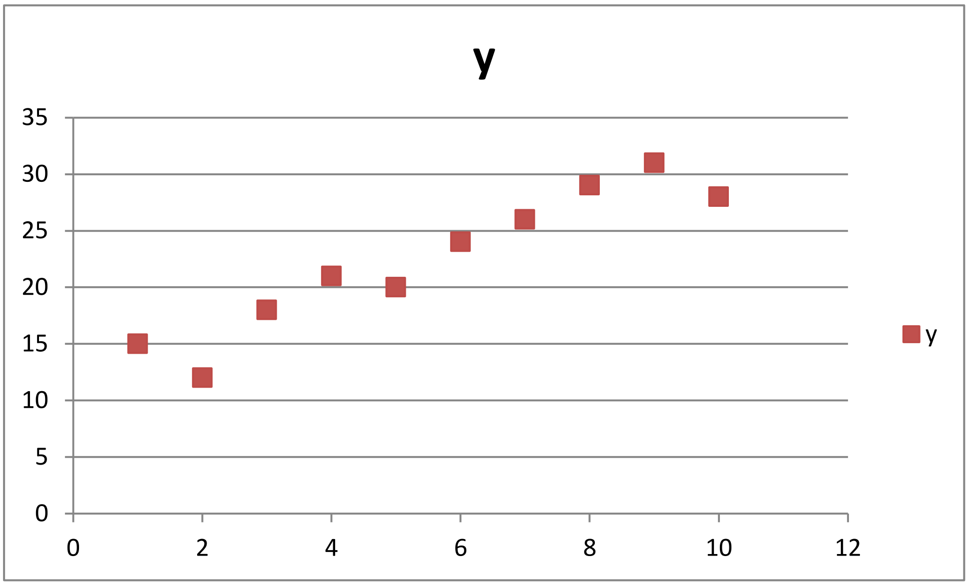
Вариант 8

Задача 1.Имеются следующие данные:

Номер семьи	1	2	3	4	5	6	7	8	9	10
Число совместно проживающих членов семьи, чел.	3	3	2	4	4	4	5	6	7	7
Годовое потребление электроэнергии, тыс. кв.-час	15	12	18	21	20	24	26	29	31	28

1. **Постройте поле корреляции результата и фактора и сформулируйте гипотезу о форме связи.**

Y – годовое потребление электроэнергии – результат, X – число совместно проживающих членов семьи – это фактор, тогда получаем поле корреляции:



По разбросу значений видно, что что связь результат и фактора имеет линейную зависимость - $y = a + bx$.

2. Оцените параметры уравнений парной регрессии и дайте интерпретацию коэффициента регрессии *b*. Рассчитайте линейный коэффициент корреляции и поясните его смысл. Определите коэффициент детерминации и дайте его интерпретацию.

Для оценки коэффициентов регрессии необходимо решить систему:

$$\begin{cases} na + b\sum x_i = \sum y_i; \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i. \end{cases}$$

Для удобства результаты промежуточных вычислений поместим во вспомогательную таблицу:

№	Y	X	X^2	XY	X_i-$X_{ср}$	Y_i-$Y_{ср}$	$(X_i$-$X_{ср})$*$(Y_i$-$Y_{ср})$	$(X_i$-$X_{ср})^2$	$(Y_i$-$Y_{ср})^2$	Yпр	Y_i-$Y_{пр}$	$(Y_i$-$Y_{пр})^2$
1	15	3	99	45	-1,5	-7.4	11.1	2.25	54.76	17.4	-2.4	5.76
2	12	3	99	36	-1,5	-10.4	15.6	2.25	108.16	17.4	-5.4	29.16
3	18	2	44	36	-2,5	-4.4	11	6.25	19.36	14.2	3.8	14.44
4	21	4	116	84	-0,5	-1.4	0.7	0.25	1.96	20.6	0.4	0.16
5	20	4	116	80	-0,5	-2.4	1.2	0.25	5.76	20.6	-0.6	0.36
6	24	4	116	96	-0,5	1.6	-0.8	0.25	2.56	20.6	3.4	11.56
7	26	5	225	130	0,5	3.6	1.8	0.25	12.96	23.8	2.2	4.84
8	29	6	336	174	1,5	6.6	9.9	2.25	43.56	27	2	4
9	31	7	449	217	2,5	8.6	21.5	6.25	73.96	30.2	0.8	0.64

10	28	7	449	196	2,5	5.6	14	6.25	31.36	30.2	-2.2	4.84
Суммы	224	45	229	1094	-	-	86	26.5	354.4	222	2	75.76
Средние	22.4	4.5	22.9	109.4	-	-	-	-	-	22.2	0.2	7.6

Находим оценки параметров a, b, которые получаются в результате решения системы нормальных уравнений:

$$10a + 45b = 224;$$
$$45a + 229b = 1094.$$

Следовательно, оценки параметров регрессии будут равны: $a = 7.8, b = 3.2$. Таким образом, получили оценку уравнения регрессии:$y = 7.8 + 3.2x$.

