Энгельсский технологический институт (филиал)
федерального государственного бюджетного образовательного учреждения
высшего образования
«Саратовский государственный технический университет
имени Гагарина Ю.А.»

Кафедра «Оборудование и технологии обработки материалов»

ПРАКТИКУМ ПО МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Электронное издание локального распространения

Методические указания к выполнению практических работ
по дисциплине «Математическое моделирование технологических процессов»
для студентов направлений подготовки:

15.03.05 «Конструкторско-технологическое обеспечение машиностроительных производств»

очной и заочной форм обучения

Энгельс 2025 г

Задача 1

Найти линейное уравнение регрессии.

X	Y
1	2,5n
2	3,7n
3	4,9n
4	6,4n
5	9,3n
6	11,5n

Вычислить следующие параметры (уровень значимости 0,05):

1. Коэффициент корреляции и его значимость

2. Коэффициент эластичности

3. Ошибка аппроксимации

4. Значимость коэффициентов a,b (t-статистика. Критерий Стьюдента)

5. Значимость уравнения в целом (F-статистика. Критерий Фишера. коэффициент детерминации).

Пример

Совокупность точек результативного и факторного признаков называется полем корреляции.

На основании поля корреляции можно выдвинуть гипотезу (для генеральной совокупности) о том, что связь между всеми возможными значениями X и Y носит линейный характер.

Линейное уравнение регрессии имеет вид $y = bx + a + \varepsilon$

Здесь ε - случайная ошибка (отклонение, возмущение).

Причины существования случайной ошибки:

1. Невключение в регрессионную модель значимых объясняющих переменных;

2. Агрегирование переменных. Например, функция суммарного потребления – это попытка общего выражения совокупности решений отдельных индивидов о расходах. Это лишь аппроксимация отдельных соотношений, которые имеют разные параметры.

3. Неправильное описание структуры модели;

4. Неправильная функциональная спецификация;

5. Ошибки измерения.

Так как отклонения ε_i для каждого конкретного наблюдения i – случайны и их значения в выборке неизвестны, то:

1) по наблюдениям x_i и y_i можно получить только оценки параметров α и β

2) Оценками параметров α и β регрессионной модели являются соответственно величины a и b, которые носят случайный характер, т.к. соответствуют случайной выборке;

Оценочное уравнение регрессии (построенное по выборочным данным) будет иметь вид $y = bx + a + \varepsilon$, где e_i – наблюдаемые значения (оценки) ошибок ε_i, a и b соответственно оценки параметров α и β регрессионной модели, которые следует найти.

Для оценки параметров α и β - используют МНК (метод наименьших квадратов).

Метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки относительно случайного члена (ε) и независимой переменной (x).

Формально критерий МНК можно записать так:

$S = \sum(y_i - y^*_i)^2 \rightarrow \min$

Система нормальных уравнений.

$a \bullet n + b\sum x = \sum y$

$a\sum x + b\sum x^2 = \sum y \bullet x$

Для наших данных система уравнений имеет вид

$6a + 21\,b = 383$

$21\,a + 91\,b = 1657$

Из первого уравнения выражаем *a* и подставим во второе уравнение:

Получаем эмпирические коэффициенты регрессии: b = 18.0857, a = 0.5333

Уравнение регрессии (эмпирическое уравнение регрессии):

$y = 18.0857\,x + 0.5333$

Эмпирические коэффициенты регрессии *a* и *b* являются лишь оценками теоретических коэффициентов β_i, а само уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных.

Для расчета параметров регрессии построим расчетную таблицу (табл. 1)

x	y	x^2	y^2	x • y
1	25	1	625	25
2	37	4	1369	74
3	49	9	2401	147
4	64	16	4096	256
5	93	25	8649	465
6	115	36	13225	690
21	383	91	30365	1657

1. Параметры уравнения регрессии.

Выборочные средние.

$$\overline{x} = \frac{\sum x_i}{n} = \frac{21}{6} = 3.5$$

$$\overline{y} = \frac{\sum y_i}{n} = \frac{383}{6} = 63.83$$

$$\overline{xy} = \frac{\sum x_i y_i}{n} = \frac{1657}{6} = 276.17$$

Выборочные дисперсии:

$$S^2(x) = \frac{\sum x_i^2}{n} - \overline{x}^2 = \frac{91}{6} - 3.5^2 = 2.92$$

$$S^2(y) = \frac{\sum y_i^2}{n} - \overline{y}^2 = \frac{30365}{6} - 63.83^2 = 986.14$$

Среднеквадратическое отклонение

$$S(x) = \sqrt{S^2(x)} = \sqrt{2.92} = 1.71$$

$$S(y) = \sqrt{S^2(y)} = \sqrt{986.14} = 31.4$$

1.1. Коэффициент корреляции

Ковариация.

$$\text{cov}(x,y) = \overline{x \bullet y} - \overline{x} \bullet \overline{y} = 276.17 - 3.5 \bullet 63.83 = 52.75$$

Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

$$r_{xy} = \frac{\overline{x \bullet y} - \overline{x} \bullet \overline{y}}{S(x) \bullet S(y)} = \frac{276.17 - 3.5 \bullet 63.83}{1.71 \bullet 31.4} = 0.98$$

Линейный коэффициент корреляции принимает значения от –1 до +1.

