Задача №2 . По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x2 (%) (смотри таблицу своего варианта).

Требуется:

1. Построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат.

2. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.

3. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F --критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации R^2 (yx1x2)

5. С помощью частных F --критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x1 после фактора x2 и фактора x2после фактора x1

6. Составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Номер предприятия	y	x1	x2
1	6	3,5	10
2	6	3,6	12
3	7	3,9	15
4	7	4,1	17
5	7	4,2	18
6	8	4,5	19
7	8	5,3	19
8	9	5,3	20
9	9	5,6	20
10	10	6	21
11	10	6,3	21
12	11	6,4	22
13	11	7	23
14	12	7,5	25
15	12	7,9	28
16	13	8,2	30
17	13	8,4	31
18	14	8,6	31
19	14	9,5	35
20	15	10	36

Решение

№	y	X1	X2	X2*X2	X1*X1	y^2	y*X1	y*x2	X1*X2	y-ycp	x1-x1cp	x2-x2cp
1	6	3,5	10	100	12,25	36	21	60	35	-4,1	-2,79	-12,65
2	6	3,6	12	144	12,96	36	21,6	72	43,2	-4,1	-2,69	-10,65
3	7	3,9	15	225	15,21	49	27,3	105	58,5	-3,1	-2,39	-7,65
4	7	4,1	17	289	16,81	49	28,7	119	69,7	-3,1	-2,19	-5,65
5	7	4,2	18	324	17,64	49	29,4	126	75,6	7	4,2	18
6	8	4,5	19	361	20,25	64	36	152	85,5	-2,1	-1,79	-3,65
7	8	5,3	19	361	28,09	64	42,4	152	100,7	-2,1	-0,99	-3,65
8	9	5,3	20	400	28,09	81	47,7	180	106	-1,1	-0,99	-2,65
9	9	5,6	20	400	31,36	81	50,4	180	112	-1,1	-0,69	-2,65
10	10	6	21	441	36	100	60	210	126	-0,1	-0,29	-1,65
11	10	6,3	21	441	39,69	100	63	210	132,3	-0,1	0,01	-1,65
12	11	6,4	22	484	40,96	121	70,4	242	140,8	0,9	0,11	-0,65
13	11	7	23	529	49	121	77	253	161	0,9	0,71	0,35
14	12	7,5	25	625	56,25	144	90	300	187,5	1,9	1,21	2,35
15	12	7,9	28	784	62,41	144	94,8	336	221,2	1,9	1,61	5,35
16	13	8,2	30	900	67,24	169	106,6	390	246	2,9	1,91	7,35
17	13	8,4	31	961	70,56	169	109,2	403	260,4	2,9	2,11	8,35
18	14	8,6	31	961	73,96	196	120,4	434	266,6	3,9	2,31	8,35
19	14	9,5	35	1225	90,25	196	133	490	332,5	3,9	3,21	12,35
20	15	10	36	1296	100	225	150	540	360	4,9	3,71	13,35
сумма	202	125,8	453	11251	868,98	2194	1378,9	4954	3120,5	10,1	6,29	22,65
среднее	10,1	6,29	22,65	562,55	43,449	109,7	68,945	247,7	156,025	0,505		
σ	2,77	1,97	7,16									
σ^2	7,69	3,88	49,53									

Для нахождения параметров линейного уравнения множественной регрессии $y = a + b_1 x_1 + b_2 x_2$

необходимо решить следующую систему линейных уравнений относительно неизвестных параметров

$$a_1, b_1, b_2$$

$$\sum y = na + b_1 \cdot \sum x_1 + b_2 \cdot \sum x_2$$

$$\sum y \cdot x_1 = a \cdot \sum x_1 + b_1 \cdot \sum x_1^{\,2} + b_2 \cdot \sum x_1 \cdot x_2$$

$$\sum y \cdot x_2 = a \cdot \sum x_2 + b_1 \cdot \sum x_1 \cdot x_2 + b_2 \cdot \sum x_2^{\,2}$$

либо воспользоваться готовыми формулами

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2} \qquad b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r_{x_1x_2}^2} \qquad a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$$

Рассчитаем сначала парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x}_1 \overline{y}}{\sigma_y \cdot \sigma_{x_1}}$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \overline{x}_2 \overline{y}}{\sigma_y \cdot \sigma_{x_2}}$$

$$r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x}_1 \cdot \overline{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}}$$

ryx1= 0,990890269

ryx2= 0,953309527

rx1x2= 0,960261352

Решим данную систему уравнений и получим:

a = 1,32
b1 = 1,36
b2 = 0,009

Таким образом уравнение множественной регресии имеет вид:

$$\hat{y} = 1{,}32 + 1{,}36\, x_1 + 0{,}009\, x_2$$

Экономический смысл коэффициентов b1 и b2 в том, что это показатели силы связи, характеризующие изменение цены акции при изменении какого-либо факторного признака на единицу своего измерения при фиксированном влиянии другого фактора

Рассчитать частные коэффициенты эластичности.

