МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Рубцовский индустриальный институт (филиал)
федерального государственного бюджетного образовательного учреждения высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»
(РИИ АлтГТУ)

Е.В. НИКИТЕНКО

ОСНОВЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Методические указания по выполнению
индивидуальных домашних заданий
для студентов всех форм обучения направления
«Информатика и вычислительная техника»

Рубцовск 2021

УДК 519.23, 519.25

Никитенко Е.В. Основы статистической обработки экспериментальных данных: методические указания по выполнению индивидуальных домашних заданий для студентов всех форм обучения направления «Информатика и вычислительная техника» / Е.В. Никитенко. – Рубцовск: РИИ, 2021. – 16 с. [ЭР].

Настоящие указания содержат необходимые элементы линейного и нелинейного регрессионного анализа. Приведен список из 24 вариантов индивидуальных заданий.

Рассмотрено и одобрено на
заседании кафедры ПМ РИИ
Протокол № 8 от 26.02.2021 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Выполнение индивидуальных домашних заданий по дисциплине «Основы статистической обработки экспериментальных данных» является важной частью организации самостоятельной работы студентов, направленной на более углубленное изучение отдельных вопросов учебной дисциплины.

Домашние задания необходимы для систематизации и закрепления ранее полученных теоретических знаний и практических умений, а также помогают сформировать умения работать с дополнительной и справочной литературой.

Использование индивидуальных домашних заданий по различным темам способствует усвоению и получению необходимых знаний и навыков для дальнейшего их использования при освоении и изучении других дисциплин.

Методические указания предназначены для студентов всех форм обучения по направлению подготовки бакалавров 09.03.01 «Информатика и вычислительная техника», изучающих дисциплину «Основы статистической обработки экспериментальных данных».

Настоящие указания содержат необходимые элементы линейного и нелинейного регрессионного анализа. Приведен список индивидуальных домашних заданий.

1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Множественная линейная регрессия

Рассмотрим задачу исследования зависимости одной переменной Y от нескольких объясняющих переменных (факторов) $X_1, X_2, \dots, X_k$. Обозначим -е наблюдение переменной y_i, а объясняющих переменных – $x_{i1}, x_{i2}, \dots, x_{ik}$. Тогда *модель множественной линейной регрессии* можно представить в виде:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n.$$

Функция линейной регрессии имеет вид: $f(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$.

Коэффициенты $\beta_0, \beta_1, \dots, \beta_k$ – неизвестные параметры, подлежащие оцениванию.

Случайная величина ε_i характеризует отклонение от функции регрессии. Эту переменную называют возмущением, ошибкой измерения или регрессионным остатком. Будем считать, что выполнены следующие условия:

1. Математическое ожидание возмущения ε_i для всех i равно нулю:
$$M(\varepsilon_i) = 0.$$
2. Дисперсия возмущения ε_i постоянна для каждого i:
$$D(\varepsilon_i) = \sigma^2.$$
3. Возмущения ε_i и ε_j не коррелированы:
$$M(\varepsilon_i \varepsilon_j) = M(\varepsilon_i) M(\varepsilon_j) = 0 \quad (i \neq j).$$

Данное предположение означает, что возмущения имеющиеся при наблюдении одного изучаемого объекта, не влияют на возмущения, характеризующие наблюдения над другими объектами.

В матричных обозначениях модель примет вид:

$$Y = X\beta + \varepsilon,$$

где $Y = (y_1, \dots, y_n)^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}.$$

$X -$ матрица размера $n \times (k+1)$ (матрица плана), составленная из наблюдаемых значений объясняющих переменных.

Основным методом построения оценок для коэффициентов регрессии β является *метод наименьших квадратов*, в соответствии с которым оценки этих параметров находят из условия минимизации суммы квадратов возмущений:

$$Q(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = (Y - X\beta)^T (Y - X\beta).$$

Точку $b = (b_0, b_1, \dots, b_n)$, удовлетворяющую равенству $Q(b) = \min\limits_{\beta} Q(\beta)$, называют оценкой метода наименьших квадратов параметра β.

