МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
Национальный исследовательский ядерный университет «МИФИ»

Институт интеллектуальных кибернетических систем

КАФЕДРА КИБЕРНЕТИКИ

БДЗ

по курсу "Математическая статистика"

студента группы Б21-524

Розинко Екатерины Дмитриевны

Вариант № 18

Оценка: ________________

Подпись: ________________

2023 г.

ОТЧЕТ № 1

по теме «Проверка статистических гипотез»

Вариант № 18

ФИО студента _Розинко Е.Д._ **группа** _Б21-524

Оценка: ______ **Подпись:** ______

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение (α = 0.1)	Вывод
4.1	$\boldsymbol{H_0}$: $\boldsymbol{F(B8) \sim N}$	Хи-квадрат	H_0 принимается	$\boldsymbol{F}(\mathrm{B}\boldsymbol{8}) \in \boldsymbol{N}$
4.2	$\boldsymbol{H_0}$: $\boldsymbol{F(B8) \sim N}$	Харке-Бера	H_0 принимается	$\boldsymbol{F}(\mathrm{B}\boldsymbol{8}) \in \boldsymbol{N}$
5.1	$\boldsymbol{H_0}$: $\boldsymbol{F_1(B16) = F_2(B18)}$	знаков	H_0 отвергается	$\boldsymbol{F_1}(\mathrm{B}\boldsymbol{16}) \neq \boldsymbol{F_2}(\mathrm{B}\boldsymbol{18})$
5.2	$\boldsymbol{H_0}$: $\boldsymbol{F_1(B16) = F_2(B18)}$	Хи-квадрат	H_0 отвергается	$\boldsymbol{F_1}(\mathrm{B}\boldsymbol{16}) \neq \boldsymbol{F_2}(\mathrm{B}\boldsymbol{18})$

Выводы:

В результате проведённого в п.4 статистического анализа обнаружено, что $\boldsymbol{F}(\mathrm{B}\boldsymbol{8})$ является нормально распределенной величиной.
В результате проведённого в п.5 статистического анализа обнаружено, что выборки *B16* и *B18* неоднородны, т.е имеют разные распределения.

ОТЧЕТ № 2

по теме «Анализ статистических взаимосвязей»

Вариант № 18

ФИО студента _Розинко Е.Д.______________________________ **группа** _Б21-524

Оценка: ___________________________________ **Подпись:**_______________________

Результаты статистических тестов:

№ задания	Проверяемая гипотеза H_0	Критерий	Статистическое решение (α = 0.1)	Вывод
6	$\boldsymbol{H_0}: \boldsymbol{F_Y(y \mid X = x_1)} = \cdots = \boldsymbol{F_Y(y \mid X = x_k)} = \boldsymbol{F_Y(y)}$ y – B6 x – B3	Хи-квадрат	H_0 принимается	B3 не влияет на B6
7	$\boldsymbol{H_0}: \boldsymbol{F_{X_1}(x)} = \ldots = \boldsymbol{F_{X_k}(x)} = \boldsymbol{F_X(x)}$ y – B8 x – B3	ANOVA	H_0 принимается	B3 не влияет на B8

Выводы:

В результате проведённого в п.6 статистического анализа обнаружено, что жировые отложения (В3) не оказывают влияние на то в каком городе проживает человек (В6), т.е между этими признаками не существует статистическая связь.
В результате проведённого в п.7 статистического анализа обнаружено, что жировые отложения (В3) не оказывают влияние на то какого человек роста (В8), т.е между этими признаками не существует статистическая связь.

ОТЧЕТ № 3

по теме «Основы регрессионного анализа»

Вариант № 18

ФИО студента _Розинко Е.Д.______________________________ **группа** _Б21-524

Оценка: ___________________________________ **Подпись:**________________________

Сводная таблица свойств различных регрессионных моделей:

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	49,3%	49,7%	49,49%
Значимость	Да	Да	Да
Адекватность	Да	Да	Да
Степень тесноты связи	слабая	средняя	средняя

Выводы:

В результате проведённого в п.8 статистического анализа обнаружено, что признак B3, B10 и B3 зависимы при уровнях значимости 0.05 и 0.1 и не зависимы при 0.01. Попарно зависимы только B8 и B10.
В результате проведённого в п.9 статистического анализа обнаружено, что все предложенные регрессионные модели адекватно отражают реальную зависимость окружности груди (см) (B10) от их веса (B7) и от их процента жира в организме из уравнения Siri (B2), при этом в данном случае точность простейшей линейной модели и линейной модели с квадратичным членом достаточно низкая, в отличии от множественной линейной модели.

1. Описательные статистики

1.1. Выборочные характеристики

Анализируемый признак 1 – B8 Height (inches)

Анализируемый признак 2 – B10 Chest circumference (cm)

Анализируемый признак 3 – B13 Thigh circumference (cm)

а) Привести формулы расчёта выборочных характеристик

Выборочная хар-ка	Формула расчета
Объём выборки	n
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	$\boldsymbol{\sigma}_X^* = \sqrt{d_X^*}$
Выборочный коэффициент асимметрии	$\boldsymbol{\gamma}_X^* = \frac{\boldsymbol{\mu}_3^*}{(\boldsymbol{\sigma}_X^*)^3}$
Выборочный эксцесс	$\boldsymbol{\varepsilon}_X^* = \frac{\boldsymbol{\mu}_4^*}{(\boldsymbol{\sigma}_X^*)^4} - 3$

б) Рассчитать выборочные характеристики

Выборочная хар-ка	Признак 1	Признак 2	Признак 3
Среднее	70.3107	100.8242	59.4059
Выборочная дисперсия	6.8345	71.0729	27.5619
Выборочное среднеквадратическое отклонение	2.6142	8.4304	5.2499
Выборочный коэффициент асимметрии	0.0982	0.6774	0.8163
Выборочный эксцесс	-0.4284	0.9440	2.5894