Будем рассчитывать частные коэффициенты эластичности для среднего значения фактора и результата:

$$Эb_1 = b_1 \cdot \frac{\overline{x}_1}{\overline{y}}$$

$$Эb_2 = b_2 \cdot \frac{\overline{x}_2}{\overline{y}}$$

$$Э_{b_1} = 0{,}85$$

$$Э_{b_2} = 0{,}02$$

Определить стандартизованные коэффициенты регрессии
формулы определения:

$$\beta_j = b_j \cdot \frac{\sigma x_1}{\sigma y},$$

β1 = 0,97

β2 = 0,02

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что ввод в действие новых основных фондов оказывает большее влияние на выработку продукции, чем удельный вес рабочих высокой квалификации.Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности. Т.е. увеличение только основных фондов (от своего среднего значения) или только удельного веса рабочих высокой квалификации на 1% увеличивает в среднем выработку продукции на 0,85% или 0,02% соответственно. Таким образом, подтверждается большее влияние на результат фактора x1 , чем фактора x2 .

Коэффициенты парной корреляции мы уже нашли

ryx1= 0,9909

ryx2= 0,95

rx1x2= 0,96

Они указывают на весьма сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы x1 и x2 явно коллинеарны, т.к. r(x1x2) = 0,96>0,7). При такой сильной межфакторной зависимости рекомендуется один из факторов исключить из рассмотрения.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии

При двух факторах частные коэффициенты корреляции рассчитываются следующим образом

r(yx1x2)=(r(yx1)-r(yx2)*r(x1x2))/√(1-r(yx2)²)*(1-r(x1x2)²)= 0,8953

r(yx2x1)=(r(yx2)-r(yx1)*r(x1x2))/√(1-r(yx1)²)*(1-r(x1x2)²)= 0,0478

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи. Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи

Коэффициент множественной корреляции

R(yx1x2) = √∑β(i)*r(yx(i))= 0,9909

Коэффициент множественной корреляции показывает на весьма сильную связь всего набора факторов с результатом

Нескорректированный коэффициент множественной детерминации

R(yx1x2)²= 0,98190

оценивает долю вариации результата за счет представленных в уравнении факторов в общей вариации результата. Здесь эта доля составляет 98% и указывает на весьма высокую степень обусловленности вариации результата вариацией факторов, иными словами – на весьма тесную связь факторов с результатом

Скорректированный коэффициент множественной детерминации

Rск²=1-(1-R²)*(n-1)/(n-m-1)= 0,9798

определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсий. Он дает такую оценку тесноты связи, которая не зависит от числа факторов и поэтому может сравниваться по разным моделям с разным числом факторов. Оба коэффициента указывают на весьма высокую (более 94 %) детерминированность результата y в модели факторами x1 и x2

Оценку надежности уравнения регрессии в целом и показателя тесноты R(yx1x2) связи дает F -критерий Фишера:

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

В нашем случае получаем

F= 228,51

Получили, что F (фак) > F (таб) = 3,49 (при n=20), т.е. вероятность случайно получить такое значение F -критерия не превышает допустимый уровень значимости 5% Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов, т.е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи R ²(yx1x2)

С помощью частных F -критериев Фишера оценим целесообразность включения в уравнение множественной регрессии фактора x1 после x2 и фактора x2 после x1 при помощи формул:

$$F_{\text{част, } x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1-R^2_{yx_1}} \cdot \frac{n-m-1}{m}$$

$$F_{\text{част, } x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1-R^2_{yx_2}} \cdot \frac{n-m-1}{m}$$

Найдем $R^2_{yx_2}$ и $R^2_{yx_1}$

R(yx1)²= 0,98186

$R(yx2)^2 = 0{,}91$

Тогда

F (час x1)= 34,26

F (час x2)= 0,004

Получили, что F (част. x2) < F (таб) = 3,49 Следовательно, включение в модель фактора x2 после того, как в модель включен фактор x1 сстатистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x2 оказывается незначительным, несущественным; фактор x2 включать в уравнение после фактора x1 не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения x1 после x2 то результат расчета частного F -критерия для x1 будет иным F (част. x1) > F (таб) = 3,49 т.е. вероятность его случайного формирования меньше принятого стандарта $\alpha = 0{,}05$ (5 %) Следовательно, значение частного F -критерия для дополнительно включенного фактора x1 не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора x1 является существенным. Фактор x1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x2 .

Общий вывод состоит в том, что множественная модель с факторами x1 и x2 с $R(yx1x2)^2 = 0{,}98190$ содержит неинформативный фактор x2

Если исключить фактор x2 , то можно ограничиться уравнением парной регрессии

$$a = \frac{\sum y \cdot \sum x^2 - \sum xy \cdot \sum x}{n \cdot \sum x^2 - \sum x \cdot \sum x} \qquad b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \sum x \cdot \sum x}$$

a = 1,33 b = 1,39

y = 1,33+1,39*x1

r(yx) = 0,99089

$r(yx)^2 =$ 0,98186