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются по шкале Чеддока:

$0.1 < r_{xy} < 0.3$: слабая;

$0.3 < r_{xy} < 0.5$: умеренная;

$0.5 < r_{xy} < 0.7$: заметная;

$0.7 < r_{xy} < 0.9$: высокая;

$0.9 < r_{xy} < 1$: весьма высокая;

В нашем примере связь между признаком Y фактором X весьма высокая и прямая.

Кроме того, коэффициент линейной парной корреляции может быть определен через коэффициент регрессии b:

$$r_{x,y} = b\frac{S(x)}{S(y)}$$

1.2. Уравнение регрессии (оценка уравнения регрессии).

$$y_x = r_{xy}\frac{x - \overline{x}}{S(x)}S(y) + \overline{y} = 0.98\frac{x - 3.5}{1.71}31.4 + 63.83 = 18.09x + 0.53$$

Линейное уравнение регрессии имеет вид y = 18.09 x + 0.53

Коэффициентам уравнения линейной регрессии можно придать экономический смысл.

Коэффициент регрессии b = 18.09 показывает среднее изменение результативного показателя (в единицах измерения y) с повышением или понижением величины фактора x на единицу его измерения. В данном примере с увеличением на 1 единицу y повышается в среднем на 18.09.

Коэффициент a = 0.53 формально показывает прогнозируемый уровень y, но только в том случае, если x=0 находится близко с выборочными значениями.

Но если x=0 находится далеко от выборочных значений x, то буквальная интерпретация может привести к неверным результатам, и даже если линия регрессии довольно точно описывает значения наблюдаемой выборки, нет гарантий, что также будет при экстраполяции влево или вправо.

Подставив в уравнение регрессии соответствующие значения x, можно определить выровненные (предсказанные) значения результативного показателя y(x) для каждого наблюдения.

Связь между y и x определяет знак коэффициента регрессии b (если > 0 – прямая связь, иначе - обратная). В нашем примере связь прямая.

1.4. Ошибка аппроксимации.

Оценим качество уравнения регрессии с помощью ошибки абсолютной аппроксимации. Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{\sum|y_i - y_x| : y_i}{n}100\%$$

Ошибка аппроксимации в пределах 5%-7% свидетельствует о хорошем подборе уравнения регрессии к исходным данным.

$$\overline{A} = \frac{0.59}{6}100\% = 9.9\%$$

Поскольку ошибка больше 7%, то данное уравнение не желательно использовать в качестве регрессии.

1.6. Коэффициент детерминации.

Квадрат (множественного) коэффициента корреляции называется коэффициентом детерминации, который показывает долю вариации результативного признака, объясненную вариацией факторного признака.

Чаще всего, давая интерпретацию коэффициента детерминации, его выражают в процентах.

$R^2 = 0.98^2 = 0.9674$

т.е. в 96.74 % случаев изменения x приводят к изменению y. Другими словами - точность подбора уравнения регрессии - высокая. Остальные 3.26 % изменения Y объясняются факторами, не учтенными в модели (а также ошибками спецификации).

Для оценки качества параметров регрессии построим расчетную таблицу (табл. 2)

x	y	y(x)	$(y_i-y_{ср})^2$	$(y-y(x))^2$	$(x_i-x_{ср})^2$	$\lvert y - y_x\rvert:y$
1	25	18.62	1508.03	40.72	6.25	0.26
2	37	36.7	720.03	0.0872	2.25	0.00798
3	49	54.79	220.03	33.53	0.25	0.12
4	64	72.88	0.0278	78.79	0.25	0.14
5	93	90.96	850.69	4.15	2.25	0.0219
6	115	109.05	2618.03	35.43	6.25	0.0518
21	383	383	5916.83	192.7	17.5	0.59

2. Оценка параметров уравнения регрессии.

2.1. Значимость коэффициента корреляции.

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_1 \neq 0$, надо вычислить наблюдаемое значение критерия

$$t_{набл} = r_{xy}\frac{\sqrt{n-2}}{\sqrt{1 - r^2_{xy}}}$$

и по таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы k = n - 2 найти критическую точку $t_{крит}$ двусторонней критической области. Если $t_{набл} < t_{крит}$ оснований отвергнуть нулевую гипотезу. Если $|t_{набл}| > t_{крит}$ — нулевую гипотезу отвергают.

$$t_{набл} = 0.98\frac{\sqrt{4}}{\sqrt{1 - 0.98^2}} = 15.35$$

По таблице Стьюдента с уровнем значимости α=0.05 и степенями свободы k=4 находим $t_{крит}$:

$t_{крит}$ (n-m-1;α/2) = (4;0.025) = 2.776

где m = 1 - количество объясняющих переменных.