Вектор-столбец оценок метода наименьших квадратов b неизвестных параметров β находится по формуле:

$$b = (X^T X)^{-1} X^T Y.$$

Введем обозначение для значений оценочной функции регрессии:

$$\hat{y}_i = \hat{f}(X_i) = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}.$$

Приведем общий вид симметричной матрицы $X^T X$:

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ik} \\ \dots & \dots & \dots & \dots \\ \sum x_{ik} & \sum x_{i1} x_{ik} & \dots & \sum x_{ik}^2 \end{pmatrix}.$$

$X^T Y$ – матрица размера $(k+1) \times 1$ (т.е. вектор-столбец) имеет следующий вид:

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ik} \end{pmatrix}.$$

В частном случае, для одной объясняющей переменной ($k = 1$) имеем:

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}.$$

Мы предполагаем, что матрица $X^T X$ является невырожденной, т.е. ее определитель отличен от нуля.

1.2 Значимость оценок и доверительные интервалы

Статистика

$$S^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

является несмещенной оценкой дисперсии σ^2.

Эмпирические дисперсии оценок коэффициентов регрессии найдем по следующим формулам:

$$S_{b_j}^2 = S^2[(X^TX)^{-1}]_{(j+1)(j+1)},$$

где $[(X^TX)^{-1}]_{(j+1)(j+1)}$ – элемент главной диагонали матрицы $(X^TX)^{-1}$, расположенный на пересечении $j+1$ строки и $j+1$ столбца.

Выберем уровень значимости α и вычислим статистику

$$t_{\beta_j} = \frac{b_j}{S_{b_j}}.$$

Для сравнения будем использовать $t_{1-\frac{\alpha}{2},n-k-1}$ – квантиль уровня $1-\frac{\alpha}{2}$ распределения Стьюдента при числе степеней свободы $n-k-1$.

Если $\left|t_{\beta_j}\right| > t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза $H_0: \beta_j = 0$ отклоняется и оценка b_j признается статистически значимой на уровне значимости α.

Если $\left|t_{\beta_j}\right| \le t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза $H_0: \beta_j = 0$ не отклоняется и оценка b_j признается статистически незначимой на уровне значимости α.

Доверительный интервал с уровнем доверия $\gamma = 1-\alpha$ для любого параметра β_j имеет вид

$$b_j - t_{1-\frac{\alpha}{2},n-k-1} \cdot S_{b_j} < \beta_j < b_j + t_{1-\frac{\alpha}{2},n-k-1} \cdot S_{b_j}.$$

1.3 Линейная регрессионная модель

Рассмотрим более подробно линейную (однофакторную) регрессионную модель

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \ \ i = 1, \dots, n,$$

где β_0 и β_1 – неизвестные параметры. Функциональная зависимость здесь

$$f(x) = \beta_0 \ + \beta_1 x \text{ – прямая.}$$

Рассмотрим сумму квадратов отклонений (или ошибок)

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Оценкой метода наименьших квадратов для неизвестных параметров (β_0, β_1) уравнения регрессии называется набор значений параметров (b_0, b_1), доставляющий минимум сумме $Q(\beta_0, \beta_1)$.

Используя необходимое условие экстремума, приравняем нулю частные производные:

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} = -2\sum\limits_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \\ \dfrac{\partial Q}{\partial \beta_1} = -2\sum\limits_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0 \end{cases}.$$

Проведя преобразования и поделив каждое уравнение на n получим:

$$\begin{cases} \dfrac{1}{n}\sum\limits_{i=1}^{n}\beta_0 + \beta_1\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i = \dfrac{1}{n}\sum\limits_{i=1}^{n}y_i, \\ \beta_0\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i + \beta_1\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i y_i, \end{cases}$$

или в других обозначениях

$$\begin{cases} \beta_0 + \beta_1\bar{x} = \bar{y}, \\ \beta_0\bar{x} + \beta_1\overline{x^2} = \overline{xy}, \end{cases}$$

где

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i, \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n}x_i y_i.$$

Решив систему, получим искомые оценки коэффициентов регрессии:

$$b_1 = \frac{n\cdot\sum x_i y_i - \sum x_i\cdot\sum y_i}{n\cdot\sum x_i^2 - (\sum x_i)^2} = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \qquad b_0 = \frac{\sum y_i - b_1\sum x_i}{n},$$

или

$$b_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad b_0 = \bar{y} - b_1\cdot\bar{x}.$$

Количественной мерой рассеяния значений y_i вокруг регрессии $\hat{f}(x)$ является несмещенная оценка остаточной дисперсии