1.2. Группировка и гистограммы частот

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

а) Выбрать число групп

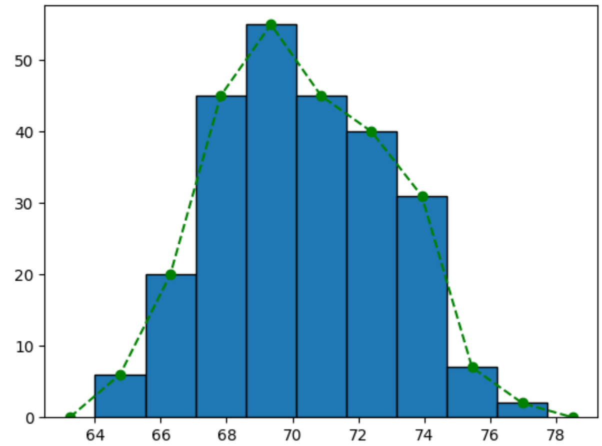
Число групп	Обоснование выбора числа групп	Ширина интервалов
9	Формула Стерджесса: $k = [1 + log_2 n]$	1.53

б) Построить таблицу частот

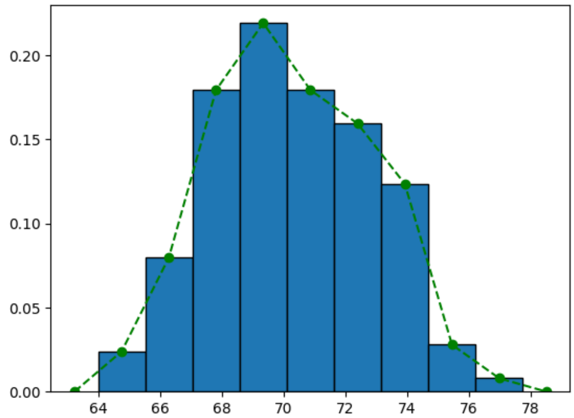
Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Накопл. частота	Относит. накопл. частота
1	64	65.53	6	0.024	6	0.024
2	65.53	67.06	20	0.08	26	0.104
3	67.06	68.58	45	0.179	71	0.283
4	68.58	70.11	55	0.219	126	0.502
5	70.11	71.64	45	0.179	171	0.681
6	71.64	73.17	40	0.159	211	0.84
7	73.17	74.69	31	0.124	242	0.964
8	74.69	76.22	7	0.028	249	0.992
9	76.22	77.75	2	0.008	251	1

в) Построить гистограммы частот и полигоны частот

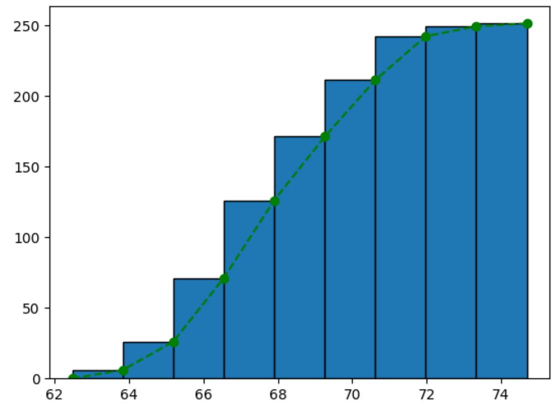
Гистограмма и полигон частот



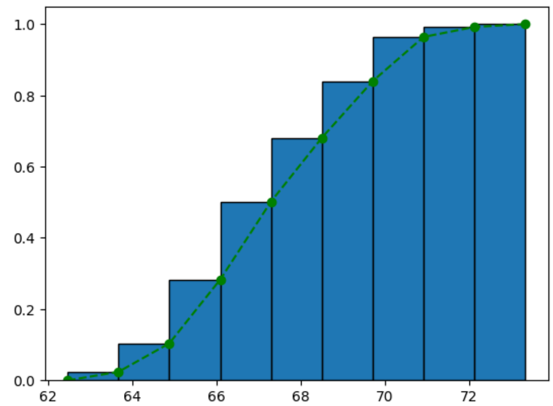
Гистограмма и полигон относительных частот



Гистограмма и полигон накопленных частот

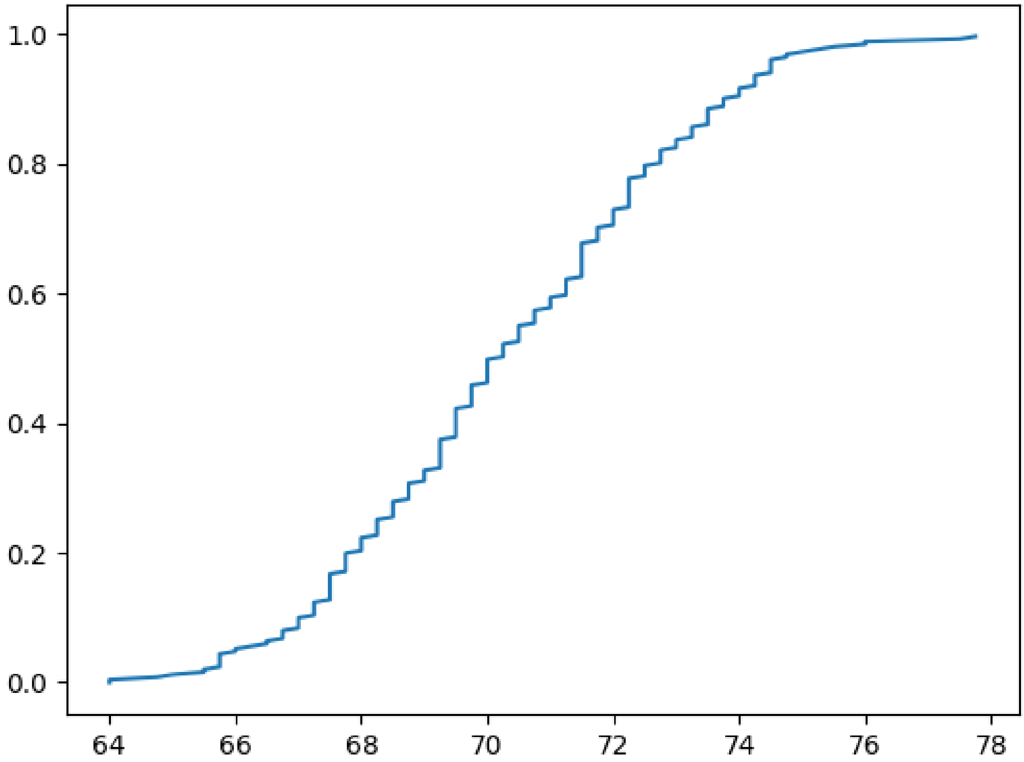


Гистограмма и полигон накопленных относительных частот



г) Построить график эмпирической функции распределения

Эмпирическая функция распределения



2. Интервальные оценки

2.1. Доверительные интервалы для мат. ожидания

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

Оцениваемый параметр – m

а) Привести формулы расчёта доверительных интервалов

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

б) Рассчитать доверительные интервалы

Граница доверительного интервала	α = 0.01	α = 0.05	α = 0.1
Нижняя граница	69.843	69.938	69.985
Верхняя граница	70.778	70.682	70.635