Коэффициент b означает, что при увеличении числа совместно проживающих членов семьи на 1 человека годовое потребление электроэнергии увеличится на 3.2 тысячи кв/час.

Определим коэффициент корреляции по формуле$r_{xy} = \frac{cov(x,y)}{s_x s_y}$, где$cov\ x,y\ = \frac{x_i-\bar{x}\ (y_i-\bar{y})}{n}$ – ковариация, а $s_x = \sqrt{\frac{1}{n-1}\ (x_i-\bar{x})^2}, s_y = \sqrt{\frac{1}{n-1}\ (y_i-\bar{y})^2}$ - средние квадратические отклонения. Необходимые суммы найдены в таблице выше, откуда получаем: $cov\ x,y\ = \frac{86}{10} = 8{,}6$; $s_x = \sqrt{\frac{1}{9} * 26{,}5} \approx 1{,}7, s_y = \sqrt{\frac{1}{9} * 354{,}4} \approx 6{,}3$. Таким образом, можем вычислить коэффициент корреляции: $r_{xy} = \frac{cov(x,y)}{s_x s_y} = \frac{8{,}6}{1{,}7*6{,}3} \approx 0{,}8$.

Коэффициент корреляции изменяется от –1 до +1 и характеризует относительную силу связи между переменными, образующими двумерную выборку. Поскольку коэффициент положителен, то связь между числом совместно проживающих членов семьи и годовым потреблением электроэнергии прямая. А так как r_{xy} по абсолютной величине близок к единице, то связь между ними также очень тесная.

Коэффициентом детерминации называется коэффициент:$R^2 = 1 - \frac{(y_i - y_{пр})^2}{(y_i-\bar{y})^2}$, где $y_{пр}$- это прогнозные значения, полученные подстановкой значений переменной x в полученное уравнение регрессии. Необходимые для вычисления коэффициента детерминации данные приведены в таблице выше.

Получаем: $R^2 = 1 - \frac{(y_i - y_{пр})^2}{(y_i-\bar{y})^2} = 1 - \frac{75.76}{354.4} \approx 0.79$. Так как R^2 практически близок к 1, то линия регрессии соответствует всем наблюдениям, это означает, что 79% вариации годового потребления электроэнергии (y) объясняется вариацией фактора x – числа совместно проживающих членов семьи.

3. На уровне значимости 0,05 оцените статистическую значимость коэффициента *b* и коэффициента корреляции. Сделайте выводы.

Оценку статистической значимости параметра b регрессии проведем с помощью t -статистики Стьюдента. Табличное значение t -критерия для числа степеней свободы $df = n - 2 = 10 - 2 = 8$ и уровня значимости α = 0,05 составит $t_{табл} = 2{,}31$. Стандартная ошибка регресии рассчитывается по формуле:$m_b = \sqrt{\frac{y_i - y_{пр}\ ^2}{(n-2)\ \ x_i-\bar{x}\ ^2}} = \sqrt{\frac{75{,}76}{8*26{,}5}} \approx 0.6$. Тогда $t_b = \frac{b}{m_b} = \frac{3.2}{0.6} \approx 5.3$. Получаем: $t_b = 5.3 > t_{крит} = 2{,}31$, следовательно, **параметрbзначим** в уравнении регрессии.

Наблюдаемое значение коэффициента корреляции находим по формуле:

$$t_{набл} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0.8 * \sqrt{\frac{8}{1-0.8^2}} \approx 3.8.$$

Так как$t_{набл}$=3,8 >$t_{крит}$=2,31, то полученное значение **коэффициента корреляции** признается **значимым** (нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается).

4. На уровне значимости 0,05 оцените статистическую значимость уравнения регрессии в целом.

Оценку статистической значимости уравнения регрессии в целом проведем с помощью F -критерия Фишера. Фактическое значение F -критерия составит:$F_{факт} = \frac{R^2}{1-R^2}\cdot\ n-2\ = \frac{0.79}{1-0.79}\cdot 8 \approx 30$. Табличное значение критерия при пятипроцентном уровне значимости и степенях свободы $k_1 = 1$ и $k_2 = n - 2 = 10 - 2 = 8$ составляет $F_{крит} = 5{,}32$. Так как $F_{факт} = 30 > F_{крит} = 5{,}32$, то **уравнение регрессии** признается **статистически значимым**.

