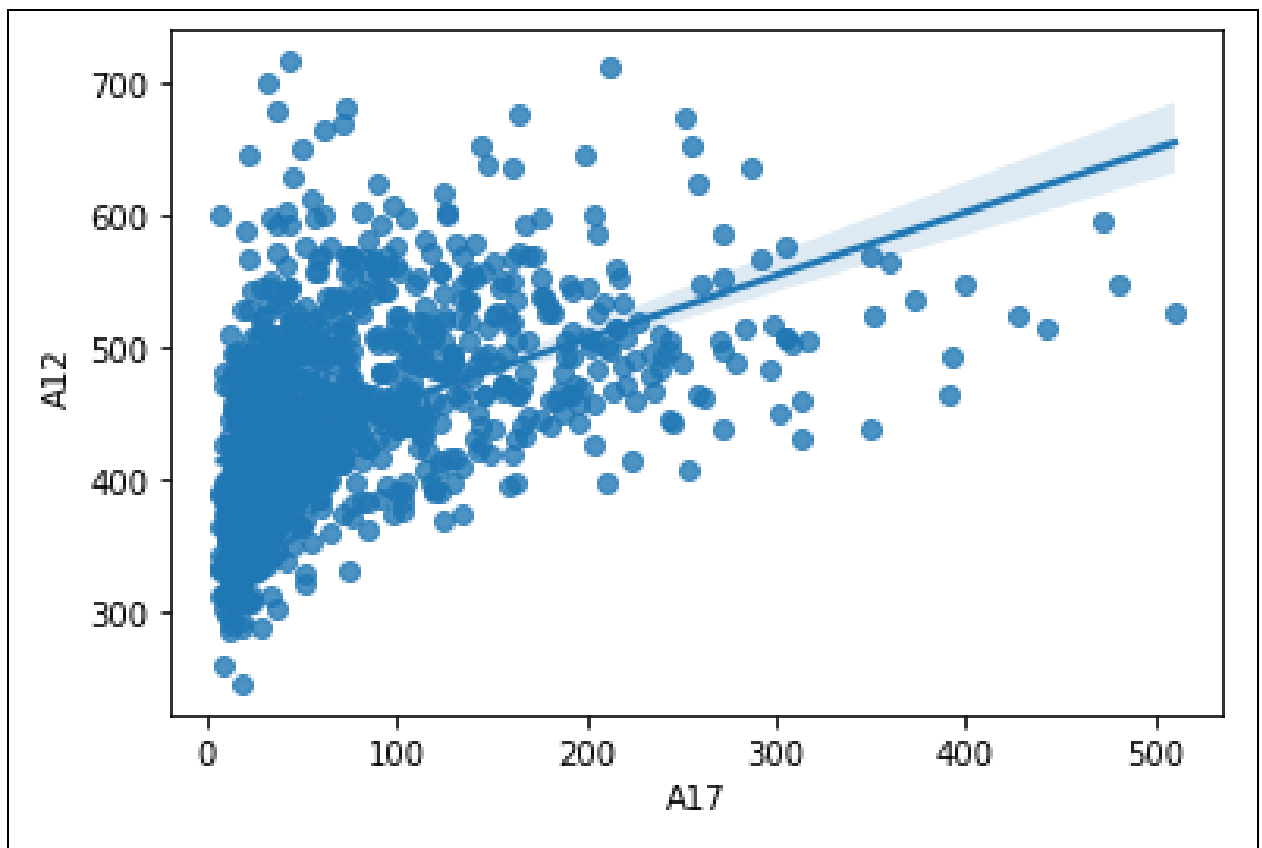
б) Рассчитать доверительные интервалы для параметров линейной регрессионной модели

Параметр	Границы доверительного интервала	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
β_0	Нижняя граница	404.22223024483986	405.9567583386892	406.84304618822824
	Верхняя граница	418.70133060183497	416.9668025079856	416.0805146584466
β_1	Нижняя граница	0.4086963542809441	0.4253824157262198	0.43390845434620484
	Верхняя граница	0.5479844748284539	0.5312984133831782	0.5227723747631932