2.2. Доверительные интервалы для дисперсии

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

Оцениваемый параметр – σ^2

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	α = 0.01	α = 0.05	α = 0.1
Нижняя граница	5.487	5.643	5.778
Верхняя граница	8.710	8.434	8.210

2.3. Доверительные интервалы для разности мат. ожиданий

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n =252

Оцениваемый параметр – m_1-m_2

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$(\bar{X}_1 - \bar{X}_2) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
Верхняя граница	$(\bar{X}_1 - \bar{X}_2) + t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	α = 0.01	α = 0.05	α = 0.1
Нижняя граница	13.482	13.595	13.652
Верхняя граница	14.605	14.491	14.435

2.4. Доверительные интервалы для отношения дисперсий

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n = 252

Оцениваемый параметр – $\frac{\boldsymbol{\sigma_1}}{\boldsymbol{\sigma_2}}$

а) Привести формулы расчёта доверительных интервалов

Граница доверительного интервала	Формула расчета
Нижняя граница	$\frac{s_1^2}{s_2^2}\ F_{1-\frac{\alpha}{2}}(n_2 - 1, n_1 - 1)$
Верхняя граница	$\frac{s_1^2}{s_2^2}\ F_{\frac{\alpha}{2}}(n_2 - 1, n_1 - 1)$

б) Рассчитать доверительные интервалы

Граница доверительного интервала	α = 0.01	α = 0.05	α = 0.1
Нижняя граница	7.557	8.172	8.505
Верхняя граница	14.512	13.420	12.895

3. Проверка статистических гипотез о математических ожиданиях и дисперсиях

3.1. Проверка статистических гипотез о математических ожиданиях

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

Статистическая гипотеза – $\begin{matrix} H_0 : m = m_0 \\ H' : m \neq m_0 \end{matrix}$

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

	Выражение
Формула расчета статистики критерия	$Z = \dfrac{\bar{X} - m_0}{S\sqrt{n}}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$T(n-1)$
Формулы расчета критических точек	$\pm t_{1-\frac{\alpha}{2}}(n-1)$
Формула расчета *p-value*	$2 * \min(F_Z(z), 1 - F_Z(z))$

б) Выбрать произвольные значения m_0 и проверить статистические гипотезы

m_0	Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение $\alpha = 0.1$	Вывод
70	0.1	1.883	0.060	H_0 принимается	m = 70
71	0.1	-4.176	0.00004	H_0 отвергается	$m \neq 71$
69	0.1	7.943	0	H_0 отвергается	$m \neq 69$

3.2. Проверка статистических гипотез о дисперсиях

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

Статистическая гипотеза – $\begin{matrix} H_0 : \sigma = \sigma_0 \\ H' : \sigma \neq \sigma_0 \end{matrix}$

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

	Выражение
Формула расчета статистики критерия	$Z = \dfrac{(n-1)S^2}{\boldsymbol{\sigma}_0^2}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(n-1)$
Формулы расчета критических точек	$\chi^2_{\frac{\alpha}{2}}(n-1);\ \chi^2_{1-\frac{\alpha}{2}}(n-1);$

Формула расчета *p-value*	$2 * \min(F_Z(z), 1 - F_Z(z))$

б) Выбрать произвольные значения σ_0 и проверить статистические гипотезы

σ_0	Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение $\alpha = 0.1$	Вывод
2.6	0.1	252.756	0.878	H_0 принимается	$\boldsymbol{\sigma} = 2.6$
2	0.1	427.158	0	H_0 отвергается	$\boldsymbol{\sigma} \neq 2$
3	0.1	189.848	0.003	H_0 отвергается	$\boldsymbol{\sigma} \neq 3$

3.3. Проверка статистических гипотез о равенстве математических ожиданий

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n = 252

Статистическая гипотеза – $\begin{matrix} H_0 : m_1 = m_2 \\ H' : m_1 \neq m_2 \end{matrix}$

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

	Выражение
Формула расчета статистики критерия	$Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$T(n_1 + n_2 - 2)$
Формулы расчета критических точек	$\pm T_{1-\alpha}(n_1 + n_2 - 2)$
Формула расчета *p-value*	$2 * \min(F_Z(z), 1 - F_Z(z))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	70.499	0	H_0 отвергается	$m_1 \neq m_2$
0.05			H_0 отвергается	$m_1 \neq m_2$
0.1			H_0 отвергается	$m_1 \neq m_2$

3.4. Проверка статистических гипотез о равенстве дисперсий

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n = 252

Статистическая гипотеза – $\begin{matrix} H_0 : \sigma_1 = \sigma_2 \\ H' : \sigma_1 \neq \sigma_2 \end{matrix}$

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	$F(n_1 - 1, n_2 - 1)$
Формулы расчета критических точек	$F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1), F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$
Формула расчета *p-value*	$2 * \min(F_Z(z), 1 - F_Z(z))$

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	10.473	0	H_0 отвергается	$\boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$
0.05			H_0 отвергается	$\boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$
0.1			H_0 отвергается	$\boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$

4. Критерии согласия

Анализируемый признак – B8 Height (inches)

Объём выборки – n = 251

4.1. Критерий хи-квадрат

Теоретическое распределение – нормальное.