Если $t_{набл} > t_{критич}$, то полученное значение коэффициента корреляции признается значимым (нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается).

Поскольку $t_{набл} > t_{крит}$, то отклоняем гипотезу о равенстве 0 коэффициента корреляции. Другими словами, коэффициент корреляции статистически - значим

В парной линейной регрессии $t^2_r = t^2_b$ и тогда проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

2.2. Интервальная оценка для коэффициента корреляции (доверительный интервал).

$$(r - t_{крит}\frac{1-r^2}{\sqrt{n}};\ r + t_{крит}\frac{1-r^2}{\sqrt{n}})$$

Доверительный интервал для коэффициента корреляции

$$(0.98 - 2.776\frac{1-0.98^2}{\sqrt{6}};\ 0.98 + 2.776\frac{1-0.98^2}{\sqrt{6}})$$

r(0.95;1.02)

2.3. Анализ точности определения оценок коэффициентов регрессии.

Несмещенной оценкой дисперсии возмущений является величина:

$$S^2_y = \frac{\sum(y_i - y_x)^2}{n - m - 1}$$

$$S^2_y = \frac{192.7}{4} = 48.18$$

$S^2_y = 48.18$ - необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии).

$$S_y = \sqrt{S^2_y} = \sqrt{48.18} = 6.94$$

$S_y = 6.94$ - стандартная ошибка оценки (стандартная ошибка регрессии).

S_a - стандартное отклонение случайной величины a.

$$S_a = S_y\frac{\sqrt{\sum x^2}}{n\ S(x)}$$

$$S_a = 6.94\frac{\sqrt{91}}{6 \cdot 1.71} = 6.46$$

S_b - стандартное отклонение случайной величины b.

$$S_b = \frac{S_y}{\sqrt{n}\ S(x)}$$

$$S_b = \frac{6.94}{\sqrt{6} \cdot 1.71} = 1.66$$

2.5. Проверка гипотез относительно коэффициентов линейного уравнения регрессии.

1) t-статистика. Критерий Стьюдента.

С помощью МНК мы получили лишь оценки параметров уравнения регрессии, которые характерны для конкретного статистического наблюдения (конкретного набора значений x и y).

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются t-критерий Стьюдента и доверительные интервалы каждого из показателей. Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля.

Чтобы проверить, значимы ли параметры, т.е. значимо ли они отличаются от нуля для генеральной совокупности используют статистические методы проверки гипотез.

В качестве основной (нулевой) гипотезы выдвигают гипотезу о незначимом отличии от нуля параметра или статистической характеристики в генеральной совокупности. Наряду с основной (проверяемой) гипотезой выдвигают альтернативную (конкурирующую) гипотезу о неравенстве нулю параметра или статистической характеристики в генеральной совокупности.

Проверим гипотезу H_0 о равенстве отдельных коэффициентов регрессии нулю (при альтернативе H_1 не равно) на уровне значимости $\alpha=0.05$.

В случае если основная гипотеза окажется неверной, мы принимаем альтернативную. Для проверки этой гипотезы используется *t-критерий Стьюдента*.

Найденное по данным наблюдений значение t-критерия (его еще называют наблюдаемым или фактическим) сравнивается с табличным (критическим) значением, определяемым по таблицам распределения Стьюдента (которые обычно приводятся в конце учебников и практикумов по статистике или эконометрике).

Табличное значение определяется в зависимости от уровня значимости (α) и числа степеней свободы, которое в случае линейной парной регрессии равно (n-2), n-число наблюдений.

Если фактическое значение t-критерия больше табличного (по модулю), то основную гипотезу отвергают и считают, что с вероятностью ($1-\alpha$) параметр или статистическая характеристика в генеральной совокупности значимо отличается от нуля.

Если фактическое значение t-критерия меньше табличного (по модулю), то нет оснований отвергать основную гипотезу, т.е. параметр или статистическая характеристика в генеральной совокупности незначимо отличается от нуля при уровне значимости α.

$t_{крит}$ (n-m-1;α/2) = (4;0.025) = 2.776

$$t_b = \frac{b}{S_b}$$

$$t_b = \frac{18.09}{1.66} = 10.9$$

Поскольку 10.9 > 2.776, то статистическая значимость коэффициента регрессии b подтверждается (отвергаем гипотезу о равенстве нулю этого коэффициента).

$$t_a = \frac{a}{S_a}$$

$$t_a = \frac{0.53}{6.46} = 0.0825$$

Поскольку 0.0825 < 2.776, то статистическая значимость коэффициента регрессии а не подтверждается (принимаем гипотезу о равенстве нулю этого коэффициента). Это означает, что в данном случае коэффициентом *a* можно пренебречь.

Доверительный интервал для коэффициентов уравнения регрессии.

Определим доверительные интервалы коэффициентов регрессии, которые с надежность 95% будут следующими:

$(b - t_{крит} S_b; b + t_{крит} S_b)$

(18.09 - 2.776 • 1.66; 18.09 + 2.776 • 1.66)

(13.48;22.69)

С вероятностью 95% можно утверждать, что значение данного параметра будут лежать в найденном интервале.

