Министерство образования Российской Федерации
Новосибирский государственный технический университет

Кафедра прикладной математики

Контрольная работа по дисциплине
Эконометрика

Выполнил:	Проверил:
Студент группы ОТЗ-001	
__________________ Ганжа К.И.	Бериков В.Б.
«____» ____________ 2003 г.	«____» ____________ 2003 г.
специальность: 060100	
шифр: 316 601 27	оценка ________________________

Новосибирск

2003

В каждой естественной науке заключено
столько истины, сколько в ней математики
(И. Кант)

Содержание

Задача № 1

В таблице 1 приведены следующие данные: единичная стоимость процесса добычи газа (*Y*), процент жидкости в добываемом из скважины газе (*X*) для различных месторождений.

1. В рамках линейной модели найдите регрессионную зависимость *Y* от *X*.
2. Вычислите коэффициент корреляции между *X* и *Y*.
3. Определите значимость регрессии для $\alpha = 0{,}05$.
4. Найдите 95 % доверительные интервалы для параметров модели.
5. Найдите интервал, в котором с вероятностью 0.95 находится значение единичной стоимости добычи газа при наличии 25 % жидкости.
6. Вычислите коэффициент детерминации R^2.

Таблица 1

X	13,2	16,4	19,4	23,5	26,2	30,5	42,3
Y	3,3	5,5	4,3	6,4	6,2	9,7	8,3

Решение

1. **Уравнение регрессии** – это формула статистической связи между переменными. Если эта формула линейна, то имеем **линейную регрессию**. Формула статистической связи *двух* переменных называется **парной регрессией.**

В модели **парной линейной регрессии** зависимость между переменными в генеральной совокупности представляется в виде[1]:

$$Y = \alpha + \beta X + \varepsilon \tag{1.1}$$

На основании выборочного наблюдения оценивается выборочное уравнение регрессии (*линия регрессии*):

$$y = a + bx\,, \tag{1.2}$$

[1] Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с. 22

где (a, b) – оценки параметров (α, β).

Неизвестные значения (a, b) определяются **методом наименьших квадратов (МНК)**, суть которого заключается в минимизации суммы квадратов остатков:

$$Q = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 \to \min$$

Здесь (x_i, y_i) – известные значения наблюдения, (a, b) – неизвестные. Записав необходимое условие экстремума и сделав необходимые преобразования, получим систему уравнений[2].

$$\begin{cases} \hat{b} = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases} \quad (1.3)$$

Для удобства, сделаем промежуточные расчёты и внесём их в таблицу 1.1

Таблица 1.1

I	1	2	3	4	5	6	7	**ср.**
x	13,2	16,4	19,4	23,5	26,2	30,5	42,3	**24,5**
y	3,3	5,5	4,3	6,4	6,2	9,7	8,3	**6,24**
*x*y*	43,56	90,20	83,42	150,40	162,44	295,85	351,09	**168,14**
*сум.(x*y)*	**1176,96**							
x ^{2}	174,24	268,96	376,36	552,25	686,44	930,25	1789,29	**682,54**
сум. x ^{2}	**4777,79**							

Подставим данные из таблицы 1.1 в (1.3) и вычислим:

$$\hat{b} = \frac{168{,}14 - 24{,}5 \cdot 6{,}24}{682{,}54 - 24{,}5^2} = 0{,}1846$$