5. На уровне значимости 0,05 проверьте гипотезу о гетероскедастичности остатков модели с помощью критерия Спирмена.

Упорядочим по величинам X_i и остатки $E_i = Y_i - Y_{пр}$, и вычислим разность между рангами X_i и E_i, получаем:

Номер	3	1	2	4	5	6	7	8	9	10
X_i	2	3	3	4	4	4	5	6	7	7
Номер	2	1	10	5	4	9	8	7	6	3
E_i	-5.4	-2.4	-2.2	-0.6	0.4	0.8	2	2.2	3.4	3.8
Номер	1	2	3	4	5	6	7	8	9	10
D_i	0	2	-9	-1	1	-3	-1	1	3	7
D_i^2	0	4	81	1	1	9	1	1	9	49

Теперь найдем коэффициент ранговой корреляции по формуле:

$$r_{x,e} = 1 - 6 * \frac{D_i^2}{n\ n^2 - 1} = 1 - 6 * \frac{156}{10 * 99} \approx 0.05.$$

Теперь найдем статистику:

$$t_{набл} = r_{x,e}\sqrt{\frac{n-2}{1-r_{x,e}^2}} = 0.05 * \sqrt{\frac{8}{1-0.05^2}} \approx 1.14.$$

Знаем, что табличное значение t -критерия для числа степеней свободы $df = n - 2 = 10 - 2 = 8$ и уровня значимости α = 0,05 составляет$t_{табл} = 2{,}31$. И так как как$t_{набл}$=1,14 <$t_{крит}$=2,31,следовательно, гипотеза об отсутствиигетероскедастичности принимается.

6. На уровне значимости 0,1 проверьте предположение об автокорреляции остатков.

Построим вспомогательную таблицу:

№	Y	$e_t = Y_i - Y_{пр}$	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
1	15	-2.4	-	-	-	5.8
2	12	-5.4	-2.4	-3	9	29.2
3	18	3.8	-5.4	9.2	84.6	14.4
4	21	0.4	3.8	-3.4	11.6	0.2
5	20	-0.6	0.4	-1	1	0.4
6	24	3.4	-0.6	4	16	11.6
7	26	2.2	3.4	-1.2	1.4	4.8
8	29	2	2.2	-0.2	0.04	4.0
9	31	0.8	2	-1.2	1.4	0.6
10	28	-2.2	0.8	-3	9	4.8
Суммы	224	2	4.2	-	134	75.8

Теперь можем рассчитать фактическое значение критерия Дарбина-Уотсона для данной модели:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=2}^{n} e_t^2} = \frac{134}{75.8} \approx 1.77.$$

Сформулируем гипотезы: H_0 – в остатках нет автокорреляции; H_1–в остатках есть положительная автокорреляция; H_1^* - в остатках есть отрицательная автокорреляция. При уровне значимости α = 0.05 и для числа наблюдений n=10 и числа независимых параметров k=1 по таблице можем найти критические значения критерия Дарбина-Уотсона: $d_L = 0.88, d_U = 1.32$. Фактическое значение d- критерия Дарбина-Уотсона попадает в интервал $d_U < d < 4 - d_U$ $(1.32 < 1.77 < 2.68$.Следовательно, нет основания отклонять гипотезу H_0об отсутствии автокорреляции в остатках.

7. С вероятностью 0,9 постройте доверительный интервал ожидаемого значения результативного признака, если факторный признак увеличится на 10 % от своего среднего значения.

Интервальная оценка для уравнения регрессии при x_0= 4,5*1,1 = 4,95 вычисляется так:

$$Y \in \left(a + bx_0 \pm t_{крит} S\sqrt{\frac{1}{n} + \frac{(x_0 - x_{ср})^2}{\sum(x_i - x_{ср})^2}}\right), \text{где } S = \sqrt{\frac{\sum(Y_i - Y_{пр})^2}{n-2}} = \sqrt{\frac{75.76}{8}} \approx 3.1; \text{а}t_{крит} = 1.4 \text{ при } \alpha = 0.1$$

$$Y \in \left(7{,}8 + 3{,}2 * 4{,}95 \pm 1.4 * 3.1\sqrt{\frac{1}{10} + \frac{(4.95 - 4.5)^2}{26.5}}\right);$$

$$Y \in \ 23.64 \pm 1.42 \ ; Y \in \ 22.22; 25.06 \ .$$

Следовательно, результативный признак Y попадет в интервал [22,22; 25,06] при факторном признаке, увеличенном на 10% от своего среднего значения, с вероятностью 0.9.