$(a - t_{крит} S_a; a + t_{крит} S_a)$

(0.53 - 2.776 • 6.46; 0.53 + 2.776 • 6.46)

(-17.4;18.47)

С вероятностью 95% можно утверждать, что значение данного параметра будут лежать в найденном интервале.

Так как точка 0 (ноль) лежит внутри доверительного интервала, то интервальная оценка коэффициента *a* статистически незначима.

2) F-статистика. Критерий Фишера.

Коэффициент детерминации R^2 используется для проверки существенности уравнения линейной регрессии в целом.

Проверка значимости модели регрессии проводится с использованием F-критерия Фишера, расчетное значение которого находится как отношение дисперсии исходного ряда наблюдений изучаемого показателя и несмещенной оценки дисперсии остаточной последовательности для данной модели.

Если расчетное значение с k_1=(m) и k_2=(n-m-1) степенями свободы больше табличного при заданном уровне значимости, то модель считается значимой.

$$R^2 = 1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \overline{y})^2} = 1 - \frac{192.7}{5916.83} = 0.9674$$

где m – число факторов в модели.

Оценка статистической значимости парной линейной регрессии производится по следующему алгоритму:

1. Выдвигается нулевая гипотеза о том, что уравнение в целом статистически незначимо: H_0: R^2=0 на уровне значимости α.

2. Далее определяют фактическое значение F-критерия:

$$F = \frac{R^2}{1 - R^2} \frac{(n - m - 1)}{m}$$

$$F = \frac{0.9674}{1 - 0.9674} \frac{(6-1-1)}{1} = 118.82$$

где m=1 для парной регрессии.

3. Табличное значение определяется по таблицам распределения Фишера для заданного уровня значимости, принимая во внимание, что число степеней свободы для общей суммы квадратов (большей дисперсии) равно 1 и число степеней свободы остаточной суммы квадратов (меньшей дисперсии) при линейной регрессии равно n-2.

$F_{табл}$ - это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α. Уровень значимости α - вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно α принимается равной 0,05 или 0,01.

4. Если фактическое значение F-критерия меньше табличного, то говорят, что нет основания отклонять нулевую гипотезу.

В противном случае, нулевая гипотеза отклоняется и с вероятностью (1-α) принимается альтернативная гипотеза о статистической значимости уравнения в целом.

Табличное значение критерия со степенями свободы k_1=1 и k_2=4, $F_{табл}$ = 7.71

Поскольку фактическое значение F > $F_{табл}$, то коэффициент детерминации статистически значим (найденная оценка уравнения регрессии статистически надежна).

Связь между F-критерием Фишера и t-статистикой Стьюдента выражается равенством:

$$t^2_r = t^2_b = \sqrt{F}$$

Показатели качества уравнения регрессии.

Показатель	Значение
Коэффициент детерминации	0.97
Средняя ошибка аппроксимации	9.9

Задача 2

Найти уравнение регрессии (для нечетных вариантов экспоненциальная y = a exp(bx), для четных – степенная y = a x^b).

X	Y
1	2,5n
2	5,7n
3	9,9n
4	16,4n

X	Y
5	29,3n
6	51,5n

Вычислить следующие параметры (уровень значимости 0,05):

1. Коэффициент корреляции и его значимость

2. Коэффициент эластичности

3. Ошибка аппроксимации

4. Значимость коэффициентов a,b (t-статистика. Критерий Стьюдента)

5. Значимость уравнения в целом (F-статистика. Критерий Фишера. коэффициент детерминации).

Пример

Экспоненциальное уравнение регрессии имеет вид $y = a\ e^{bx}$ ($\ln y = \ln a + bx + \varepsilon$)

Оценочное уравнение регрессии (построенное по выборочным данным) будет иметь вид $y = a\ e^{bx}$ ($\ln y = \ln a + bx + \varepsilon$), где e_i – наблюдаемые значения (оценки) ошибок ε_i, a и b соответственно оценки параметров α и β регрессионной модели, которые следует найти.

Для оценки параметров α и β - используют МНК (метод наименьших квадратов).

Система нормальных уравнений.

$a \bullet n + b\sum x = \sum y$

$a\sum x + b\sum x^2 = \sum y \bullet x$

Для наших данных система уравнений имеет вид

$6a + 21\ b = 29.35$

$21\ a + 91\ b = 113.11$

Из первого уравнения выражаем *a* и подставим во второе уравнение:

Получаем эмпирические коэффициенты регрессии: b = 0.593, a = 2.8168

Уравнение регрессии (эмпирическое уравнение регрессии):

$y = e^{2.81679915}e^{0.593x} = 16.72324e^{0.593x}$

Для расчета параметров регрессии построим расчетную таблицу (табл. 1)

x	ln(y)	x^2	y^2	x • y
1	3.22	1	10.36	3.22
2	4.2	4	17.68	8.41
3	4.68	9	21.92	14.05
4	5.15	16	26.56	20.61
5	5.74	25	32.95	28.7

6	6.35	36	40.38	38.13
21	29.35	91	149.84	113.11

1. Параметры уравнения регрессии.