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Эмпирические дисперсии оценок коэффициентов регрессии найдем по следующим формулам:

$$S_{b_0}^2 = S^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right) = \frac{S^2\sum x_i^2}{n\cdot\sum x_i^2 - (\sum x_i)^2}.$$

$$S_{b_1}^2 = \frac{nS^2}{n\cdot\sum x_i^2 - (\sum x_i)^2} = \frac{S^2}{\sum(x_i - \bar{x})^2}.$$

Статистики

$$t_{\beta_0} = \frac{b_0 - \beta_0}{S_{b_0}} \quad \text{и} \quad t_{\beta_1} = \frac{b_1 - \beta_1}{S_{b_1}}$$

при справедливости нулевой гипотезы H_0: $\beta_0 = b_0$ и $\beta_1 = b_1$ имеют распределение Стьюдента с $f = n - 2$ степенями свободы. Следовательно, с помощью распределения Стьюдента можно проверить гипотезу о значимости коэффициентов регрессии и построить доверительные интервалы.

С достоверностью (уровнем доверия) $\gamma = 1 - \alpha$, значения оценок коэффициентов регрессии b_0 и b_1 являются значимыми, если

$$|b_0| > t_{\frac{1+\gamma}{2}} \cdot S_{b_0} \text{ и } |b_1| > t_{\frac{1+\gamma}{2}} \cdot S_{b_1}.$$

Двусторонние $\gamma \cdot 100\%$ – ные доверительные интервалы для β_0 и β_1 имеют вид:

$$b_0 - t_{\frac{1+\gamma}{2}} \cdot S_{b_0} < \beta_0 < b_0 + t_{\frac{1+\gamma}{2}} \cdot S_{b_0},$$

$$b_1 - t_{\frac{1+\gamma}{2}} \cdot S_{b_1} < \beta_1 < b_1 + t_{\frac{1+\gamma}{2}} \cdot S_{b_1}.$$

1.4 Нелинейные регрессии

В случае нелинейной регрессии $y = f(x)$ используют различные линеаризующие преобразования переменных y и x. Наиболее распространенные из них приведены в нижеследующей таблице.

Линеаризующие функциональные преобразования
$(y^* = a^* + b^* x^*)$

Исходная зависимость $\boldsymbol{y = f(x)}$	Преобразование переменных		Преобразование коэффициентов	
	$\boldsymbol{y^*}$	$\boldsymbol{x^*}$	$\boldsymbol{a^*}$	$\boldsymbol{b^*}$
$y = a + \frac{b}{x}$	y	$\frac{1}{x}$	a	b
$y = a \cdot b^x$	$\lg y$	x	$\lg a$	$\lg b$
$y = a \cdot x^b$	$\lg y$	$\lg x$	$\lg a$	b
$y = a \cdot e^{bx}$	$\ln y$	x	$\ln a$	b

Рассмотрим квадратичную (параболическую) регрессию

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

Система нормальных уравнений для нахождения оценок коэффициентов имеет следующий вид:

$$\begin{cases} \beta_0 + \beta_1 \bar{x} + \beta_2 \overline{x^2} = \bar{y} \\ \beta_0 \bar{x} + \beta_1 \overline{x^2} + \beta_2 \overline{x^3} = \overline{xy} \\ \beta_0 \overline{x^2} + \beta_1 \overline{x^3} + \beta_2 \overline{x^4} = \overline{x^2 y} \end{cases}.$$

Введем следующие обозначения:

$$\Delta=\begin{vmatrix} 1 & \bar{x} & \overline{x^2} \\ \bar{x} & \overline{x^2} & \overline{x^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^4} \end{vmatrix};\ \Delta_0=\begin{vmatrix} \bar{y} & \bar{x} & \overline{x^2} \\ \overline{xy} & \overline{x^2} & \overline{x^3} \\ \overline{x^2y} & \overline{x^3} & \overline{x^4} \end{vmatrix};\ \Delta_1=\begin{vmatrix} 1 & \bar{y} & \overline{x^2} \\ \bar{x} & \overline{xy} & \overline{x^3} \\ \overline{x^2} & \overline{x^2y} & \overline{x^4} \end{vmatrix};$$

$$\Delta_2=\begin{vmatrix} 1 & \bar{x} & \bar{y} \\ \bar{x} & \overline{x^2} & \overline{xy} \\ \overline{x^2} & \overline{x^3} & \overline{x^2y} \end{vmatrix}.$$

Оценки коэффициентов регрессии найдем методом Крамера:

$$b_0=\frac{\Delta_0}{\Delta},\qquad b_1=\frac{\Delta_1}{\Delta},\qquad b_2=\frac{\Delta_2}{\Delta}.$$

1.5 Выборочный коэффициент детерминации

Рассмотрим следующие суммы:

$$TSS=\sum_{i=1}^{n}(y_i-\bar{y})^2;$$

$$ESS=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2;$$

$$RSS=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2.$$

TSS (Total Sum of Squares) – общая сумма квадратов отклонения наблюдаемых значений Y от среднего значения.