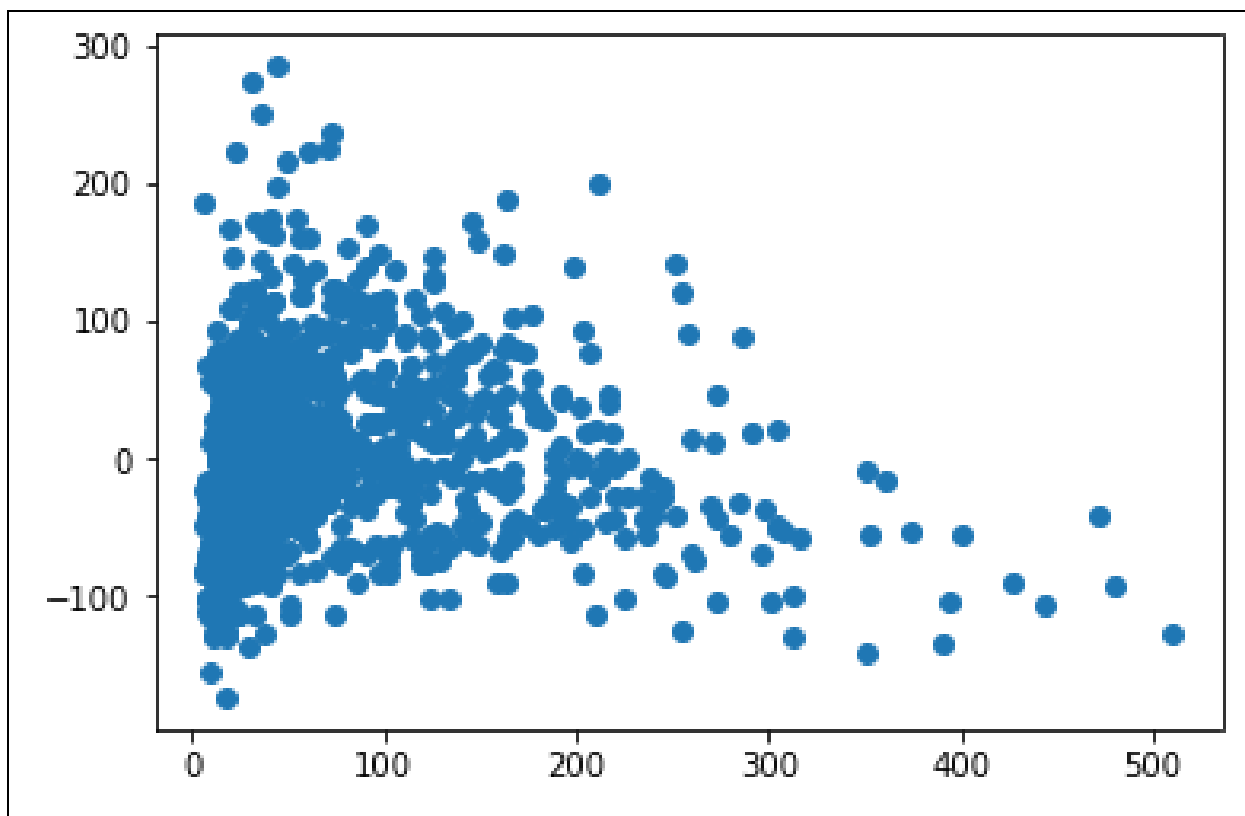
в) Привести формулы расчёта доверительного интервала для значений регрессии $f(x)$

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

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	$Z = \frac{R_{Y X}^{2*}}{(1 - R_{Y X}^{2*})/(n-2)}$	$R_{Y X}^{2*} = \frac{D_{\text{рег } Y X}^*}{D_Y^*};$ n – объем выборки.
Закон распределения статистики критерия при условии истинности основной гипотезы	$f_Z(z H_0) \sim F(1, n-2)$	
Формула расчета критической точки	$f_{1-\alpha}(1, n-2)$	
Формула расчета p -value	$1 - F(Z, 1, n-2)$	