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

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 i выборки, принадлежащих интервалу J_i, $i = \overline{1,k}$ p_i - вероятность попадания в каждый интервал
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(k - r - 1)$	k - число интервалов r - число неизвестных параметров распределения
Формула расчета критической точки	$\chi^2_{1-\alpha}(k - r - 1)$	Критическая область выбирается правосторонней
Формула расчета *p-value*	$2 * min(F_Z(z), 1 - F_Z(z)\ )$	

б) Выбрать число групп

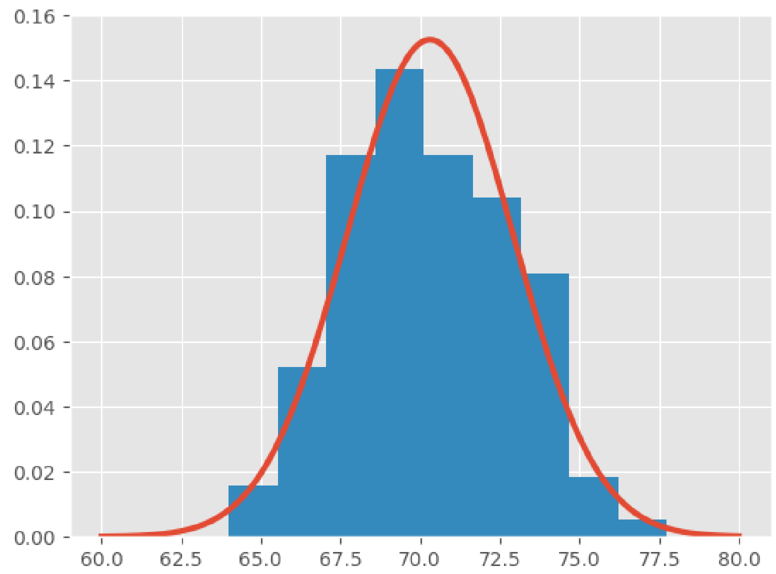
Число групп	Обоснование выбора числа групп	Ширина интервалов
9	Формула Стерджесса: $k \approx 1 + log_2 n$	1.53

в) Построить таблицу частот

Номер интервала	Нижняя граница	Верхняя граница	Частота	Относит. частота	Вероятность попадания в интервал при условии истинности основной гипотезы
1	64	65.52	6	0.024	0.025
2	65.52	67.05	20	0.08	0.072
3	67.05	68.58	45	0.179	0.147

4	68.58	70.11	55	0.219	0.215
5	70.11	71.63	45	0.179	0.224
6	71.63	73.16	40	0.159	0.168
7	73.16	74.69	31	0.124	0.090
8	74.69	76.22	7	0.028	0.034
9	76.22	77.75	2	0.008	0.009

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	7.759	0.256	H_0 принимается	$F(B8) \in N$
0.05			H_0 принимается	$F(B8) \in N$
0.1			H_0 принимается	$F(B8) \in N$

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

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

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$, где	$\hat{\mu}_3$ - третий момент (ассиметрия), $\hat{\mu}_4$ - четвертый момент (эксцесс), n - число наблюдений

	$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{3}{2}}}$ $K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$	
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(2)$	
Формула расчета критической точки	$\chi^2_{1-\alpha}(2)$	α - уровень значимости
Формула расчета *p-value*	$2 * min(F_Z(z), 1 - F_Z(z)\ )$	

б) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	2.324	0.312	H_0 принимается	$F(B8) \in N$
0.05			H_0 принимается	$F(B8) \in N$
0.1			H_0 принимается	$F(B8) \in N$

Вывод (в терминах предметной области)

В результате проведённого в п.4 статистического анализа обнаружено, что рост людей является нормально распределенной величиной.

5. Проверка однородности выборок

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n = 252

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	$Z = 2\sqrt{n}\,(H - 1/2)$	n – объем выборок $H = K/n$ – частота успеха K – число знаков «+» в последовательности знаков разностей $x_1 - y_1, \ldots, x_n - y_n$
Закон распределения статистики критерия при условии истинности основной гипотезы	$N(0,1)$	
Формула расчета критической точки	$N_{1-\frac{\alpha}{2}}(0,1)$	
Формула расчета *p-value*	$2 * min\left(F_Z(z), 1 - F_Z(z)\right)$	

б) Проверить статистические гипотезы

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

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	$Z = n_X n_Y \sum_{i=1}^{k} \frac{1}{m_i^{(X)} + m_i^{(Y)}} \left(\frac{m_i^{(X)}}{n_X} - \frac{m_i^{(Y)}}{n_Y} \right)^2$	$m_i^{(X)}/n_X$ и $m_i^{(Y)}/n_Y$ – относительные частоты, где n_X и n_Yобъемы выборок X и Y соответственно. k – число интервалов
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(k-1)$	
Формула расчета критической точки	$\chi^2{}_{1-\frac{\alpha}{2}}(k-1)$	
Формула расчета *p-value*	$2 * min\big(F_Z(z), 1 - F_Z(z)\big)$	

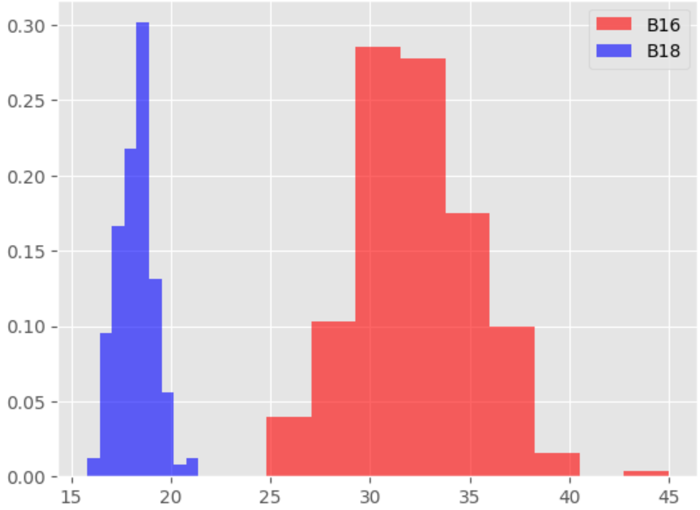
б) Выбрать число групп

Число групп	Обоснование выбора числа групп	Ширина интервалов
9	$k \approx 1 + log_2 n$	3.24

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	15,8	19,04	1	210	0,003953	0,83004
2	19,04	22,28	0	42	0	0,166008
3	22,28	25,53	2	0	0,007905	0
4	25,53	28,77	24	0	0,094862	0
5	28,77	32,02	100	0	0,395257	0
6	32,02	35,26	79	0	0,312253	0
7	35,26	38,51	45	0	0,177866	0
8	38,51	41,75	1	0	0,003953	0
9	41,75	45	1	1	0,003953	0,003953

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	76.958	0	H_0 отвергается	$F_1(x) \neq F_2(x)$
0.05			H_0 отвергается	$F_1(x) \neq F_2(x)$
0.1			H_0 отвергается	$F_1(x) \neq F_2(x)$

Вывод (в терминах предметной области)

В результате проведённого в п.5 статистического анализа обнаружено, что выборки окружности бицепса (B16) и окружности талии (B17) имеют разное распределение, т.к. критерий знаков и критерий хи-квадрат отвергли гипотезу однородности.