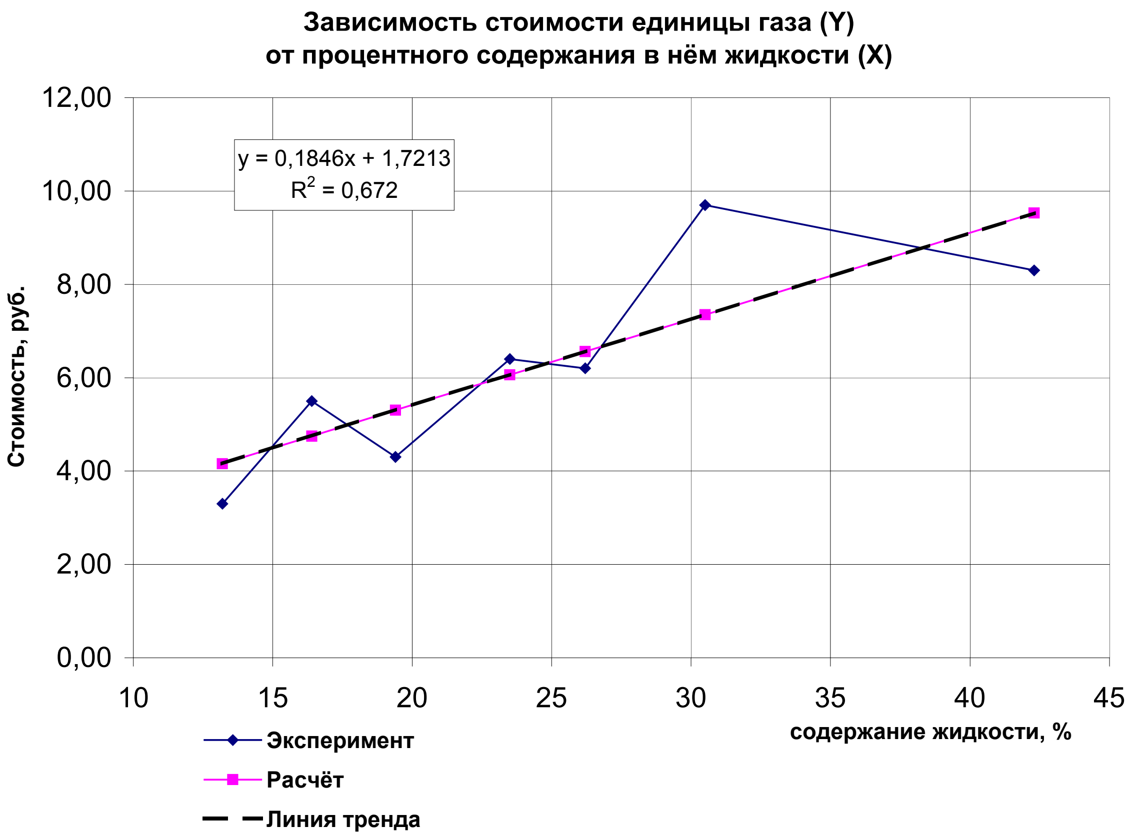
$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x} = 6{,}24 - 0{,}18 \cdot 25{,}4 = 1{,}7213$$

Таким образом линейная модель зависимости y от x имеет вид:

$$\boldsymbol{y_i = 1{,}7213 + 0{,}1846 \cdot x_i}$$

Построим график полученной зависимости и нанесём на него экспериментальные точки.

[2] Подробнее о преобразованиях см.: Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с. 23.



2. **Коэффициент корреляции** (ρ) является безразмерной величиной и показывает степень линейной связи двух переменных. Коэффициент корреляции определяется выражением:

$$\hat{\rho} = \frac{pop.\mathrm{cov}(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \hat{y}) \cdot (x_i - \hat{x})}{\hat{\sigma}_x \cdot \hat{\sigma}_y} = \hat{b} \cdot \frac{\hat{\sigma}_x}{\hat{\sigma}_y}, \ -1 \le \rho \le 1 \qquad (1.4)$$

где

$$\hat{\sigma}_x = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2}\,, \quad \hat{\sigma}_y = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \overline{y})^2} \qquad (1.5)$$

Для удобства, сделаем промежуточные расчёты и внесём их в таблицу 1.2

Таблица 1.2

I	1	2	3	4	5	6	7	**ср.**
x	13,2	16,4	19,4	23,5	26,2	30,5	42,3	**24,5**
y	3,3	5,5	4,3	6,4	6,2	9,7	8,3	**6,24**
$(x - x_{ср})^2$	127,69	65,61	26,01	1,00	2,89	36,00	316,84	
сумма	**576,04**							
$(y - y_{ср})^2$	8,66	0,55	3,77	0,02	0,00	11,95	4,23	
сумма	**29,20**							