Выборочные средние.

$$\overline{x} = \frac{\sum x_i}{n} = \frac{21}{6} = 3.5$$

$$\overline{y} = \frac{\sum y_i}{n} = \frac{29.35}{6} = 4.89$$

$$\overline{xy} = \frac{\sum x_i y_i}{n} = \frac{113.11}{6} = 18.85$$

Выборочные дисперсии:

$$S^2(x) = \frac{\sum x^2_i}{n} - \overline{x}^2 = \frac{91}{6} - 3.5^2 = 2.92$$

$$S^2(y) = \frac{\sum y^2_i}{n} - \overline{y}^2 = \frac{149.84}{6} - 4.89^2 = 1.04$$

Среднеквадратическое отклонение

$$S(x) = \sqrt{S^2(x)} = \sqrt{2.92} = 1.71$$

$$S(y) = \sqrt{S^2(y)} = \sqrt{1.04} = 1.02$$

1.3. Коэффициент эластичности.

Коэффициент эластичности находится по формуле:

$$E = \frac{\partial y}{\partial x}\frac{x}{y} = \overline{x}\ln(b)$$

$$E = 3.5\ln(0.59) = -1.83$$

В нашем примере коэффициент эластичности больше 1. Следовательно, при изменении X на 1%, Y изменится более чем на 1%. Другими словами - X существенно влияет на Y.

1.4. Ошибка аппроксимации.

$$\overline{A} = \frac{\sum |y_i - y_x| : y_i}{n} 100\%$$

$$\overline{A} = \frac{0.14}{6} 100\% = 2.39\%$$

Поскольку ошибка меньше 7%, то данное уравнение можно использовать в качестве регрессии.

1.6. Индекс детерминации.

$$R^2 = 1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \overline{y})^2}$$

$$R^2 = 1 - \frac{0.0881}{6.24} = 0.99$$

т.е. в 98.59 % случаев изменения x приводят к изменению y. Другими словами - точность подбора уравнения регрессии - высокая. Остальные 1.41 % изменения Y объясняются факторами, не учтенными в модели (а также ошибками спецификации).

Для оценки качества параметров регрессии построим расчетную таблицу (табл. 2)

x	ln(y)	y(x)	$(y_i-y_{ср})^2$	$(y-y(x))^2$	$(x_i-x_{ср})^2$	$\lvert y - y_x\rvert:y$
1	3.22	3.41	2.8	0.0364	6.25	0.0593
2	4.2	4	0.47	0.0408	2.25	0.048
3	4.68	4.6	0.0441	0.00747	0.25	0.0185
4	5.15	5.19	0.0682	0.00125	0.25	0.00687
5	5.74	5.78	0.72	0.00175	2.25	0.00729
6	6.35	6.37	2.14	0.00041	6.25	0.00319
21	29.35	29.35	6.24	0.0881	17.5	0.14

2. Оценка параметров уравнения регрессии.

2.1. Значимость коэффициента корреляции.

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_1 \neq 0$, надо вычислить наблюдаемое значение критерия

$$t_{набл} = r_{xy}\frac{\sqrt{n-2}}{\sqrt{1 - r^2_{xy}}}$$

и по таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы k = n - 2 найти критическую точку $t_{крит}$ двусторонней критической области. Если $t_{набл} < t_{крит}$ оснований отвергнуть нулевую гипотезу. Если $|t_{набл}| > t_{крит}$ — нулевую гипотезу отвергают.

$$t_{набл} = 0.99\frac{\sqrt{4}}{\sqrt{1 - 0.99^2}} = 23.59$$

По таблице Стьюдента с уровнем значимости α=0.05 и степенями свободы k=4 находим $t_{крит}$:

$t_{крит}$ (n-m-1;α/2) = (4;0.025) = 2.776

где m = 1 - количество объясняющих переменных.

Если $t_{набл} > t_{критич}$, то полученное значение коэффициента корреляции признается значимым (нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается).

Поскольку $t_{набл} > t_{крит}$, то отклоняем гипотезу о равенстве 0 коэффициента корреляции. Другими словами, коэффициент корреляции статистически - значим

В парной линейной регрессии $t^2_r = t^2_b$ и тогда проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

2.2. Интервальная оценка для коэффициента корреляции (доверительный интервал).

$$(r - t_{крит}\frac{1-r^2}{\sqrt{n}};\ r + t_{крит}\frac{1-r^2}{\sqrt{n}})$$

Доверительный интервал для коэффициента корреляции

$$(0.99 - 2.776\frac{1-0.99^2}{\sqrt{6}};\ 0.99 + 2.776\frac{1-0.99^2}{\sqrt{6}})$$

r(0.98;1.01)

2.3. Анализ точности определения оценок коэффициентов регрессии.