ESS (Explained Sum of Squares) – сумма квадратов отклонений, объясненная регрессией (факторная сумма квадратов).

RSS (Residual Sum of Squares) – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Эти суммы связаны соотношением:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

или

$$TSS=ESS+RSS.$$

Коэффициент детерминации определяется следующим образом:

$$R^2=\frac{ESS}{TSS}=1-\frac{RSS}{TSS}.$$

В матричной форме записи он имеет вид:

$$R^2=\frac{b^TX^TY-n(\overline{y})^2}{Y^TY-n(\overline{y})^2}.$$

Коэффициент детерминации показывает долю вариации зависимой переменной, обусловленную вариацией выборочной функции регрессии $\hat{f}(x) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$. В зависимости от тесноты связи может принимать значения от 0 до 1. Коэффициент детерминации R^2 можно рассматривать как меру качества уравнения регрессии: чем ближе он к единице, тем лучше регрессия описывает зависимость между зависимой и объясняющими переменными.

Иногда, предпочтительнее использовать скорректированный коэффициент детерминации $\hat{R}^2$, который определяется следующим образом:

$$\hat{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2).$$

1.6 Оценка значимости модели в целом

Наиболее важным применением коэффициента детерминации является его использование при проверке гипотезы статистической значимости линейной регрессионной модели в целом.

Говорят, что множественная линейная регрессия статистически значима на уровне α, если гипотеза $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ отклоняется на уровне значимости α и принимается альтернативная гипотеза H_1: хотя бы один $\beta_j \neq 0$.

Пусть $T_{1-\alpha;k;n-k-1}$ – квантиль уровня $1 - \alpha$ распределения Фишера с k и $n - k - 1$ степенями свободы. Если

$$F = \frac{R^2 \cdot (n-k-1)}{(1 - R^2) \cdot k} > T_{1-\alpha;k;n-k-1},$$

то гипотеза H_0 отклоняется на уровне значимости α и построенная линейная регрессия является статистически значимой.

Если же $F \leq T_{1-\alpha;k;n-k-1}$, то гипотеза H_0 принимается и построенная линейная регрессия является статистически незначимой.

2. ВАРИАНТЫ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

Задача 1.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	83	88	75	89	85	79	81	97	79	90	84	112
Y	137	142	128	140	133	153	142	154	132	150	132	166

Задача 2.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	10	14	21	23	27	32	39	45	55	61	62	168
Y	3,8	4,8	5,9	6,1	6,2	6,3	6,6	7,4	8,5	9,7	10,5	12,4

Задача 3.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	0,69	0,68	0,65	0,74	0,72	0,66	0,72	0,72	0,75	0,62	0,69	0,72
Y	53,9	48,9	46,8	52,2	53,5	48,7	49,2	53,7	57,6	45,7	49,6	49,4

Задача 4.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	6	7	8	9	10	11	12	13	14	15	16	17
Y	13	16	15	20	19	21	26	24	30	32	30	35

Задача 5.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	751,6	779,2	810,3	865,3	858,4	875,8	906,8	942,9	988,8	1015,5
Y	129,4	130,0	132,4	129,4	128,1	132,3	139,7	145,2	146,1	149,3

Задача 6.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	78	82	87	79	89	106	67	88	73	87	76	115
Y	133	148	134	154	162	195	139	158	152	162	159	173

Задача 7.

Найти оценки коэффициентов показательной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	78	82	87	79	89	106	67	88	73	87	76	115
Y	133	148	134	154	162	195	139	158	152	162	159	173

Задача 8.

Найти оценки коэффициентов множественной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X_1	5350	6880	7045	7250	5255	12090	3525	5430	7680	8225
X_2	420	550	570	880	430	840	930	525	675	685
X_3	330	485	500	790	360	725	820	430	610	620
Y	5000	6920	6900	10095	8090	195	4880	7355	10060	7885

Задача 9.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	50	81	104	120	139	154	180	208	241	250	269	301
Y	8	10	14	16	20	19	23	26	30	31	36	37

Задача 10.