Подставим данные в (1.5):

$$\hat{\sigma}_x = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2} = \sqrt{\frac{1}{7} \cdot 576{,}04} = 9{,}07$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2} = \sqrt{\frac{1}{7} \cdot 29{,}2} = 2{,}04$$

Подставим полученные данные в (1.4) и рассчитаем коэффициент корреляции:

$$\hat{\rho} = \hat{b} \cdot \frac{\hat{\sigma}_x}{\hat{\sigma}_y} = 0{,}18 \cdot \frac{9{,}07}{2{,}04} = 0{,}8$$

На основании этого можно сделать вывод, что имеется положительная связь между содержанием жидкости и стоимостью добычи, т.е. с ростом содержания жидкости растёт стоимость добычи.

3.1. **Проверка адекватности регрессионной модели.**[3]

Значимость коэффициентов простой линейной регрессии для нашего примера (для N < 30) осуществляют с помощью *t-критерия Стьюдента*. При этом вычисляют расчётные (фактические) значения *t*-критерия:

для параметра *a*

$$t_a = |a| \frac{\sqrt{N-2}}{\sigma_{ост}} \tag{1.6}$$

для параметра *b*

---

[3] Подробнее см.: Гусаров В.М. Статистика: Учебное пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001, с. 192–199. А так же: Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с. 39–42.

$$t_b = |b| \frac{\sqrt{N-2}}{\sigma_{ост}} \sigma_x \qquad (1.7)$$

Для удобства, сделаем промежуточные расчёты и внесём их в таблицу 1.3

Таблица 1.3

I	1	2	3	4	5	6	7	ср.
x	13,2	16,4	19,4	23,5	26,2	30,5	42,3	**24,5**
y	3,3	5,5	4,3	6,4	6,2	9,7	8,3	**6,24**
$(x-x_{ср})^2$	127,69	65,61	26,01	1,00	2,89	36,00	316,84	
сумма	**576,04**							
$y_{оц}$	4,16	4,75	5,30	6,06	6,56	7,35	9,53	
$(y-y_{оц})^2$	0,74	0,56	1,01	0,12	0,13	5,51	1,51	
сумма	**9,58**							

Среднеквадратичное отклонение результативного признака y от выравненных значений $\hat{y}$:

$$\sigma_{ост} = \sqrt{\frac{\sum (y-\hat{y})^2}{N}} = \sqrt{\frac{9{,}58}{7}} = 1{,}1699 \qquad (1.8)$$

Среднеквадратичное отклонение факторного признака x от общей средней $\bar{x}$:

$$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{N}} = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2} = \sqrt{\frac{576{,}04}{7}} = 9{,}071 \qquad (1.9)$$

Теперь по (1.6) и (1.7) рассчитаем фактические значения *t*-критерия:

$$t_a = 1{,}7213 \frac{\sqrt{(7-2)}}{1{,}1699} = 3{,}29$$

$$t_b = 0{,}1846 \cdot \frac{\sqrt{7-2}}{1{,}1699} \cdot 9{,}071 = 3{,}20$$

Вычисленные значения t_a и t_b сравнивают с критическими $t_{кр}$, которые определяют по таблице распределения Стьюдента с учётом принятого уровня значимости α и числом степеней свободы вариации ν = N – 2. В социально-экономических исследованиях уровень значимости α обычно принимают равным 0,05. Параметр признаётся значимым (существенным) при условии, если $t_{расч} > t_{табл}$.

По таблице распределения Стьюдента (см. приложение 1) для $\nu = 5$ находим критическое значение *t*-критерия: $t_{табл} = 2{,}01$

Поскольку расчётные значения t_a и $t_b > t_{табл}$, то оба параметра *a* и *b* признаются значимыми.

3.2. **F-тест на качество оценивания.** Для определения статистической значимости проверяется гипотеза H_0: F=0 (нулевая гипотеза о значимости параметров) для F-статистики, рассчитываемой по формуле:

$$F = \frac{\hat{\rho}^2 (N-2)}{(1-\hat{\rho}^2)} \tag{1.10}$$

Величина F имеет распределение Фишера с $\nu_1 = 1$, $\nu_2 = N - 2$ степенями свободы.

