**Контрольная работа по эконометрике**

**Вариант 3**

Задача 1.По территории региона приводятся данные за 199Х год

|  |  |  |
| --- | --- | --- |
| Номер региона | Среднедушевой прожиточный минимум в день на одного трудоспособного, руб.,х | Среднедневная заработная плата, руб., у |
| 1 | 77 | 123 |
| 2 | 85 | 152 |
| 3 | 79 | 140 |
| 4 | 93 | 142 |
| 5 | 89 | 157 |
| 6 | 81 | 181 |
| 7 | 79 | 133 |
| 8 | 97 | 163 |
| 9 | 73 | 134 |
| 10 | 95 | 155 |
| 11 | 84 | 132 |
| 12 | 108 | 165 |

Требуется:

1. Построить линейное уравнение парной регрессии y по x.

2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F-критерия Фишера и t-критерия Стьюдента

4. Выполнить прогноз заработной платы при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.

5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

6. На одном графике отложить исходные данные и теоретическую прямую.

**Решение**

1. Построим линейное уравнение парной регрессии y по x. Ищем уравнение в виде: *у* = *а+ bх.*

Составим расчетную таблицу для дальнейшего исследования:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Месяц |  *х* |  *у* |  *х*2 |  *ху* |  *у*2 | $$\hat{у\_{х}}$$ | $$у-\hat{у\_{х}}$$ | $$\left|\frac{у-\hat{у\_{х}}}{у}\right|$$ | $$\left(y-\hat{у\_{х}}\right)^{2}$$ |  |
| 1 | 77 | 123 | 5929 | 9471 | 15129 | 139,28 | 264,927 | -16,28 | 0,1323 | 629,01 |
| 2 | 85 | 152 | 7225 | 12920 | 23104 | 146,56 | 29,54 | 5,44 | 0,0358 | 15,37 |
| 3 | 79 | 140 | 6241 | 11060 | 19600 | 141,10 | 1,20705 | -1,10 | 0,0078 | 65,29 |
| 4 | 93 | 142 | 8649 | 13206 | 20164 | 153,85 | 140,5 | -11,85 | 0,0835 | 36,97 |
| 5 | 89 | 157 | 7921 | 13973 | 24649 | 150,21 | 46,1163 | 6,79 | 0,0433 | 79,57 |
| 6 | 81 | 181 | 6561 | 14661 | 32761 | 142,92 | 1450,03 | 38,08 | 0,2104 | 1083,73 |
| 7 | 79 | 133 | 6241 | 10507 | 17689 | 141,10 | 65,5882 | -8,10 | 0,0609 | 227,41 |
| 8 | 97 | 163 | 9409 | 15811 | 26569 | 157,50 | 30,2779 | 5,50 | 0,0338 | 222,61 |
| 9 | 73 | 134 | 5329 | 9782 | 17956 | 135,63 | 2,66469 | -1,63 | 0,0122 | 198,25 |
| 10 | 95 | 155 | 9025 | 14725 | 24025 | 155,68 | 0,45613 | -0,68 | 0,0044 | 47,89 |
| 11 | 84 | 132 | 7056 | 11088 | 17424 | 145,65 | 186,428 | -13,65 | 0,1034 | 258,57 |
| 12 | 108 | 165 | 11664 | 17820 | 27225 | 167,52 | 6,34514 | -2,52 | 0,0153 | 286,29 |
| Сумма | **1040** | **1777** | **91250** | **155024** | **266295** | **1777** | **2224,08** | **-9E-14** | **0,7429** | **3150,92** |
| Ср.знач. | ***86,6667*** | ***148,0833*** | ***7604*** |  |  |  |  |  |  |  |
| σ | ***9,55*** | ***16,92*** |  |  |  |  |  |  |  |  |
| σ2 | ***91,15*** | ***286,45*** |  |  |  |  |  |  |  |  |