6. Таблицы сопряжённости

Факторный признак x – B3 Body fat

Результативный признак y – B6 Town

Объёмы выборок – 252

Статистическая гипотеза – $H_0: F_Y(y|X = x_1) =\ldots= F_Y(y|X = X_k) = F_Y(y)$

$H': \neg H_0$

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}- частота пары (x_i, y_j) в выборке, m_{ij}- теоретические частоты $m_{ij} = \frac{1}{n}\sum_{j=1}^{l} n_{ij} \sum_{i=1}^{k} n_{ij}$
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2((k-1)(l-1))$	k - число вариантов случайной величины X, l - число вариантов случайной величины Y
Формула расчета критической точки	$\chi^2_{1-\alpha}((k-1)(l-1))$	Критическая область для статистики критерия Z выбирается правосторонней. α - уровень значимости
Формула расчета *p-value*	$2 * min(F_Z(z), 1 - F_Z(z))$	

б) Построить эмпирическую таблицу сопряжённости

x \ y	high	low	normal	Σ
Arlington	39	8	26	73
Norwood	20	4	17	41
Revere	39	11	32	82
Somerville	22	16	18	56
Σ	120	39	93	252

в) Построить теоретическую таблицу сопряжённости

x \ y	high	low	normal	Σ
Arlington	34.76	11.29	26.94	73
Norwood	19.52	6.34	15.13	41

Revere	39.04	12.69	30.26	82
Somerville	26.66	8.66	20.66	56
Σ	120	39	93	252

г) Проверить статистические гипотезы

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	10.312	0.224	H_0 принимается	B3 не влияет на B6
0.05			H_0 принимается	B3 не влияет на B6
0.1			H_0 принимается	B3 не влияет на B6

Вывод (в терминах предметной области)

В результате проведённого в п.6 статистического анализа обнаружено, что тип B3 не влияет на значение выражения B6: нахождение людей в конкретном городе не влияет на его жировые отложения.

7. Дисперсионный анализ

Факторный признак x – B3 Body fat

Результативный признак y – B8 Height (inches)

Число вариантов факторного признака – 3

Объём выборки – n = 251

Статистическая гипотеза – H_0: $F_{X_1}(x) = \ldots = F_{X_k}(x) = F_X(x)$

а) Рассчитать групповые выборочные характеристики

№ п/п	Вариант факторного признака	Объём выборки	Групповые средние	Групповые дисперсии
1	high	120	69.99	19.98
2	low	39	70.43	7.58
3	normal	93	70.23	7.56

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	$D_\omega^* = \frac{1}{n}\sum_{k=1}^{K} n_k\tilde{\sigma}_k^2$	$n - K$	$\frac{n}{n-K}D_\omega^*$
Все признаки	$D_X^* = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(x_i^{(k)} - \bar{x}\right)^2$	$n - 1$	$\frac{n}{n-1}D_X^*$

в) Рассчитать показатели вариации, используемые в дисперсионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	0.026	2	3.294
Остаточные признаки	13.337	249	13.497
Все признаки	13.363	251	13.416

г) Проверить правило сложения дисперсий

Показатель	$D_{межгр}$	$D_{внутригр}$	$D_{общ}$	$D_{межгр} + D_{внутригр}$
Значение	0.026	13.337	13.363	13.363

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Эмпирический коэффициент детерминации	$\eta^2 = \frac{D_b^*}{D_X^*}$	0.002
Эмпирическое корреляционное отношение	$\eta = \sqrt{\frac{D_b^*}{D_X^*}}$	0.044

е) Охарактеризовать тип связи между факторным и результативным признаками

По шкале Чеддока степень тесноты связи между выборками B3 и B8 **слабая**.

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{D_b^*/(K-1)}{D_\omega^*(n-K)}$	К – кол-во вар-в факторного признака n-объем выборки D_b^*- межгрупповая дисперсия D_X^* -общая 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*	$2 * min(F_Z(z),\ 1 - F_Z(z))$	

з) Проверить статистическую гипотезу дисперсионного анализа

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	0.244	0.783	H_0 принимается	B3 не влияет на B8
0.05			H_0 принимается	B3 не влияет на B8
0.1			H_0 принимается	B3 не влияет на B8

Вывод (в терминах предметной области)

В результате проведённого в п.7 статистического анализа обнаружено, что тип В3 не влияет на В8.

8. Корреляционный анализ

8.1. Расчёт парных коэффициентов корреляции

Анализируемый признак 1 – B16 Biceps circumference (cm)

Анализируемый признак 2 – B18 Wrist circumference (cm)

Объёмы выборок – n = 252

а) Рассчитать точечные оценки коэффициентов корреляции

	Формула расчета	Значение
Линейный коэффициент корреляции	$\rho_{XY} = \frac{k_{XY}}{\sigma_X \sigma_Y}$ $k_{XY}^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$	0.632
Ранговый коэффициент корреляции по Спирмену	$\rho_{XY}^{(sp)} = \frac{\sum_{i=1}^{n}(R_i - \bar{r})(S_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{r})^2 \sum_{i=1}^{n}(S_i - \bar{s})^2}}$, где $r_i = s_i = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}$	0.602
Ранговый коэффициент корреляции по Кендаллу	$\tau_{XY} = \frac{N^+ - N^-}{\frac{1}{2}n(n-1)}$, где N^+- число пар наблюдений $(x_i, y_i), (x_j, y_j), i > j$, для которых выполнено условие $(x_i - x_j)(y_i - y_j) > 0$ N^-- число пар наблюдений $(x_i, y_i), (x_j, y_j), i > j$, для которых выполнено условие $(x_i - x_j)(y_i - y_j) < 0$	0.436

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	α = 0.01	α = 0.05	α = 0.1
Нижняя граница	0.535	0.558	0.570
Верхняя граница	0.730	0.707	0.695