Вычисленный критерий F сравнивается с критическим значением $F_{кр}$:
если $F < F_{кр}$, то гипотеза H_0 принимается, т.е. параметры незначимы;
если $F > F_{кр}$, то гипотеза H_0 отклоняется, т.е. параметры значимы.

Вычислим критерий F по формуле (1.10):

$$F = \frac{\hat{\rho}^2 (N-2)}{(1-\hat{\rho}^2)} = \frac{0{,}8^2 (7-2)}{(1-0{,}8^2)} = 8{,}9$$

Теоретический критерий Фишера определим по таблице, принимая во внимание, что $\nu_1 = 1$, $\nu_2 = (7 - 2) = 5$:

$F_{теор}$=6,61.

Т.к. $F > F_{теор}$, то гипотеза о неслучайной связи между X и Y подтверждается.

4. 95 % доверительный интервал для параметров модели.

Несмещённой оценкой дисперсии σ^2 является величина (остаточная дисперсия)[4]

$$S^2 = \frac{n}{n-2}\operatorname{var}(e) = \frac{1}{n-2}\sum e_i^2 \tag{1.11}$$

[4] Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с.36.

которая служит *мерой разброса* зависимой переменной вокруг линии регрессии. Величина S называется **стандартной ошибкой регрессии.**

Заменив в теоретических дисперсиях неизвестную σ^2 на оценку S^2 получим оценки дисперсии:

$$S_b^2 = \frac{\overline{x^2} S^2}{n\,\mathrm{var}(x)} = \frac{S^2}{\sum (x-\bar{x})^2},\quad S_a^2 = \frac{S^2}{n\,\mathrm{var}(x)} = \overline{x^2} S_b^2 \qquad (1.12)$$

Величины S_a, S_b называются **стандартными ошибками коэффициентов регрессии.**

Подставим данные из табл. 1.3 в (1.11) и вычислим:

$$S^2 = \frac{1}{7-2}\cdot 9{,}58 = 1{,}916$$

А теперь по (1.12) вычислим стандартные ошибки коэффициентов регрессии:

$$S_b^2 = \frac{S^2}{\sum (x-\bar{x})^2} = \frac{1{,}916}{576{,}04} = 3{,}326\cdot 10^{-3}$$

$$S_a^2 = \overline{x^2} S_b^2 = 682{,}54\cdot 3{,}326\cdot 10^{-3} = 2{,}27$$

Учитывая, что значения статистики Стьюдента для α = 0,05/2 и ν = 5 равна $a_t = b_t$ = 2,57 (см. приложение 1)

$$b^{\pm} = b \pm b_t \cdot S_b = 0{,}1846 \pm 2{,}57\cdot 3{,}326\cdot 10^{-3} = 0{,}1846 \pm 0{,}00855$$

$$a^{\pm} = a \pm a_t \cdot S_a = 1{,}7213 \pm 2{,}57\cdot 2{,}27 = 1{,}7213 \pm 5{,}8533$$

5. Рассчитаем интервал, в который с вероятностью 95 % находится значение единичной стоимости добычи газа при наличии в нём 25 % жидкости.

Исходя из найденных параметров регрессионной модели, стоимость извлечения газа будет равна:

$y = 1{,}7213 + 0{,}1846 \cdot x = 1{,}7213 + 0{,}1846 \cdot 25 = 6{,}3363$

Доверительный интервал рассчитаем по формуле:

$$y^{\pm} = y \pm a_t S\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum (x-\bar{x})^2}} = 6{,}3363 \pm 2{,}57\cdot\sqrt{1{,}916}\cdot\sqrt{\frac{1}{7} + \frac{(25-24{,}5)^2}{567{,}04}} = 6{,}3363 \pm 1{,}3464$$

Т.е. с вероятностью 95 % стоимость добычи газа с 25 % содержанием жидкости будет находиться в диапазоне от 4,99 до 7,68.