Вычислим параметры линейного уравнения регрессии по формулам:

$$b=\frac{n\sum\_{i=1}^{n}xᵢyᵢ-\sum\_{i=1}^{n}xᵢ\sum\_{i=1}^{n}yᵢ}{n\sum\_{i=1}^{n}xᵢ²-(\sum\_{i=1}^{n}xᵢ)²}=\frac{12∙155024-1040∙1777}{12∙91250-1040²}≈0,9110;$$

$$a=\frac{\sum\_{i=1}^{n}yᵢ-b\sum\_{i=1}^{n}xᵢ}{n}=\frac{1777-9110∙1040}{12}≈69,13.$$

Получено выборочное уравнение линейной регрессии

 $ \hat{у\_{х}}=69,13+0,911x$.

2. Рассчитаем линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

 Вычислим линейный коэффициент корреляции по формуле:

$$r=\frac{n\sum\_{i=1}^{n}xᵢyᵢ-\sum\_{i=1}^{n}xᵢ\sum\_{i=1}^{n}yᵢ}{\sqrt{(n\sum\_{i=1}^{n}xᵢ²-(\sum\_{i=1}^{n}xᵢ)²)(n\sum\_{i=1}^{n}yᵢ²-(\sum\_{i=1}^{n}yᵢ)²)}}=$$

$$=\frac{12∙155024-1040∙1777}{\sqrt{(12 ∙91250-1040²)(12∙266295-1777²)}}≈0,5424.$$

 Значение коэффициента корреляции не очень близко к 1, что свидетельствует об умеренной линейной положительной связи (с ростом *х* значения *у* возрастают).

Коэффициент детерминации ***r2***= 0,542420,2941.

Это означает, что 29% вариации заработной платы () объясняется вариацией фактора  – среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации:

Коэффициент аппроксимации определим по формуле:



Средняя ошибка аппроксимации:



Допустимый предел значений  – не более 10 %.

Качество построенной модели оценивается как хорошее, так как  не превышает 8-10%.

3. Оценим статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F-критерия Фишера и t-критерия Стьюдента.

Проведем оценку качества уравнения регрессии в целом с помощью F-критерия Фишера. Найдем фактическое значение F-критерия:

$$F=\frac{r^{2}}{1-r^{2}}∙\left(n-2\right)=\frac{0,2941}{1-0,2941}∙10=4,17.$$

Табличное значение ($k\_{1}=1, k\_{2}=n-2=10, α=0,05); F\_{табл.}=4,96.$

Так как $F\_{факт.}<F\_{табл.}$, поэтому уравнение статистически не значимо.

Оценку статистической значимости параметров регрессии проведем с помощью -статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Табличное значение -критерия для числа степеней свободы  и  составит .

Определим случайные ошибки , , :



;

;

.

Тогда

;

;

.

Фактические значения -статистики превосходят табличное значение:

; ; ,

поэтому параметры ,  и  случайно отличаются от нуля и статистически незначимы.

Рассчитаем доверительные интервалы для параметров регрессии  и . Для этого определим предельную ошибку для каждого показателя:

;

.

Доверительные интервалы













Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью  параметры  и , находясь в указанных границах, принимают нулевые значения, т.е. являются статистически незначимыми и несущественно отличны от нуля.

Вывод: статистически незначимы параметры уравнения регрессии и коэффициент корреляции.

4. Выполним прогноз заработной платы при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.

Вычислим прогнозное значение результативного фактора $\hat{у}\_{0}$ при значении признака-фактора $х\_{0}=1,07∙86,6667=92,73$(руб)

$\hat{у}\_{0}=69,126+0,911∙92,73=153,6(руб)$.