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	p-value	Статистическое решение	Вывод
0.01	313.23779981210424	$1.1102230246251565e-16$	H_0 отклоняется	$\beta_1 \neq 0$
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 – A17

Результативный признак y – A12

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	$\tilde{\beta} = (F^T F)^{-1} F^T y;$ $y = (y_1, \dots, y_n)^T;$ $\beta = (\beta_0, \beta_1, \beta_2)^T;$ $F = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{pmatrix}$	389.1832587353914
β_1		1.0724954869733492
β_2		-0.0019638148003871254

б) Записать точечную оценку уравнения регрессии

$$f(x) = 389.1832587353914 + 1.0724954869733492 * x - 0.0019638148003871254 * x^2$$

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	1599.0702136745588	2	857901.1696364008
Остаточные признаки	3907.4087198014568	1070	3918.364071352302
Все признаки	5506.478933476012	1072	5511.615574272165

г) Проверить правило сложения дисперсий

Показатель	$D_{\text{регр}}$	$D_{\text{ост}}$	$D_{\text{общ}}$	$D_{\text{регр}} + D_{\text{ост}}$
Значение	1599.0702136745588	3907.4087198014568	5506.478933476012	5506.478933476015

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R_{Y X}^{2*} = \frac{D_{\text{регр } Y X}^*}{D_Y^*} = 1 - \frac{D_{\text{ост } Y}^*}{D_Y^*}$	0.29039795357305254
Корреляционное отношение	$R_{Y X}^* = \sqrt{\frac{D_{\text{регр } Y X}^*}{D_Y^*}} = \sqrt{1 - \frac{D_{\text{ост } Y}^*}{D_Y^*}}$	0.5388858446582657

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

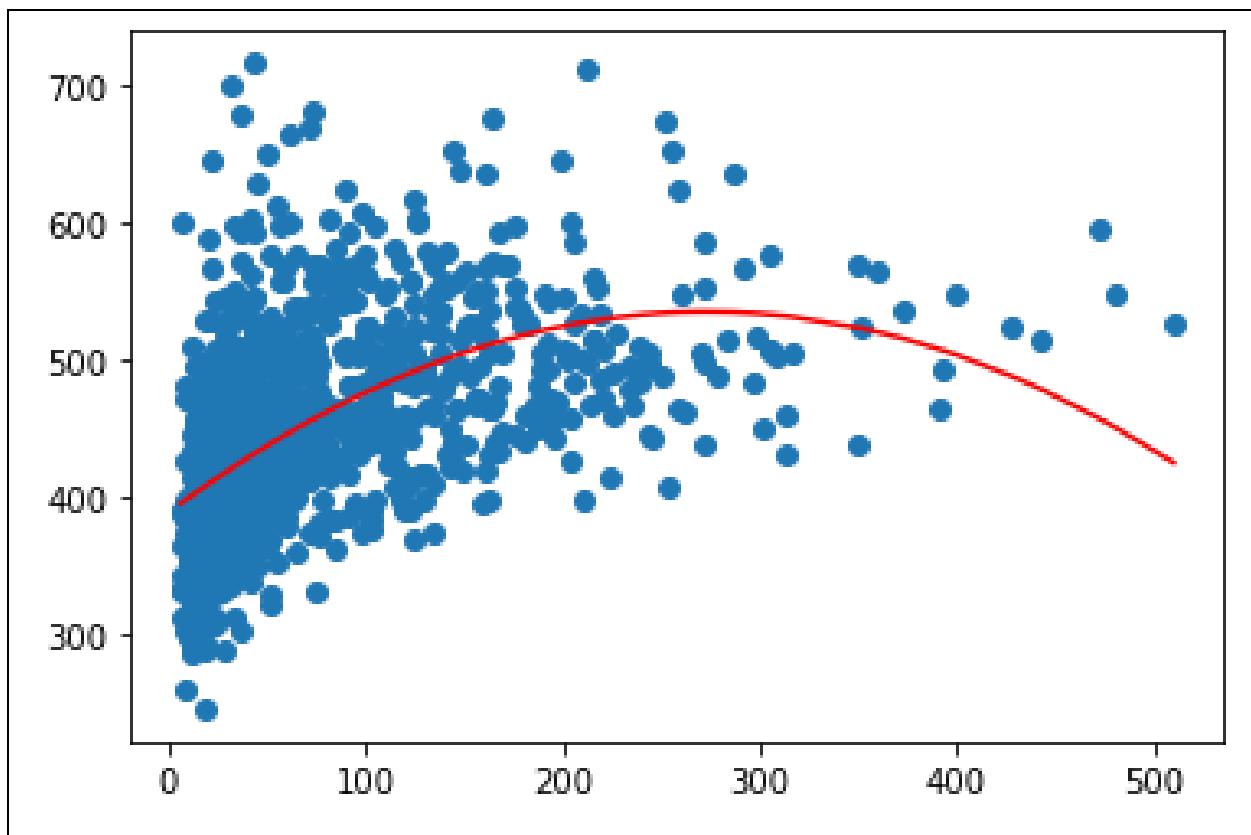
Связь между факторным и результативным признаками заметная.