Задача 2. По выборке из 10 почтовых отправлений изучается зависимость стоимости отправки корреспонденции экспресс-почтой от веса конверта и дальности перевозки:

№	Стоимость доставки, тыс. руб.	Вес конверта, г	Дальность перевозки, тыс. км
1	**26**	**590**	**0,5**
2	**39**	**320**	**1,5**
3	**80**	**440**	**2**
4	**92**	**660**	**1,6**
5	**44**	**75**	**2,8**
6	**15**	**70**	**0,8**
7	**145**	**650**	**2,4**
8	**19**	**450**	**0,5**
9	**10**	**60**	**1**
10	**140**	**750**	**1,8**

1. Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.

Линейное уравнение множественной регрессии для двух переменных имеет вид - $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i}$.Для оценки коэффициентов регрессии необходимо решить систему:

$$n * b_0 + b_1 \quad x_{1i} + b_2 \quad x_{2i} = \quad y_i \ ;$$

$$b_0 \quad x_{1i} + b_1 \quad x_{1i}^2 + b_2 \quad x_{1i} * x_{2i} = \quad y_i * x_{1i} \ ;$$

$$b_0 \quad x_{2i} + b_1 \quad x_{1i} * x_{2i} + b_2 \quad x_{2i}^2 = \quad y_i * x_{2i} \ .$$

Для удобства результаты промежуточных вычислений поместим их в вспомогательную таблицу:

№	Y	X_1	X_2	X_1^2	X_2^2	$X_1 * X_2$	$Y * X_1$	$Y * X_2$	$(x_{1i} - \overline{x_1})^2$	$(x_{2i} - \overline{x_2})^2$	$(y_i - \overline{y})^2$
1	26	590	0.50	348100	0.25	295	15340	13	33672.3	0.98	1225
2	39	320	1.50	102400	2.25	480	12480	58.5	7482.3	0.00	484
3	80	440	2.00	193600	4.00	880	35200	160	1122.3	0.26	361
4	92	660	1.60	435600	2.56	1056	60720	147.2	64262.3	0.01	961
5	44	75	2.80	5625	7.84	210	3300	123.2	109892.3	1.72	289
6	15	70	0.80	4900	0.64	56	1050	12	113232.3	0.48	2116
7	145	650	2.40	422500	5.76	1560	94250	348	59292.3	0.83	7056
8	19	450	0.50	202500	0.25	225	8550	9.5	1892.3	0.98	1764
9	10	60	1.00	3600	1.00	60	600	10	120062.3	0.24	2601
10	140	750	1.80	562500	3.24	1350	105000	252	117992.3	0.10	6241
Суммы	610	4065	14.90	2281325	27.79	6172	336490	1133.4	628902.5	5.59	23098
Средние	61	406.5	1.49	-	-	-	-	-	-	-	-

Находим оценки параметров, которые получаются в результате решения системы нормальных уравнений:

$$10b_0 + 4065b_1 + 14{,}9b_2 = 610;$$
$$4065b_0 + 2281325b_1 + 6172b_2 = 336490;$$
$$14{,}9b_0 + 6172b_1 + 27{,}79b_2 = 1133{,}4.$$

Следовательно, оценки параметров регрессии будут равны: $b_0 = -49.2, b_1 = 0.13, b_2 = 37.4$. Таким образом, получили оценку уравнения регрессии:$y_i = -49.2 + 0.13x_{1i} + 37.4x_{2i}$.

Полученные коэффициенты характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне. То есть, если вес конверта x_1 увеличиться на 1 г при неизменном среднем значении дальности перевозки x_2 = 1,49 тыс.км, то стоимость доставкиУ увеличиться на 0,13тыс.руб.. И, аналогично, если дальность доставки x2 увеличиться на 1 тыс.км при неизменном среднем значении веса конверта x1 = 406,5г, то стоимость перевозкиУувеличиться на 37,4 тыс.руб.

2. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции, сделайте выводы.