Несмещенной оценкой дисперсии возмущений является величина:

$$S^2_y = \frac{\sum(y_i - y_x)^2}{n - m - 1}$$

$$S^2_y = \frac{0.0881}{4} = 0.022$$

$S^2_y = 0.022$ - необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии).

$$S_y = \sqrt{S^2_y} = \sqrt{0.022} = 0.15$$

$S_y = 0.15$ - стандартная ошибка оценки (стандартная ошибка регрессии).

S_a - стандартное отклонение случайной величины a.

$$S_a = S_y\frac{\sqrt{\sum x^2}}{n\ S(x)}$$

$$S_a = 0.15\frac{\sqrt{91}}{6 \cdot 1.71} = 0.14$$

S_b - стандартное отклонение случайной величины b.

$$S_b = \frac{S_y}{\sqrt{n}\ S(x)}$$

$$S_b = \frac{0.15}{\sqrt{6} \cdot 1.71} = 0.0355$$

2.5. Проверка гипотез относительно коэффициентов линейного уравнения регрессии.

1) t-статистика. Критерий Стьюдента.

$t_{крит}$ (n-m-1;α/2) = (4;0.025) = 2.776

$$t_b = \frac{b}{S_b}$$

$$t_b = \frac{0.59}{0.0355} = 16.71$$

Поскольку 16.71 > 2.776, то статистическая значимость коэффициента регрессии b подтверждается (отвергаем гипотезу о равенстве нулю этого коэффициента).

$$t_a = \frac{a}{S_a}$$

$$t_a = \frac{2.82}{0.14} = 20.39$$

Поскольку 20.39 > 2.776, то статистическая значимость коэффициента регрессии a подтверждается (отвергаем гипотезу о равенстве нулю этого коэффициента).

Доверительный интервал для коэффициентов уравнения регрессии.

Определим доверительные интервалы коэффициентов регрессии, которые с надежность 95% будут следующими:

$(b - t_{крит} S_b; b + t_{крит} S_b)$

(0.59 - 2.776 • 0.0355; 0.59 + 2.776 • 0.0355)

(0.49;0.69)

С вероятностью 95% можно утверждать, что значение данного параметра будут лежать в найденном интервале.

$(a - t_{крит} S_a; a + t_{крит} S_a)$

(2.82 - 2.776 • 0.14; 2.82 + 2.776 • 0.14)

(2.43;3.2)

С вероятностью 95% можно утверждать, что значение данного параметра будут лежать в найденном интервале.

2) F-статистика. Критерий Фишера.

$$R^2 = 1 - \frac{\sum(y_i - y_x)^2}{\sum(y_i - \overline{y})^2} = 1 - \frac{0.0881}{6.24} = 0.9859$$

$$F = \frac{R^2}{1 - R^2} \frac{(n - m - 1)}{m}$$

$$F = \frac{0.9859}{1 - 0.9859} \frac{(6-1-1)}{1} = 279.35$$

Табличное значение критерия со степенями свободы k_1=1 и k_2=4, $F_{табл}$ = 7.71

Поскольку фактическое значение F > $F_{табл}$, то коэффициент детерминации статистически значим (найденная оценка уравнения регрессии статистически надежна).

Связь между F-критерием Фишера и t-статистикой Стьюдента выражается равенством:

$$t^2_r = t^2_b = \sqrt{F}$$

Показатели качества уравнения регрессии.

Показатель	Значение
Коэффициент детерминации	0.99
Средний коэффициент эластичности	-1.83
Средняя ошибка аппроксимации	2.39

Задача 3

Необходимо найти максимальное/ минимальное значение целевой функции $F = 4nx_1 + 6nx_2 \rightarrow \max$ (для нечетного варианта) и $F = 4nx_1 + 6nx_2 \rightarrow \min$ (для четного варианта), при системе ограничений:

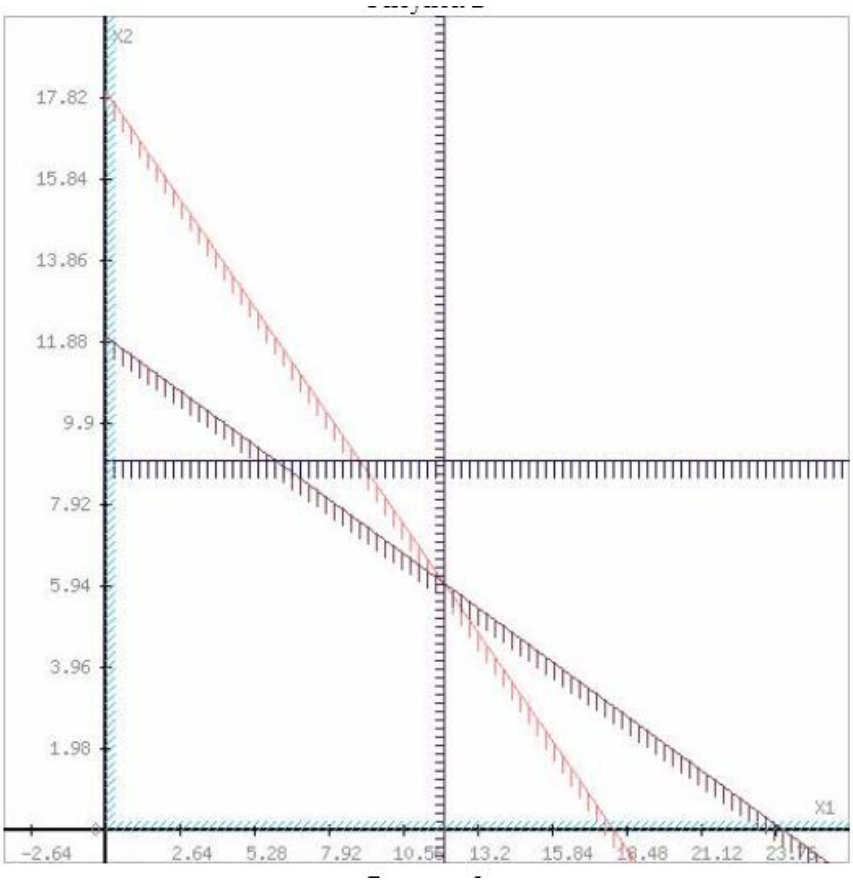
$$\begin{cases} nx_1 + nx_2 \leq 16n; \\ 0{,}5nx_1 + nx_2 \leq 11n; \\ x_1 \leq 11n; \\ x_2 \leq 7n; \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Пример

Необходимо найти максимальное значение целевой функции $F = 4x + 6y \rightarrow \max$, при системе ограничений:

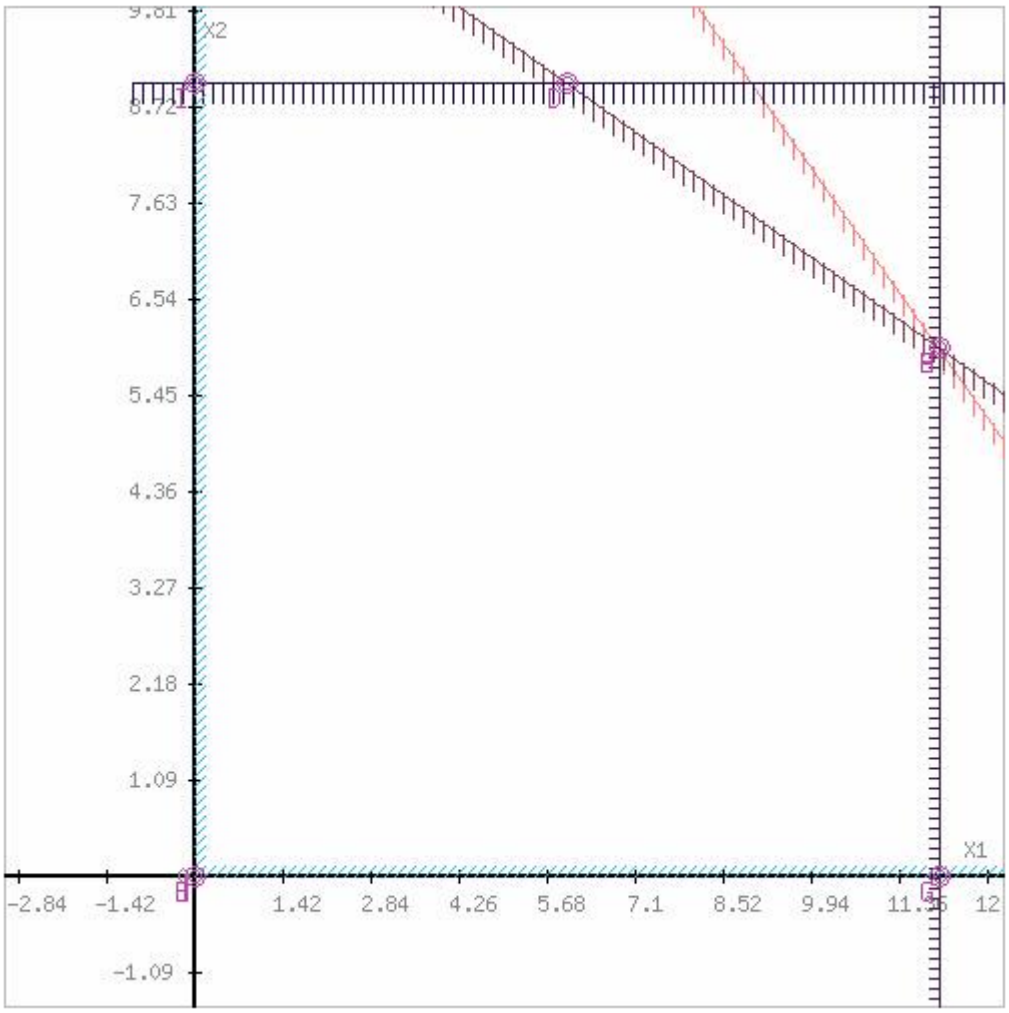
$$\begin{cases} x + y \leq 18; \\ 0{,}5x + y \leq 12; \\ x \leq 12; \\ y \leq 9; \\ x \geq 0, y \geq 0. \end{cases}$$

Построим область допустимых решений, т. е. решим графически систему неравенств.