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

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	12.898	0	H_0 отвергается	$\rho \neq 0$
$H_0: r^{(сп)} = 0$ $H': r^{(сп)} \neq 0$	0.1	11.912	0	H_0 отвергается	$r^{(\text{сп})} \neq 0$
$H_0: r^{(кен)} = 0$ $H': r^{(кен)} \neq 0$	0.1	10.350	0	H_0 отвергается	$r^{(\text{кен})} \neq 0$

8.2. Расчёт множественных коэффициентов корреляции

Анализируемый признак 1 – B8 Height (inches)

Анализируемый признак 2 – B10 Chest circumference (cm)

Анализируемый признак 3 – B3 Body fat

Объёмы выборок – 252

а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу

Признак \ Признак	B8	B10	B3
B8	1	0.178	-0.022
B10	0.178	1	-0.334
B3	-0.022	-0.334	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	B8	B10	B3
B8	–	0	0.652
B10	0	–	0
B3	0.652	0	–

в) Рассчитать точечную оценку коэффициента конкордации

	Формула расчета	Значение
Коэффициент конкордации	$W = \frac{12}{n^3-n}\sum_{i=1}^{n}(\frac{1}{k}\sum_{j=1}^{k} r_{ij} - \frac{n+1}{2})^2$, где n - объем выборок k - число выборок r_{ij}- ранг i-го элемента выборки в j-ом признаке	0.277

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = k(n-1)W$	W - коэффициент конкордации n - размер выборки k - число выборок
Закон распределения статистики критерия при условии истинности основной гипотезы	$\chi^2(n-1)$	
Формула расчета критической точки	$\chi^2_{1-\frac{\alpha}{2}}(n-1)$	α- уровень значимости
Формула расчета *p-value*	$2 * min(F_Z(z), 1 - F_Z(z))$	

д) Проверить значимость коэффициента конкордации

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	208.624	0.048	H_0 принимается	B8,B10,B3 – не зависимы
0.05			H_0 отвергается	B8,B10,B3 - зависимы
0.1			H_0 отвергается	B8,B10,B3 - зависимы

Вывод (в терминах предметной области)

В результате проведённого в п.8 статистического анализа обнаружено, что признак В3, В10 и В3 зависимы при уровнях значимости 0.05 и 0.1 и не зависимы при 0.01. Попарно зависимы только В8 и В10

9. Регрессионный анализ

9.1 Простейшая линейная регрессионная модель

Факторный признак x – B10 Chest circumference (cm)

Результативный признак y – B2 Percent body fat from Siri's equation

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	$\widetilde{\beta_0} = \bar{y} - \rho^*_{XY}\frac{\sigma^*_Y}{\sigma^*_X}\bar{x}$	-51.171
β_1	$\widetilde{\beta_1} = \rho^*_{XY}\frac{\sigma^*_Y}{\sigma^*_X}$	0.697

б) Записать точечную оценку уравнения регрессии

$f(x)$ = -51.171 + 0.697x

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

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	$D^*_{Y\|X} = \frac{1}{n}\sum_{i=1}^{n} n_i(\bar{y}_i - \bar{y})^2$	$k - 1$	$\frac{n}{k-1}D^*_{Y\|X}$
Остаточные признаки	$D^*_{resY} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f\left(x_i, \tilde{\beta}_0, \dots, \tilde{\beta}_{k-1}\right)\right)^2$	$n - k$	$\frac{n}{n-k}D^*_{resY}$
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	34.437	1	8678.313
Остаточные признаки	69.757	250	35.602
Все признаки	35.320	251	70.035

д) Проверить правило сложения дисперсий

Показатель	$D_{регр}$	$D_{ост}$	$D_{общ}$	$D_{регр} + D_{ост}$
Значение	34.437	35.320	69.757	69.757

е) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R^{2*}_{Y\|X} = \frac{D^*_{Y\|X}}{D^*_Y}$	0.493
Корреляционное отношение	$R^*_{Y\|X} = \sqrt{\frac{D^*_{Y\|X}}{D^*_Y}}$	0.702

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

Тип связи между факторным и результативным признаками, определяемой рассчитанной простейшей линейной регрессией – **слабая**.

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	$\bar{\beta}_0 - t_{1-\frac{\alpha}{2}}(n-2)\sqrt{\bar{D}_{resY}}\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n^2 D^*_X}}$
	Верхняя граница	$\bar{\beta}_0 + t_{1-\frac{\alpha}{2}}(n-2)\sqrt{\bar{D}_{resY}}\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n^2 D^*_X}}$
β_1	Нижняя граница	$\bar{\beta}_1 - t_{1-\frac{\alpha}{2}}(n-2)\sqrt{\bar{D}_{resY}}\sqrt{\frac{1}{\mathrm{n} D^*_X}}$
	Верхняя граница	$\bar{\beta}_1 + t_{1-\frac{\alpha}{2}}(n-2)\sqrt{\bar{D}_{resY}}\sqrt{\frac{1}{\mathrm{n} D^*_X}}$