Парные коэффициенты корреляции, измеряющие влияние на y фактора x_i , при исключении влияния других факторов, можно определить по формуле: $r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{1-r_{yx_2}^2 \cdot 1-r_{x_1x_2}^2}}$; $r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{1-r_{yx_1}^2 \cdot 1-r_{x_1x_2}^2}}$ и влияние факторных признаков друг на друга при фиксированном признаке y: $r_{yx_1 \cdot x_2} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{1-r_{yx_2}^2 \cdot 1-r_{yx_1}^2}}$, где коэффициенты корреляции вычисляются по формулам$r_{yx_i} = \frac{cov(x_i, y)}{S_{xi} S_y}$, где$cov\ x_i, y\ = \frac{x_{ij} - \overline{x_i}\ (y_j - \overline{y})}{n}$ – ковариация, а $s_{xi} = \sqrt{\frac{1}{n-1}\ (x_{ij} - \overline{x_i})^2}$, $s_y = \sqrt{\frac{1}{n-1}\ (y_j - \overline{y})^2}$ - средние квадратические отклонения. Необходимые суммы внесем в таблицу:

№	Y	X_1	X_2	$(x_{1i} - \overline{x_1})^2$	$(x_{2i} - \overline{x_2})^2$	$(y_i - \overline{y})^2$	$(x_{1i} - \overline{x_1})(y_i - \overline{y})$	$(x_{2i} - \overline{x_2})(y_i - \overline{y})$	$(x_{1i} - \overline{x_1})(x_{2i} - \overline{x_2})$
1	26	590	0.50	33672.3	0.98	1225	-6422.5	34.65	-181.67
2	39	320	1.50	7482.3	0.00	484	1903.0	-0.22	-0.87
3	80	440	2.00	1122.3	0.26	361	636.5	9.69	17.09
4	92	660	1.60	64262.3	0.01	961	7858.5	3.41	27.89
5	44	75	2.80	109892.3	1.72	289	5635.5	-22.27	-434.27
6	15	70	0.80	113232.3	0.48	2116	15479.0	31.74	232.19
7	145	650	2.40	59292.3	0.83	7056	20454.0	76.44	221.59
8	19	450	0.50	1892.3	0.98	1764	-1827.0	41.58	-43.07
9	10	60	1.00	120062.3	0.24	2601	17671.5	24.99	169.79
10	140	750	1.80	117992.3	0.10	6241	27136.5	24.49	106.49
Суммы	610	4065	14.90	628902.5	5.59	23098	88525.0	224.50	115.15
Средние	61	406.5	1.49	-	-	-	-	-	-

Откуда получаем: $cov\ x_1, y\ = \frac{88525}{10} \approx 8852{,}5$; $cov\ x_2, y\ = \frac{224{,}5}{10} \approx 22{,}5$; $s_{x1} = \sqrt{\frac{1}{9} * 628902{,}5} \approx 264{,}3$; $s_{x2} = \sqrt{\frac{1}{9} * 5.59} \approx 0.79$; $s_y = \sqrt{\frac{1}{9} * 23098} \approx 50.7$. Таким образом, можем вычислить коэффициенты корреляции: $r_{yx_1} = \frac{8852.5}{264.3*50.7} \approx 0.66$, $r_{yx_2} = \frac{22.5}{0.79*50.7} \approx 0.56$. Аналогично находим $r_{x_1x_2} = \frac{cov(x_1, x_2)}{S_{x1} S_{x2}}$, где$cov\ x_1, x_2\ = \frac{x_{1i} - \overline{x_1}\ \ x_{2i} - \overline{x_2}}{n}$. Откуда: $cov\ x_1, x_2\ = \frac{115{,}15}{10} \approx 11{,}5$; $r_{x_1x_2} = \frac{cov\ x_1, x_2}{S_{x1} S_{x2}} = \frac{11{,}5}{264{,}3*0{,}79} \approx 0{,}06$.

Теперь можем вычислить парные коэффициенты корреляции:

$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{1-r_{yx_2}^2 \cdot 1-r_{x_1x_2}^2}} = \frac{0{,}66-0{,}56*0{,}06}{\sqrt{(1-0{,}56^2)(1-0{,}06^2)}} \approx 0{,}76$; $r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{1-r_{yx_1}^2 \cdot 1-r_{x_1x_2}^2}} = \frac{0{,}56-0{,}66*0{,}06}{\sqrt{1-0{,}66^2\ \ 1-0{,}06^2}} \approx 0{,}69$; $r_{yx_1 \cdot x_2} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{1-r_{yx_2}^2 \cdot 1-r_{yx_1}^2}} = \frac{0{,}06-0{,}66*0{,}56}{\sqrt{1-0{,}56^2\ \ 1-0{,}66^2}} \approx -0{,}5$.

На основании частных коэффициентов корреляции можносделать вывод об обоснованности включения переменной в регрессионную модель: так как частные коэффициенты $r_{yx_1 \cdot x_2}, r_{yx_2 \cdot x_1}$ близки к 1, следовательно, связь между обоими факторами и результативной переменной сильна и никакой из признаков исключать из модели не стоит.