6. **Коэффициентом детерминации $\mathbf{R^2}$** называют отношение[5]

$$R^2 = \frac{\operatorname{var}(\hat{y})}{\operatorname{var}(y)} = 1 - \frac{\operatorname{var}(e)}{\operatorname{var}(y)} = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (1.13)$$

характеризующее долю вариации (разброса) зависимой переменной, *объяснённую* с помощью уравнения регрессии.

Подставим в (1.13) данные из табл. 1.2 и 1.3 и получим:

$$R^2 = 1 - \frac{9,58}{29,20} = 0,672$$

Т.е. изменение исследуемого признака (стоимость добычи газа) на 67 % зависит от анализируемого фактора (наличия в газе жидкости).

5 Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с. 25.

Задача № 3

У семи сотрудников предприятия собраны данные (табл. 3) об их среднемесячной зарплате (Y), возрасте (X_1) и стаже работы (X_2).

1. С помощью метода наименьших квадратов (МНК) оценить параметры линейной модели вида $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$, влияния возраста и стажа работы на среднемесячную зарплату;

2. Оценить параметры построенной модели;

3. Рассчитать коэффициент детерминации.

Таблица 3

X_1	35	45	25	55	30	42	25
X_2	5	10	3	12	1	8	2
Y	1 600	2 000	1 450	2 200	1 400	1 800	1 350

Решение

1. МНК–оценки параметров a_0, a_1 и a_2 можно получить, решив систему нормальных уравнений, которая в данном случае имеет вид[6]:

$$\begin{cases} a_0 N + a_1 \sum_{i=1}^{N} x_1^i + a_2 \sum_{i=1}^{N} x_2^i = \sum_{i=1}^{N} y^i \\ a_0 \sum_{i=1}^{N} x_1^i + a_1 \sum_{i=1}^{N} (x_1^i)^2 + a_2 \sum_{i=1}^{N} x_1^i \cdot x_2^i = \sum_{i=1}^{N} y^i \cdot x_1^i \\ a_0 \sum_{i=1}^{N} x_2^i + a_1 \sum_{i=1}^{N} x_1^i \cdot x_2^i + a_2 \sum_{i=1}^{N} (x_2^i)^2 = \sum_{i=1}^{N} y^i \cdot x_2^i \end{cases} \qquad (3.1)$$

Для удобства, сделаем промежуточные расчёты и внесём их в таблицу 3.1.

---

6 Гусаров В.М. Статистика: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001. – с. 202.

Таблица 3.1

I	1	2	3	4	5	6	7
y	1 600	2 000	1 450	2 200	1 400	1 800	1 350
сум. y	**11 800**						
x_1	35	45	25	55	30	42	25
сум. x_1	**257**						
x_1^2	1 225	2 025	625	3 025	900	1 764	625
сум. x_1^2	**10 189**						
x_2	5	10	3	12	1	8	2
сум. x_2	**41**						
x_2^2	25	100	9	144	1	64	4
сум. x_2^2	**347**						
$x_1 \cdot x_2$	175	450	75	660	30	336	50
сум. $x_1 \cdot x_2$	**1 776**						
$y \cdot x_1$	56 000	90 000	36 250	121 000	42 000	75 600	33 750
сум. $y \cdot x_1$	**454 600**						
$y \cdot x_2$	8 000	20 000	4 350	26 400	1 400	14 400	2 700
сум. $y \cdot x_2$	**77 250**						

Подставив данные из табл. 3.1 в (3.1) получим систему уравнений:

$$\begin{cases} a_0 \cdot 7 + a_1 \cdot 257 + a_2 \cdot 41 = 11\,800 \\ a_0 \cdot 257 + a_1 \cdot 10\,189 + a_2 \cdot 1\,776 = 454\,600 \\ a_0 \cdot 41 + a_1 \cdot 1\,776 + a_2 \cdot 347 = 77\,250 \end{cases}$$

решив которую с помощью пакета MathCAD 2001 Professional получим:

1. Матрица системы

$$A := \begin{pmatrix} 7 & 257 & 41 \\ 257 & 10189 & 1776 \\ 41 & 1776 & 347 \end{pmatrix}$$