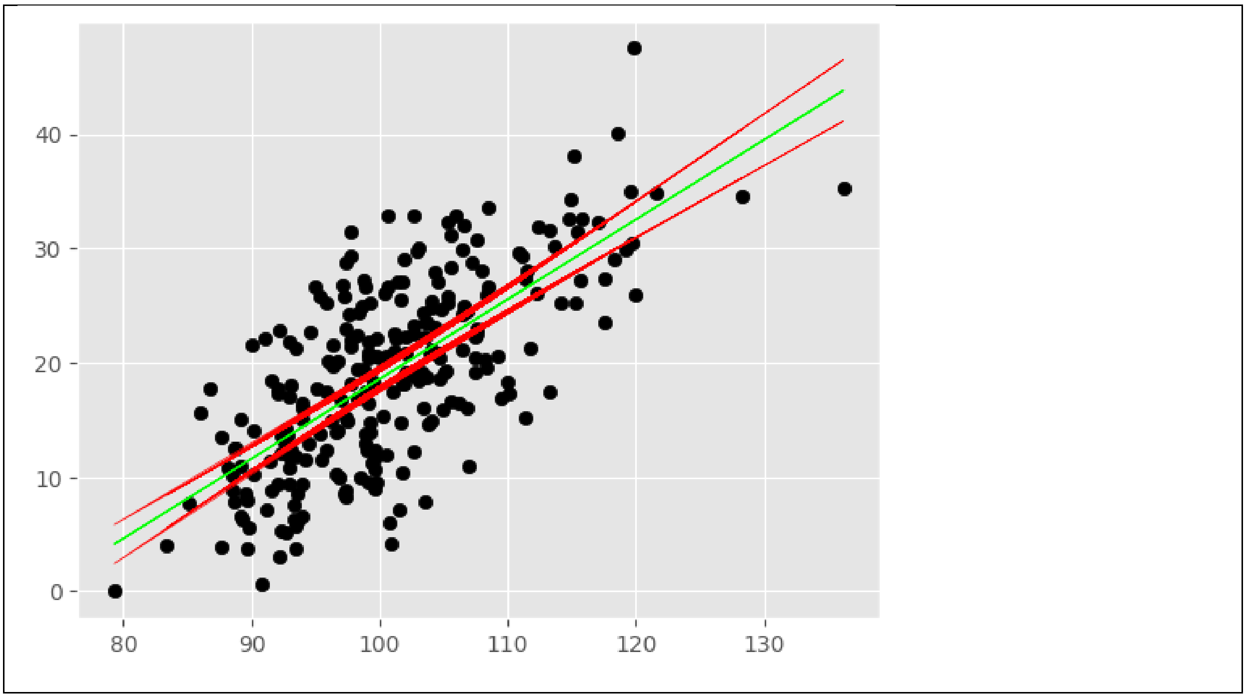
б) Рассчитать доверительные интервалы для параметров линейной регрессионной модели

Параметр	Границы доверительного интервала	α = 0.01	α = 0.05	α = 0.1
β_0	Нижняя граница	-62.903	-60.073	-58.634
	Верхняя граница	-39.44	-42.27	-43.709
β_1	Нижняя граница	0.69	0.692	0.693
	Верхняя граница	0.705	0.703	0.702

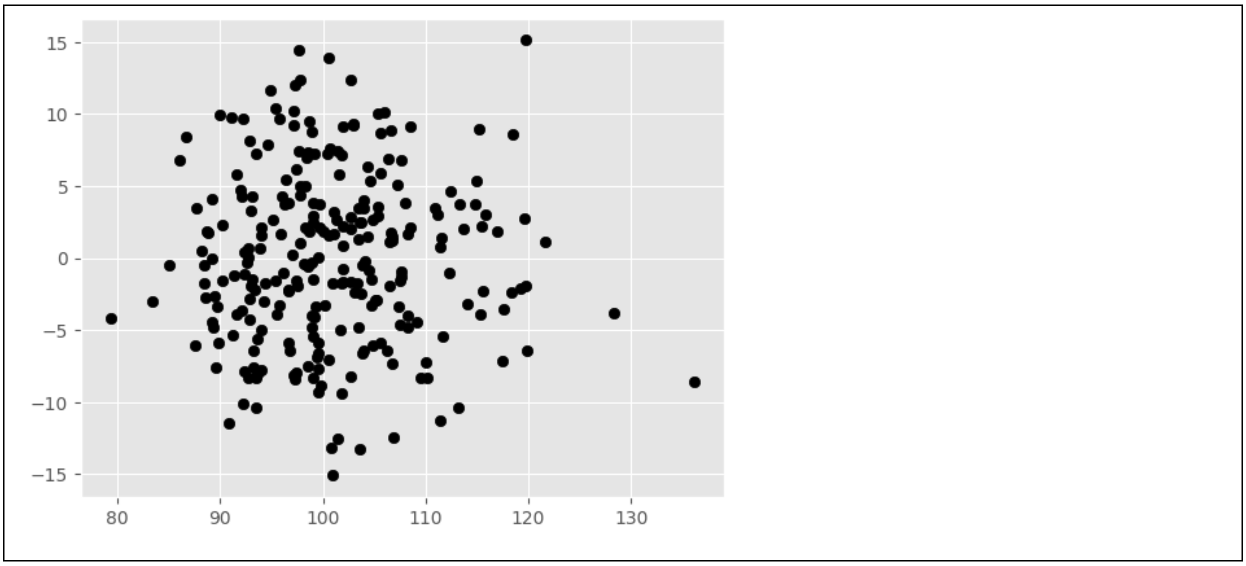
в) Привести формулы расчёта доверительного интервала для значений регрессии f(x)

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

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	$Z = \dfrac{R^{2*}_{Y\|X}}{(1 - R^{2*}_{Y\|X})/(n-2)}$	$R^{2*}_{Y\|X}$ - статистика, для точечной оценки коэффициента детерминации
Закон распределения статистики критерия при условии истинности основной гипотезы	$f(1, n-2)$	Распределение Фишера с 1 и n−2 степенями свободы
Формула расчета критической точки	$f_{1-\alpha}(1, n-2)$	α - уровень значимости
Формула расчета *p-value*	$1 - F_Z(z)$	

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	243.754	0	H_0 отвергается	Регрессионная модель значима
0.05			H_0 отвергается	Регрессионная модель значима
0.1			H_0 отвергается	Регрессионная модель значима

9.2 Линейная регрессионная модель общего вида

Факторный признак x – B10 Chest circumference (cm)

Результативный признак y – B2 Percent body fat from Siri's equation

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	$\tilde{\beta} = (F^T F)^{(-1)} F^T y$, где $\beta = (\beta_0, \dots, \beta_{k-1})^T$—вектор параметров модели $y = (y_1, \dots, y_n)^T$— вектор откликов модели	-102.654
β_1		1.697
β_2		-0.004

б) Записать точечную оценку уравнения регрессии

$f(x)$ = -102.654 + 1.697x - 0.004x^2

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	34.727	2	4375.691
Остаточные признаки	35.030	249	35.452
Все признаки	69.757	250	70.315

г) Проверить правило сложения дисперсий

Показатель	$D_{регр}$	$D_{ост}$	$D_{общ}$	$D_{регр} + D_{ост}$
Значение	34.727	35.030	69.757	69.757

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Коэффициент детерминации	$R^{2*}_{Y\|X} = \frac{D^*_{Y\|X}}{D^*_Y}$	0.497
Корреляционное отношение	$R^*_{Y\|X} = \sqrt{\frac{D^*_{Y\|X}}{D^*_Y}}$	0.705