Множественный коэффициент корреляции определяется через матрицы парных коэффициентов корреляции. Получаем: $R_{yx_1x_2} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}$, где $\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_2x_1} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0{,}66 & 0{,}56 \\ 0{,}66 & 1 & 0{,}06 \\ 0{,}56 & 0{,}06 & 1 \end{vmatrix} \approx 0{,}29$, а $\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_2x_1} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0{,}06 \\ 0{,}06 & 1 \end{vmatrix} = 0.9964$. Откуда множественный коэффициент корреляции равен: $R_{yx_1x_2} = \sqrt{1 - \frac{0{,}29}{0{,}9964}} \approx 0.84$. Данный коэффициент указывает на весьма сильную связь всего набора факторов с результатом.

3. Определите коэффициенты эластичности и стандартизованные коэффициенты регрессии. Сделайте выводы.

Коэффициенты эластичности для линейной регрессии рассчитываются по формуле$\overline{Э}_{yx_j} = b_j \cdot \frac{\overline{x_j}}{\overline{y}}$. Откуда получаем: $\overline{Э}_{yx_1} = b_1 \cdot \frac{\overline{x_1}}{\overline{y}} = 0{,}13 \cdot \frac{406{,}5}{61} \approx 0{,}87$; $\overline{Э}_{yx_2} = b_2 \cdot \frac{\overline{x_2}}{\overline{y}} = 37{,}4 \cdot \frac{1{,}49}{61} \approx 0{,}91$.Частный коэффициент эластичности отражает процентноеизменение результативного признака при изменении на 1% от среднего уровня факторного признака x_i, если остальные переменные, участвующие в модели, зафиксированы. Таким образом, по нашим коэффициентам получаем: стоимость доставки увеличиться на 0,87% при изменении веса конверта на 1% от среднего значения при фиксированном значении дальности перевозки, и стоимость доставки увеличиться на 0,91% при изменении дальности перевозки на 1% от среднего значения при фиксированном значении веса конверта.

Определим стандартизированные коэффициенты регрессии: $\beta_1 = b_1 \frac{S_{x1}}{S_y} = 0.13 * \frac{264.3}{50.7} \approx 0.68; \beta_2 = b_2 \frac{S_{x2}}{S_y} = 37.4 * \frac{0.79}{50.7} \approx 0.58$.Так как стандартизированный частный регрессионный коэффициентпоказывает степень непосредственной или прямой связи междурезультативным и факторным признаками, то видим, что между результативный признаком (стоимость доставки) и обоими факторными признаками существует прямая и довольно тесная связь.

4. На уровне значимости 0,05 оцените статистическую значимость уравнения регрессии в целом.

Оценку надежности уравнения регрессии в целом дает F-критерий Фишера: $F_{факт} = \frac{R^2_{yx_1x_2}}{1-R^2_{yx_1x_2}} \cdot \frac{n-m-1}{m}$, где $R^2_{yx_1x_2}$ – множественный коэффициент корреляции, вычисленный. Вычисляем: $F_{факт} = \frac{R^2_{yx_1x_2}}{1-R^2_{yx_1x_2}} \cdot \frac{n-m-1}{m} = \frac{0{,}84^2}{1-0{,}84^2} * \frac{10-2-1}{2} \approx 8{,}4$. Получили, что $F_{факт} = 8{,}4 > F_{табл} = 3.94$(при n = 10 и α=0,05), то есть вероятность случайно получить такое значение F-критерия не превышает допустимый уровень значимости 5%. Следовательно, полученное значение не случайно и статистическая значимость всего уравнения подтверждается.

Задача 3. Ниже приводится макроэкономическая модель, характеризующая экономику:

$\boldsymbol{C_t = a_0 + a_1 Y_t + a_2 J_t + u_1}$

$\boldsymbol{J_t = b_0 + b_1 Y_{t-1} + u_2}$

$\boldsymbol{T_t = c_0 + c_1 Y_t + u_3}$

$\boldsymbol{Y_t = C_t + J_t + G_t}$,

где C_t –совокупное потребление в период *t*; Y_t, Y_{t-1} –совокупный доход в периоды *t* и *t*-1; J_t – инвестиции в период *t*; T_t налоги в период *t*; G_t – государственные доходы в период *t*; u_1, u_2, u_3 – случайные ошибки.

1. Проверьте с помощью необходимого и достаточного условий идентификации, идентифицирована ли данная модель.

В модели использованы 2 эндогенные переменные –C_t, T_tи три экзогенных переменных – J_t,G_tиY_t. Рассмотрим каждое уравнение.

1ое уравнение.