$x = 12$ – параллельна оси OY;
$y = 9$ – параллельна оси OX;
$x > = 0$ – ось OY
$y = 0$ – ось OX;
$x \geq 0$ – полуплоскость правее оси OY;
$y \geq 0$ – полуплоскость выше оси OX;
$y \leq 9$ – полуплоскость ниже $y = 9$;
$x \leq 12$ – полуплоскость левее $x = 12$;
$0{,}5x + y \leq 12$ – полуплоскость ниже прямой $0{,}5x + y = 12$;
$x + y \leq 18$ – полуплоскость ниже прямой $x + y = 18$.

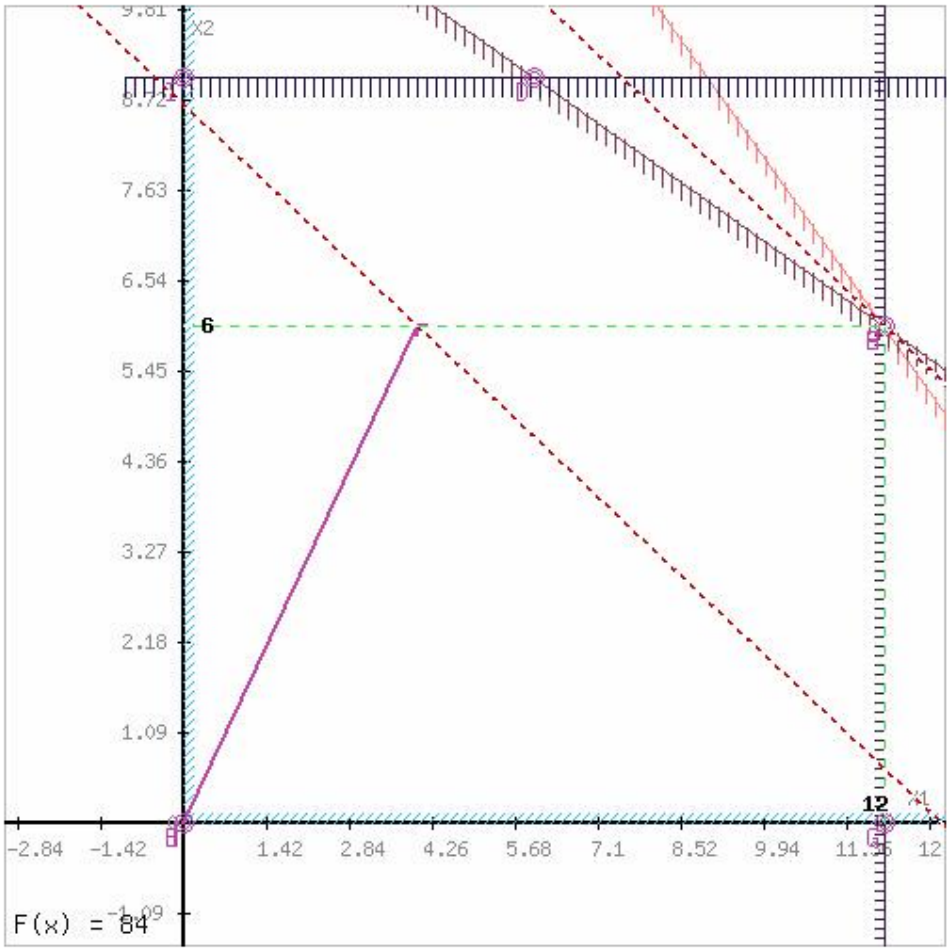
Пересечением всех этих полуплоскостей является пятиугольник $ABCDE$, с вершинами в точках $A(0;\ 0)$, $B(0;9)$, $C(6;\ 9)$, $D(12;6)$, $E(12;0)$. Этот пятиугольник и образует область допустимых решений задачи.



Рассмотрим целевую функцию задачи $F = 4x + 6y \to \max$.

x	3	0
y	–2	0

Построим прямую, отвечающую значению функции $F = 0$: $4x + 6y = 0$. Будем двигать эту прямую параллельным образом. Из всего семейства прямых $4x + 6y = \text{const}$ последней вершиной, через которую пройдет прямая при выходе за границу многоугольника, будет вершина C (12; 6). Именно в ней $F = 4x + 6y$ достигнет своего максимального значения.



Значит, при $x = 12$, $y = 6$ функция F достигает своего максимального значения $F = 4 \cdot 12 + 6 \cdot 6 = 84$, равного 84. Точка с координатами (12;6) удовлетворяет всем неравенствам системы ограничений, и в ней значение целевой функции оптимально $F^* = 84$.