Матрица правой части

$$b := \begin{pmatrix} 11800 \\ 454600 \\ 77250 \end{pmatrix}$$

2. Вычисление определеителя

$$|A| = 5.056 \times 10^4$$

Определеитель отличен от нуля,
система имеет единственное решение

3. Вычисление решения системы

$$a := A^{-1} \cdot b$$

$$a = \begin{pmatrix} 993.741 \\ 11.247 \\ 47.643 \end{pmatrix}$$

Проверка правильности решения

$$A \cdot a - b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Таким образом уравнение регрессии примет вид:

$$y = 993{,}74 + 11{,}25x_1 + 47{,}64x_2$$

2. Проверим значимость полученного уравнения регрессии по критерию Фишера. Расчётный критерий Фишера для нашей выборки равен[7]:

$$F_{расч} = \frac{\sigma_y^2}{\sigma_{ост}^2} \cdot \frac{(n-m)}{m-1}, \qquad (3.2)$$

где m – число параметров регрессии, в нашем случае $m = 2$.

Произведём промежуточные расчёты и заполним таблицу 3.2

Таблица 3.2

I	1	2	3	4	5	6	7	**ср.**
y	1 600	2 000	1 450	2 200	1 400	1 800	1 350	**1685,71**
$y_{оц}$	1 626	1 976	1 418	2 184	1 379	1 847	1 370	
x_1	35	45	25	55	30	42	25	**36,71**
x_2	5	10	3	12	1	8	2	**5,86**
$(y - y_{оц})^2$	660	557	1 030	251	446	2 243	411	
сумма	**5 598**							
$(y - y_{ср})^2$	7 347	98 776	55 561	264 490	81 633	13 061	112 704	
сумма	**633 571**							
$(x_1 - x_{1ср})^2$	3	69	137	334	45	28	137	
сумма	**753**							
$(x_2 - x_{2ср})^2$	1	17	8	38	24	5	15	
сумма	**107**							

По формулам (1.5) и (1.8) вычислим:

$$\hat{\sigma}_{ост} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \hat{y})^2} = \sqrt{\frac{5598}{7}} = 28{,}28$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2} = \sqrt{\frac{633571}{7}} = 300{,}85$$

Подставим полученные данные в (3.2) и вычислим расчётный критерий Фишера:

---

[7] Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – с. 51–54.

$$F_{расч} = \frac{\sigma_y^2}{\sigma_{ост}^2} \cdot \frac{(n-m)}{m-1} = \frac{300{,}85^2}{28{,}28^2} \cdot \frac{(7-2)}{2-1} = 565{,}86$$

По таблице распределения Фишера с $\nu_1 = 1$ и $\nu_2 = 5$ найдём критической значение $F_{кр} = 6{,}61$. Поскольку $F_{расч} > F_{кр}$, то гипотеза H_0 отвергается, т.е. полученное уравнение регрессии статистически значимо.

Рассчитаем среднеквадратичные отклонения.

$$\hat{\sigma}_{x1} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_{1i} - \bar{x}_1)^2} = \sqrt{\frac{753}{7}} = 10{,}37$$

$$\hat{\sigma}_{x2} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_{2i} - \bar{x}_2)^2} = \sqrt{\frac{107}{7}} = 3{,}91$$

Для измерения тесноты связи между двумя из рассматриваемых переменных (без учёта их взаимодействия с другими переменными) применяют парные коэффициенты корреляции. Если известны средние квадратичные отклонения анализируемых величин, то **парные коэффициенты корреляции** можно рассчитать по следующим формулам[8]:

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_{x_1} \sigma_y} \qquad (3.3)$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_{x_2} \sigma_y} \qquad (3.4)$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}} \qquad (3.5)$$

Предварительно вычислим среднеквадратические отклонения:

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} \qquad (3.6)$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^{\,2}} - \bar{x}_1^{\,2}} \qquad (3.7)$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^{\,2}} - \bar{x}_2^{\,2}} \qquad (3.8)$$

Для этого произведём промежуточные расчёты и заполним таблицу 3.3.