Необходимое условие: Н: эндогенных – 1 (C_t); D: отсутствующих экзогенных – 1 (G_t). Выполняется условие: H<D + 1; 1< 1 + 1, следовательно, уравнение сверхидентифицируемо.

2ое уравнение.

*Необходимое условие:*Н: эндогенных – 0; D: отсутствующих экзогенных – 1 (G_t). Выполняется условие: H<D + 1; 0< 1 + 1, следовательно, уравнение сверхидентифицируемо.

3ое уравнение.

Необходимое условие: Н: эндогенных – 1(T_t); D: отсутствующих экзогенных – 2 (G_t , J_t). Выполняется условие: H<D + 1; 1<2 + 1, следовательно, уравнение сверхидентифицируемо.

4ое уравнение.

Необходимое условие: Н: эндогенных – 1 (C_t); D: отсутствующих экзогенных – 0. Выполняется условие: H=D + 1; 1=0 + 1, следовательно, уравнение точно идентифицируемо.

Следовательно, исследуемая система одновременных уравнений точно идентифицируема.

2. Выпишите приведенную форму модели.

Мы должны путем подстановок уравнений, определяющих экзогенные переменные, в уравнения эндогенных переменных получить систему, в каждом уравнении которой слева будут эндогенные переменные, а справа экзогенные. Таким образом, в первом уравнении получаем: $C_t = a_0 + a_1C_t + a_1J_t + a_1G_t + a_2b_0 + a_2b_1Y_{t-1} + a_2u_2 + u_1$или$C_t = \frac{a_0+a_2b_0+a_2u_2+u_1+a_1J_t+a_1G_t+a_2b_1Y_{t-1}}{1-a_1}$. Аналогично для T_t получаем: $T_t = c_0 + \frac{c_1(a_0+ a_2b_0+a_2u_2+u_1)}{1-a_1} + \frac{c_1J_t}{1-a_1} + \frac{(2c_1+a_1-c_1a_1)G_t}{1-a_1} + \frac{c_1a_2b_1Y_{t-1}}{1-a_1} + u_3$. То есть получаем приведенную форму модели:

$$C_t = \frac{a_0 +\ a_2b_0 + a_2u_2 + u_1 + a_1J_t + a_1G_t + a_2b_1Y_{t-1}}{1 - a_1};$$

$$T_t = c_0 + \frac{c_1(a_0 +\ a_2b_0 + a_2u_2 + u_1)}{1 - a_1} + \frac{c_1J_t}{1 - a_1} + \frac{(2c_1 + a_1 - c_1a_1)G_t}{1 - a_1} + \frac{c_1a_2b_1Y_{t-1}}{1 - a_1} + u_3.$$

3. Укажите, каким методом вы будете определять структурные параметры каждого уравнения, кратко опишите методику расчета.

Так как ранее выяснили, что все уравнения точно идентифицированы, то их параметры можно оценить косвенным МНК. Для этого структурная форма преобразуется в приведенную форму модели, после чего, обычным МНК оцениваются приведенные коэффициенты. И эти коэффициенты модифицируются в параметры структурной модели.

Задача 5. Изучается зависимость между объемом инвестиций в основные производственные фонды (ОПФ) и валовой добавленной стоимостью (ВДС). Ниже представлены данные по некоторой отрасли промышленности за последние 10 лет (в сопоставимых ценах, млрд. руб.)

Время	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Объем инвестиций в ОПФ	**140**	**160**	**190**	**210**	**220**	**240**	**260**	**290**	**310**	**320**
ВДС	**300**	**345**	**405**	**445**	**480**	**535**	**595**	**639**	**677**	**704**

1. Определите коэффициент корреляции между временными рядами объема инвестиций в ОПФ и валовой добавленной стоимостью по исходным уровням ряда, по первым разностям уровней ряда.

Определим коэффициент корреляции по формуле$r_{xy} = \frac{cov(x,y)}{s_xs_y}$, где $cov\ x,y\ = \frac{x_i-\overline{x}\ (y_i-\overline{y})}{n}$ – ковариация, а $s_x = \sqrt{\frac{1}{n-1}\ \ (x_i - \overline{x})^2}, s_y = \sqrt{\frac{1}{n-1}\ \ (y_i - \overline{y})^2}$ - средние квадратические отклонения.