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

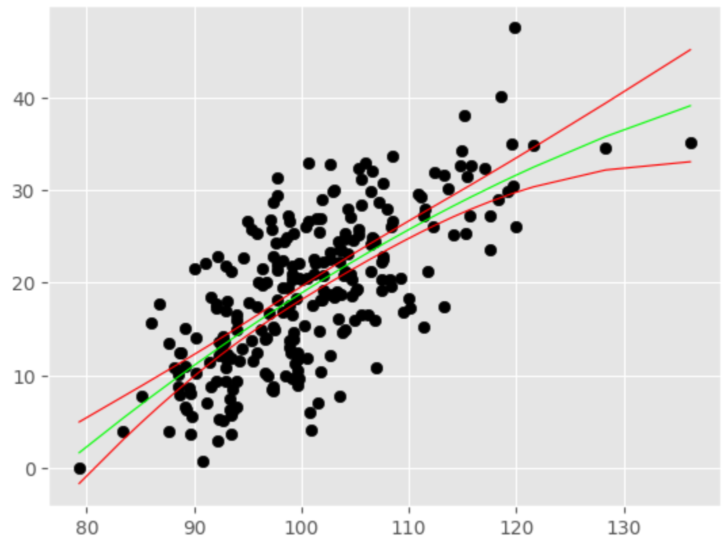
Тип связи между факторным и результативным признаками, определяемой рассчитанной простейшей линейной регрессией - **средняя**.

9.2.2. Интервальные оценки линейной регрессионной модели

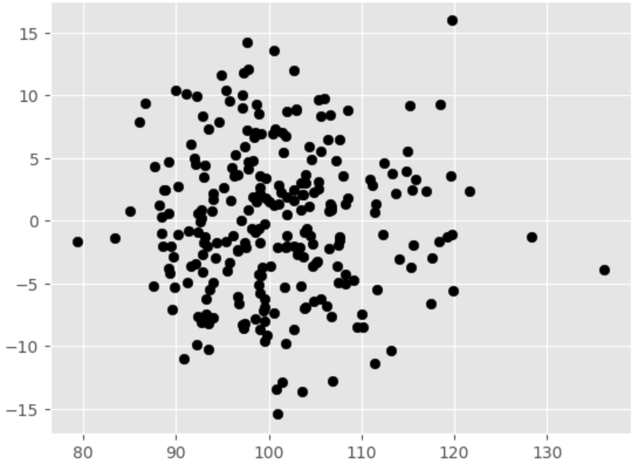
а) Привести формулы расчёта доверительного интервала для значений регрессии f(x)

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

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 – $\begin{matrix} H_0 : \beta_1 = \beta_2 = 0 \\ H' : \text{не } H_0 \end{matrix}$

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

	Выражение	Пояснение использованных обозначений
Формула расчета статистики критерия	$Z = \frac{R^{2*}_{Y\|X}/(k - 1)}{(1 - R^{2*}_{Y\|X})/(n-k)}$	$R^{2*}_{Y\|X}$ - статистика, для точечной оценки коэффициента детерминации
Закон распределения статистики критерия при условии истинности основной гипотезы	$f(k-1, n-k)$	Распределение Фишера с k–1 и n–k степенями свободы
Формула расчета критической точки	$f_{1-\alpha/2}(k-1, n-k)$	α - уровень значимости
Формула расчета *p-value*	$1 - F_Z(z)$	

б) Проверить значимость линейной регрессионной модели

Уровень значимости	Выборочное значение статистики критерия	*p-value*	Статистическое решение	Вывод
0.01	381.965	0	H_0 отвергается	Регрессионная модель значима
0.05			H_0 отвергается	Регрессионная модель значима
0.1			H_0 отвергается	Регрессионная модель значима

9.3 Множественная линейная регрессионная модель

Факторный признак 1 x_1 – B10 Chest circumference (cm)

Факторный признак 2 x_2 – B7 Weight (lbs)

Результативный признак y – B2 Percent body fat from Siri's equation

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	$\tilde{\beta} = (F^T F)^{(-1)} F^T y$, где	-54.222
β_1		0.767

β2	$\beta = (\beta_0, \dots, \beta_{k-1})^T$—вектор параметров модели$y = (y_1, \dots, y_n)^T$— вектор откликов модели	-0.022

б) Записать точечную оценку уравнения регрессии

$f(x) = -54.222 + 0.767x_1 - 0.022\ x_2$

в) Рассчитать показатели вариации, используемые в регрессионном анализе

Источник вариации	Показатель вариации	Число степеней свободы	Несмещенная оценка
Факторный признак	34.53	2	4350.191
Остаточные признаки	35.23	249	35.657
Все признаки	69.76	250	70.315

г) Проверить правило сложения дисперсий

Показатель	$D_{регр}$	$D_{ост}$	$D_{общ}$	$D_{регр} + D_{ост}$
Значение	34.53	35.23	69.76	69.76

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

Показатель	Формула расчета	Значение
Множественный коэффициент детерминации	$R^{2*}_{Y\|X} = \frac{D^*_{Y\|X}}{D^*_Y}$	0.4949
Множественное корреляционное отношение	$R^*_{Y\|X} = \sqrt{\frac{D^*_{Y\|X}}{D^*_Y}}$	0.7035

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

Тип связи между факторным и результативными признаками, определяемой рассчитанной линейной регрессией - **средняя**.

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	34.437	34.727	34.53

Остаточные признаки	69.757	35.030	35.23
Все признаки	35.320	69.757	69.76

б) Сводная таблица свойств различных регрессионных моделей

Свойство	Простейшая линейная модель	Линейная модель с квадратичным членом	Множественная линейная модель
Точность	49,3%	49,7%	49,49%
Значимость	Да	Да	Да
Адекватность	Да	Да	Да
Степень тесноты связи	слабая	средняя	средняя

Вывод (в терминах предметной области)

В результате проведённого в п.9 статистического анализа обнаружено, что все предложенные регрессионные модели адекватно отражают реальную зависимость окружности груди (см) (B10) от их веса (B7) и от их процента жира в организме из уравнения Siri (B2), при этом в данном случае точность простейшей линейной модели и линейной модели с квадратичным членом достаточно низкая, в отличии от множественной линейной модели.