8 Гусаров В.М. Статистика: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001. – с. 205

Таблица 3.3

I	1	2	3	4	5	6	7	ср.
y	1 600	2 000	1 450	2 200	1 400	1 800	1 350	**1685,71**
y^2	2 560 000	4 000 000	2 102 500	4 840 000	1 960 000	3 240 000	1 822 500	**2 932 142,86**
x_1	35	45	25	55	30	42	25	**36,71**
x_1^2	1 225	2 025	625	3 025	900	1 764	625	**1455,57**
x_2	5	10	3	12	1	8	2	**5,86**
x_2^2	25	100	9	144	1	64	4	**49,57**
$x_1 y$	56 000	90 000	36 250	121 000	42 000	75 600	33 750	**64942,86**
$x_2 y$	8 000	20 000	4 350	26 400	1 400	14 400	2 700	**11035,71**
$x_1 x_2$	175	450	75	660	30	336	50	**253,71**

Подставим данные из табл. 3.3 в (3.6) – (3.8) и получим:

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{2\,932\,143 - 1686^2} = 300{,}85$$

$$\sigma_{x_1} = \sqrt{\overline{x_1}^2 - \bar{x}_1^{\,2}} = \sqrt{1455{,}57 - 36{,}71^2} = 10{,}39$$

$$\sigma_{x_2} = \sqrt{\overline{x_2}^2 - \bar{x}_2^{\,2}} = \sqrt{49{,}57 - 5{,}86^2} = 3{,}90$$

А теперь по (3.3) – (3.5) рассчитаем парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_{x_1}\sigma_y} = \frac{64942{,}86 - 36{,}71 \cdot 1685{,}71}{10{,}39 \cdot 299{,}24} = 0{,}979$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_{x_2}\sigma_y} = \frac{11035{,}71 - 5{,}86 \cdot 1685{,}71}{3{,}90 \cdot 299{,}24} = 0.986$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1}\sigma_{x_2}} = \frac{253{,}71 - 36{,}71 \cdot 5{,}86}{10{,}39 \cdot 3{,}90} = 0{,}952$$

В реальных условиях все переменные, как правило, взаимосвязаны. Теснота этой связи определяется частными коэффициентами корреляции, которые характеризуют степень и влияние одного из аргументов на функцию при условии, что остальные независимые переменные закреплены на постоянном уровне. В зависимости от количества переменных, влияние которых исключается, частные коэффициенты корреляции могут быть различного порядка: при исключении влияния одной переменной получаем частный коэффициент корреляции первого порядка; при исключении влияния двух переменных – второго порядка и т.д. Парный коэффициент корреляции между функцией и аргументом обычно не равен соответствующему частному коэффициенту.

Частный коэффициент корреляции первого порядка между признаками x_1 и y при исключении влияния признака x_2 вычисляется по формуле[9]:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} \quad ; \tag{3.9}$$

аналогично – зависимость y от x_2 при исключении влияния признака x_1 :

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} \tag{3.10}$$

Можно рассчитать взаимосвязь факторных признаков при устранении влияния результативного признака:

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2}\right)}} \tag{3.11}$$

Выполним расчёт:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} = \frac{0{,}979 - 0{,}986 \cdot 0{,}952}{\sqrt{(1 - 0{,}986^2) \cdot (1 - 0{,}952^2)}} = 0{,}793$$

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} = \frac{0{,}986 - 0{,}979 \cdot 0{,}952}{\sqrt{\left(1 - 0{,}979^2\right) \cdot \left(1 - 0{,}952^2\right)}} = 0{,}871$$

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2}\right)}} = \frac{0{,}952 - 0.979 \cdot 0{,}986}{\sqrt{\left(1 - 0{,}979^2\right) \cdot \left(1 - 0{,}986^2\right)}} = -0.405$$

Сравнивая коэффициенты $r_{yx_1(x_2)}$ и $r_{yx_2(x_1)}$ можно сделать вывод, что влияние стажа (x_2) на зарплату выше чем влияние возраста (x_1).