9.2.2. Интервальные оценки линейной регрессионной модели

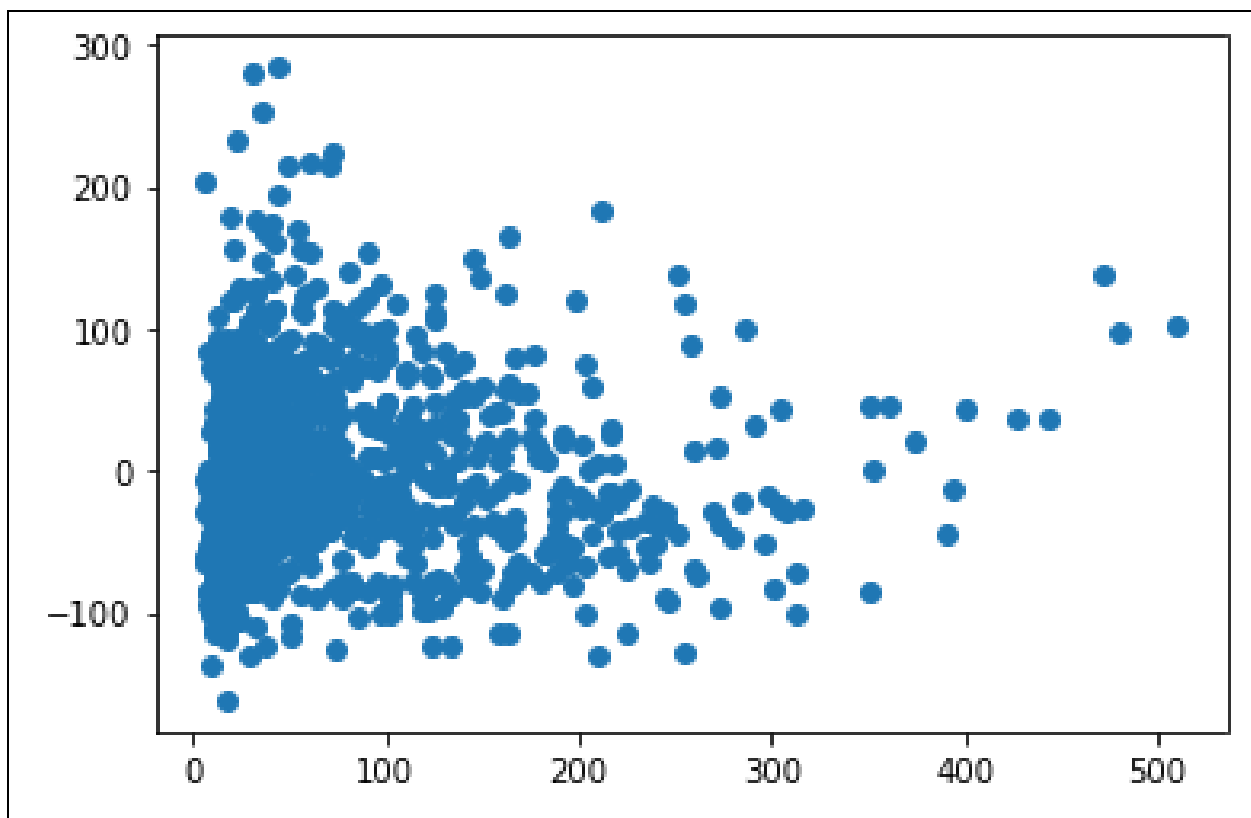
а) Привести формулы расчёта доверительного интервала для значений регрессии $f(x)$