5. Оценим точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Рассчитаем доверительный интервал прогноза. Ошибка прогноза

$$m\_{\hat{y\_{p}}}=S\_{ост.}∙\sqrt{1+\frac{1}{n}+\frac{\left(x\_{p}-\overbar{x}\right)^{2}}{n∙σ\_{x}^{2}}}$$

$$ S\_{ост.}=17,75;$$

$$σ\_{x}^{2}=\overbar{x^{2}}-\overbar{x}^{2}=7604-86,67^{2}=91,15;$$

$$m\_{\hat{y\_{0}}}=17,75∙\sqrt{1+\frac{1}{12}+\frac{\left(92,73-86,67\right)^{2}}{12∙91,15}}=18,76 (руб)$$

Доверительный интервал рассчитывается так:

$$\left(\hat{у}\_{0}-m\_{\hat{y\_{0}}}∙t\_{табл.}\leq \hat{у}\_{0}^{\*}\leq \hat{у}\_{0}+m\_{\hat{y\_{0}}}∙t\_{табл.}\right)$$

Табличное значение t-критерия Стьюдента при α=0,05 и числе степеней свободы k=n-2=10 есть $t\_{табл.}=2,23$

Поэтому доверительный интервал будет иметь вид:

$$\left(153,6-18,76∙2,23\leq \hat{у}\_{0}^{\*}\leq 153,6+18,76∙2,23\right)$$

$$111,76\leq \hat{у}\_{0}^{\*}\leq 195,43 (руб)$$

Выполненный прогноз среднемесячной заработной платы является надежным

() и находится в пределах от 111,76 руб. до 195,43 руб.

1. В заключение решения задачи построим на одном графике исходные данные и теоретическую прямую:

Задача 2.

 По 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 (%)

|  |  |  |  |
| --- | --- | --- | --- |
| Номер предприятия | у | Х1 | Х2 |
| 1 | 7 | 3,7 | 9 |
| 2 | 7 | 3,7 | 11 |
| 3 | 7 | 3,9 | 11 |
| 4 | 7 | 4,1 | 15 |
| 5 | 8 | 4,2 | 17 |
| 6 | 8 | 4,9 | 19 |
| 7 | 8 | 5,3 | 19 |
| 8 | 9 | 5,1 | 20 |
| 9 | 10 | 5,6 | 20 |
| 10 | 10 | 6,1 | 21 |
| 11 | 11 | 6,3 | 22 |
| 12 | 11 | 6,4 | 22 |
| 13 | 11 | 7,2 | 23 |
| 14 | 12 | 7,5 | 25 |
| 15 | 12 | 7,9 | 27 |
| 16 | 13 | 8,1 | 30 |
| 17 | 13 | 8,4 | 31 |
| 18 | 13 | 8,6 | 32 |
| 19 | 14 | 9,5 | 35 |
| 20 | 15 | 9,5 | 36 |

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 .

5. С помощью частных F-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора х1 после х2 и фактора х2 после х1.