3. Показателем тесноты связи, устанавливаемой между результативными и двумя или более факторными признаками, является **совокупный коэффициент множественной корреляции**. В случае линейной двухфакторной

---

9 Гусаров В.М. Статистика: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001. – с. 206.

связи (как в нашей задаче) совокупный коэффициент множественной корреляции может быть рассчитан по формуле[10]:

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}} \qquad (3.12)$$

Совокупный коэффициент множественной корреляции измеряет одновременное влияние факторных признаков на результативный. Чем меньше наблюдаемые значения изучаемого показателя отклоняются от линии множественной регрессии, тем корреляционная связь является более интенсивной, а следовательно значение R ближе к единице.

Выполним расчёт:

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}} = \sqrt{\frac{0{,}979^2 + 0{,}986^2 - 2 \cdot 0{,}979 \cdot 0{,}986 \cdot 0{,}952}{1 - 0{,}952^2}} = 0{,}995$$

Совокупный коэффициент множественной детерминации $R^2 = 0{,}99$ показывает, что вариация заработной платы на 99 % обуславливается двумя анализируемыми факторами. Это означает, что выбранные факторы существенно влияют на заработную плату.

---

[10] Гусаров В.М. Статистика: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001. – с. 207

Приложения

Приложение 1. Квантили распределения Стьюдента[11]

	0,75	0,875	0,95	0,975	0,99	0,995
1	1,00	2,41	6,31	12,7	31,82	63,7
2	0,816	1,60	2,92	4,30	6,97	9,92
3	0,765	1,42	2,35	3,18	4,54	5,84
4	0,741	1,34	2,13	2,78	3,75	4,60
5	0,727	1,30	2,01	2,57	3,37	4,03
6	0,718	1,27	1,94	2,45	3,14	3,71
7	0,711	1,25	1,89	2,36	3,00	3,50
8	0,706	1,24	1,86	2,31	2,90	3,36
9	0,703	1,23	1,83	2,26	2,82	3,25
10	0,700	1,22	1,81	2,23	2,76	3,17
11	0,697	1,21	1,80	2,20	2,72	3,11
12	0,695	1,21	1,78	2,18	2,68	3,05
13	0,694	1,20	1,77	2,16	2,65	3,01
14	0,692	1,20	1,76	2,14	2,62	2,98
15	0,691	1,20	1,75	2,13	2,60	2,95
16	0,690	1,19	1,75	2,12	2,58	2,92
17	0,689	1,19	1,74	2,11	2,57	2,90
18	0,688	1,19	1,73	2,10	2,55	2,88
19	0,688	1,19	1,73	2,09	2,54	2,86
20	0,687	1,18	1,73	2,09	2,53	2,85
25	0,684	1,18	1,71	2,06	2,49	2,79
30	0,683	1,17	1,70	2,04	2,46	2,75
40	0,681	1,17	1,68	2,02	2,42	2,70
60	0,679	1,16	1,67	2,00	2,39	2,66
∞	0,674	1,15	1,64	1,96	2,33	2,58

[11] Таблица приводится по источнику: Дерффель К. Статистика в аналитической химии. – М.: Мир, 1994.

Литература

1. Гусаров В.М. Статистика: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2001. – 464 с.
2. Дрейпер Н., Смит Г. Прикладной регрессивный анализ. Книга 1. – М.: Финансы и статистика, 1986.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. – М., 1997.
4. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. – М.: Финансы и статистика, 1982.
5. Новиков А.И. Эконометрика: Учебное пособие. – М.: ИНФРА-М, 2003. – 106 с. – (Серия «Высшее образование»).
6. Носко В.П. Эконометрика для начинающих: основные понятия, элементарные методы, границы применимости, интерпретация результатов. – М.: ИЭПП, 2000 – 252 с.
7. Программа регрессионного анализа STATISTICA http://www.statsoft.ru
8. Степанов А.В. Лекции по математике. (в формате MS-Word и Adobe Acrobat) http://www.limm.mgimo.ru/LIMM/Lections/
9. Цейтлин Н. А. Из опыта аналитического статистика. (в формате Adobe Acrobat) http://people.freenet.de/nzarchiv/buecher/