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

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	$Z = \frac{R_{Y X}^{2*}/(k-1)}{(1-R_{Y X}^{2*})/(n-k)}$	$R_{Y X}^{2*} = \frac{D_{\text{рег } Y X}^*}{D_Y^*};$ n – объём выборки; $k = 3$.
Закон распределения статистики критерия при условии истинности основной гипотезы	$F(k-1, n-k)$	
Формула расчета критической точки	$f_{1-\alpha}(k-1, n-k)$	
Формула расчета p-value	$1 - F(Z, k-1, n-k)$	

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	p-value	Статистическое решение	Вывод
0.01	218.94371069514312	1.1102230246251565e-16	H_0 отклоняется	$\beta_1 \neq \beta_2 \neq 0$
0.05			H_0 отклоняется	$\beta_1 \neq \beta_2 \neq 0$
0.1			H_0 отклоняется	$\beta_1 \neq \beta_2 \neq 0$

9.3 Множественная линейная регрессионная модель

Факторный признак 1 x_1 – A17

Факторный признак 2 x_2 – A7

Результативный признак y – A12

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	$\tilde{\beta} = (F^T F)^{-1} F^T y;$ $y = (y_1, \dots, y_n)^T;$ $\beta = (\beta_0, \beta_1, \beta_2)^T;$ $F = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{pmatrix}$	-30.703423147347948
β_1		-0.009874673313251675
β_2		1.3458845906246517

б) Записать точечную оценку уравнения регрессии

$$f(x) = -30.703423147347948 - 0.009874673313251675 * x_1 + 1.3458845906246517 * x_2$$

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	5203.96577259134	2	2791927.636995254
Остаточные признаки	302.5131608846922	1070	303.3613286254904
Все признаки	5506.478933476009	1072	5511.615574272162

г) Проверить правило сложения дисперсий

Показатель	$D_{\text{регр}}$	$D_{\text{ост}}$	$D_{\text{общ}}$	$D_{\text{регр}} + D_{\text{ост}}$
Значение	5203.96577259134	302.5131608846922	5506.478933476009	5506.478933476033

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Множественный коэффициент детерминации	$R_{Y X_1 X_2}^{2*} = \frac{D_{\text{регр}}^* Y X_1 X_2}{D_Y^*} = 1 - \frac{D_{\text{ост}}^* Y}{D_Y^*}$	0.945062323030862
Множественное корреляционное отношение	$R_{Y X_1 X_2}^* = \sqrt{\frac{D_{\text{регр}}^* Y X_1 X_2}{D_Y^*}} = \sqrt{1 - \frac{D_{\text{ост}}^* Y}{D_Y^*}}$	0.9721431597408182

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

Между факторным и результативным признаками присутствует сильная связь.

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	1246.0556604276064	1599.0702136745588	5203.96577259134
Остаточные признаки	4260.4232730484	3907.4087198014568	302.5131608846922
Все признаки	5506.478933476012	5506.478933476012	5506.478933476009

б) Сводная таблица свойств различных регрессионных моделей

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	22.6%	29%	94.5%
Значимость	да	да	да
Адекватность	нет	нет	нет
Степень тесноты связи	Умеренная	Заметная	Сильная

Вывод (в терминах предметной области)

В результате проведенного в п.9 статистического анализа обнаружено, что количество доцентов, умеренно-заметно влияет на среднюю компенсацию доцентам в университете. Однако средняя зарплата доцентов в университете оказывает сильное воздействие на среднюю компенсацию доцентам в университете.