Найти оценки коэффициентов гиперболической регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	1	2	3	5	10	20	30	50	100	200
Y	10,15	5,52	4,08	2,85	2,11	1,62	1,41	1,3	1,21	1,15

Задача 11.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	12,1	11,2	9,8	10,4	9,2	8,5	8,8	7,4	6,6	7,0	6,4	6,0
Y	10,5	9,3	8,3	9,6	8,6	7,1	6,9	5,8	5,2	5,0	5,1	4,6

Задача 12.

Найти оценки коэффициентов параболической регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	1/30	2/30	3/30	4/30	5/30	6/30	7/30	8/30	9/30	10/30
Y	11,86	15,67	20,6	26,69	33,71	41,93	51,13	61,49	72,9	85,44

Задача 13.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	83	88	75	89	85	79	81	97	79	90	84	112
Y	137	142	128	145	133	153	143	154	132	150	132	165

Задача 14.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	10	14	21	23	27	32	39	45	55	61	62	168
Y	3,9	4,8	5,9	6,1	6,2	6,3	6,7	7,4	8,5	9,7	10,5	12,5

Задача 15.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	0,69	0,68	0,65	0,74	0,72	0,66	0,72	0,72	0,75	0,62	0,69	0,72
Y	53,8	48,9	46,8	52,2	53,4	48,7	49,2	53,7	57,6	45,7	49,5	49,3

Задача 16.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	6	7	8	9	10	11	12	13	14	15	16	17
Y	12	16	15	20	19	21	25	24	30	32	30	34

Задача 17.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	751,6	779,2	810,3	865,3	868,4	875,8	906,8	941,9	988,8	1015,5
Y	129,5	130,0	132,4	129,4	128,2	132,3	139,7	145,2	146,1	149,5

Задача 18.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	78	82	87	79	89	106	67	88	73	87	76	115
Y	133	148	134	154	162	195	141	158	152	162	159	175

Задача 19.

Найти оценки коэффициентов показательной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	77	82	87	79	89	105	67	88	73	87	76	110
Y	132	148	134	154	162	192	139	158	152	162	159	170

Задача 20.

Найти оценки коэффициентов множественной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X_1	5350	6880	7045	7250	5255	12090	3525	5430	7680	8225
X_2	420	550	570	880	430	840	930	525	675	685
X_3	330	485	500	790	360	725	820	430	610	620
Y	5000	6930	6900	10090	8090	195	4890	7355	10060	7890

Задача 21.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	50	81	104	120	139	154	180	208	241	250	269	301
Y	8	10	14	16	20	19	23	26	30	31	36	37

Задача 22.

Найти оценки коэффициентов гиперболической регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	1	2	3	5	10	20	30	50	100	200
Y	10,14	5,52	4,08	2,85	2,11	1,61	1,41	1,3	1,21	1,14

Задача 23.

Найти оценки коэффициентов линейной регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	12,1	11,2	9,8	10,4	9,2	8,5	8,8	7,4	6,6	7,0	6,4	6,0
Y	10,3	9,3	8,3	9,6	8,6	7,2	6,9	5,8	5,2	5,0	5,1	4,5

Задача 24.

Найти оценки коэффициентов параболической регрессии. Построить 95% доверительные интервалы. Оценить значимость модели в целом.

X	1/30	2/30	3/30	4/30	5/30	6/30	7/30	8/30	9/30	10/30
Y	11,87	15,67	20,6	26,69	33,81	41,93	51,13	61,5	72,9	85,55

СПИСОК ЛИТЕРАТУРЫ

1. Буре, В.М. Теория вероятностей и математическая статистика: Учебник/ В.М. Буре, Е.М. Парилина. – СПб.: Лань, 2013. – 416 с.
2. Ивченко, Г. И. Введение в математическую статистику / Г. И. Ивченко, Ю. И. Медведев. Изд. 2–е, испр. и доп. – М.: ЛЕНАНД, 2017. – 608 с.
3. Ивченко, Г. И. Математическая статистика в задачах: Около 650 задач с подробными решениями / Г. И. Ивченко, Ю. И. Медведев, А.В. Чистяков. Изд. 3–е, испр. – М.: ЛЕНАНД, 2015. – 320 с.
4. Сборник задач по теории вероятностей, математической статистике и теории случайных функций: Учебное пособие / Под общей ред. А. А. Свешникова. – 5-е изд., стер. – СПб.: Лань, 2013. – 448 с.