Данные заносим в таблицу:

№	Y	X	X_i-$X_{ср}$	Y_i-$Y_{ср}$	(X_i-$X_{ср}$)*(Y_i-$Y_{ср}$)	$(X_i-X_{ср})^2$	$(Y_i-Y_{ср})^2$
1	140	300	-212,5	-94	19975	45156,25	8836
2	160	345	-167,5	-74	12395	28056,25	5476
3	190	405	-107,5	-44	4730	11556,25	1936
4	210	445	-67,5	-24	1620	4556,25	576
5	220	480	-32,5	-14	455	1056,25	196
6	240	535	22,5	6	135	506,25	36
7	260	595	82,5	26	2145	6806,25	676
8	290	639	126,5	56	7084	16002,25	3136
9	310	677	164,5	76	12502	27060,25	5776
10	320	704	191,5	86	16469	36672,25	7396
Суммы	2340	5125	-212,5	-94	77510	177428,5	34040
Средние	234	512,5	-	-	-	-	-

Необходимые суммы найдены в таблице выше, откуда получаем: $cov\ x,y\ = \frac{77510}{10} = 7751;\ s_x = \sqrt{\frac{1}{9} * 177428{,}5} \approx 140{,}4, s_y = \sqrt{\frac{1}{9} * 34040} \approx 61{,}5.$ Таким образом, можем вычислить коэффициент корреляции: $r_{xy} = \frac{cov(x,y)}{s_x s_y} = \frac{7751}{140{,}4*61{,}5} \approx 0{,}9.$

Для вычисления коэффициента корреляции по первым разностям уровней ряда имеем формулу:

$$r_1 = \frac{\sum_{t=2}^{n}(y_{t-1} - \overline{y})(x_{t-1} - \overline{x})}{\sqrt{\sum_{t=2}^{n}(y_{t-1} - \overline{y})^2 \cdot \sum_{t=2}^{n}(x_{t-1} - \overline{x})^2}}.$$

Строим вспомогательную таблицу:

t	Y	X	Y_{t-1}	X_{t-1}	$(y_{t-1} - \overline{y})$	$(x_{t-1} - \overline{x})$	$(y_{t-1} - \overline{y})*(x_{t-1} - \overline{x})$	$(y_{t-1} - \overline{y})^2$	$(x_{t-1} - \overline{x})^2$
1	140	300	-	-	-	-	-	-	-
2	160	345	140	300	-94	-212,5	19975	8836	45156,25
3	190	405	160	345	-74	-167,5	12395	5476	28056,25
4	210	445	190	405	-44	-107,5	4730	1936	11556,25
5	220	480	210	445	-24	-67,5	1620	576	4556,25
6	240	535	220	480	-14	-32,5	455	196	1056,25
7	260	595	240	535	6	22,5	135	36	506,25
8	290	639	260	595	26	82,5	2145	676	6806,25
9	310	677	290	639	56	126,5	7084	3136	16002,25
10	320	704	310	677	76	164,5	12502	5776	27060,25
Суммы	2340	5125	2020	4421	-86	-191,5	61041	26644	140756,3
Средние	234	512,5	-	-	-	-	-	-	-

Получаем: $r_1 = \frac{61041}{\sqrt{26644*140756.3}} \approx 0.9967$ - коэффициент корреляции по первым разностям уровней ряда.

2. Оцените параметры уравнения парной линейной регрессии по первым разностям и поясните его смысл. В качестве зависимой переменной используйте валовую добавленную стоимость.

Строим таблицу:

№	Y	X	X^2	XY
1	140	300	90000	42000
2	160	345	119025	55200
3	190	405	164025	76950
4	210	445	198025	93450
5	220	480	230400	105600
6	240	535	286225	128400
7	260	595	354025	154700
8	290	639	408321	185310
9	310	677	458329	209870
10	320	704	495616	225280
Суммы	2340	5125	2803991	1276760
Средние	234	512,5	-	-

Подставляем в систему:

$$\begin{cases} na + b\sum x_i = \sum y_i; \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i; \end{cases} \to \begin{cases} 10a + 512.5b = 2340; \\ 512.5a + 2803991b = 1276760. \end{cases}$$

Решаем и получаем оценки уравнения регрессии:

Следовательно, оценки параметров регрессии будут равны: $a = 212{,}7, b = 0{,}4$. Таким образом, получили оценку уравнения регрессии: $y = 212{,}7 + 0{,}4x$.

Коэффициент b означает, что при увеличении ВДС на 1 млрд. рублей объемом инвестиций в основные производственные фонды (ОПФ) увеличится на 0,4 млрд. рублей.

3. Сделайте выводы о тесноте связи между временными рядами объема инвестиций в ОПФ и валовой добавленной стоимостью.

Коэффициент корреляции характеризует тесноту линейной связи между изучаемыми признаками. Так как коэффициент корреляции был вычислен ранее и составляет 0,9, что очень близко к 1, то можем сделать вывод, что связь между объемом инвестиций в ОПФ и ВДС прямая и очень тесная.