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

**Решение**

Для удобства проведения расчетов поместим результаты промежуточных расчетов в таблицу:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № |  |  |  |  |  |  |  |  |  |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1 | 7 | 3,7 | 9 | 25,9 | 63 | 33,3 | 13,69 | 81 | 49 |
| 2 | 7 | 3,7 | 11 | 25,9 | 77 | 40,7 | 13,69 | 121 | 49 |
| 3 | 7 | 3,9 | 11 | 27,3 | 77 | 42,9 | 15,21 | 121 | 49 |
| 4 | 7 | 4,1 | 15 | 28,7 | 105 | 61,5 | 16,81 | 225 | 49 |
| 5 | 8 | 4,2 | 17 | 33,6 | 136 | 71,4 | 17,64 | 289 | 64 |
| 6 | 8 | 4,9 | 19 | 39,2 | 152 | 93,1 | 24,01 | 361 | 64 |
| 7 | 8 | 5,3 | 19 | 42,4 | 152 | 100,7 | 28,09 | 361 | 64 |
| 8 | 9 | 5,1 | 20 | 45,9 | 180 | 102 | 26,01 | 400 | 81 |
| 9 | 10 | 5,6 | 20 | 56 | 200 | 112 | 31,36 | 400 | 100 |
| 10 | 10 | 6,1 | 21 | 61 | 210 | 128,1 | 37,21 | 441 | 100 |
| 11 | 11 | 6,3 | 22 | 69,3 | 242 | 138,6 | 39,69 | 484 | 121 |
| 12 | 11 | 6,4 | 22 | 70,4 | 242 | 140,8 | 40,96 | 484 | 121 |
| 13 | 11 | 7,2 | 23 | 79,2 | 253 | 165,6 | 51,84 | 529 | 121 |
| **1** | 12 | 7,5 | 25 | 90 | 300 | 187,5 | 56,25 | 625 | 144 |
| 14 | 12 | 7,9 | 27 | 94,8 | 324 | 213,3 | 62,41 | 729 | 144 |
| 15 | 13 | 8,1 | 30 | 105,3 | 390 | 243 | 65,61 | 900 | 169 |
| 16 | 13 | 8,4 | 31 | 109,2 | 403 | 260,4 | 70,56 | 961 | 169 |
| 17 | 13 | 8,6 | 32 | 111,8 | 416 | 275,2 | 73,96 | 1024 | 169 |
| 18 | 14 | 9,5 | 35 | 133 | 490 | 332,5 | 90,25 | 1225 | 196 |
| 19 | 15 | 9,5 | 36 | 142,5 | 540 | 342 | 90,25 | 1296 | 225 |
| 20 | 7 | 3,7 | 9 | 25,9 | 63 | 33,3 | 13,69 | 81 | 49 |
| **Сумма** | **213** | **129,7** | **454** | **1417,3** | **5015** | **3117,9** | **879,19** | **11138** | **2297** |
| **Ср. знач.** | **10,143** | **6,176** | **21,619** | **67,490** | **238,810** | **148,471** | **41,866** | **530,381** | **109,381** |
| № |  |  |  |  |  |  |  |  |  |

Найдем средние квадратические отклонения признаков:

;

;

.

1. Вычисление параметров линейного уравнения множественной регрессии.

Для нахождения параметров линейного уравнения множественной регрессии



необходимо решить следующую систему линейных уравнений относительно неизвестных параметров , , :



либо воспользоваться готовыми формулами:

; ;

.

Рассчитаем сначала парные коэффициенты корреляции:

;

;

.

Находим

;

;

.

Таким образом, получили следующее уравнение множественной регрессии:

.

Коэффициенты  и  стандартизованного уравнения регрессии  находятся по формулам:

 ;

 .

Т.е. уравнение будет выглядеть следующим образом:

.

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:

 .

Вычисляем:

 ; .

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 0,74% или 0,04% соответственно. Таким образом, подтверждается большее влияние на результат  фактора , чем фактора .

1. Коэффициенты парной корреляции мы уже нашли:

; ; .

Они указывают на весьма сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы  и  явно коллинеарны, т.к. ). 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:

;

.

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 определить через матрицу парных коэффициентов корреляции:

,

где



– определитель матрицы парных коэффициентов корреляции;



– определитель матрицы межфакторной корреляции.

;

****.

Коэффициент множественной корреляции

.

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

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

Скорректированный коэффициент множественной детерминации



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 ) детерминированность результата  в модели факторами  и .

1. Оценку надежности уравнения регрессии в целом и показателя тесноты связи  дает -критерий Фишера:

.

В нашем случае фактическое значение -критерия Фишера:

 .

Получили, что  (при ), т.е. вероятность случайно получить такое значение -критерия не превышает допустимый уровень значимости . Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов, т.е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи .

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

 ;

 .

Найдем  и .

 ;

 .

Имеем

 ;

 .

Получили, что . Следовательно, включение в модель фактора  после того, как в модель включен фактор  статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака  оказывается незначительным, несущественным; фактор  включать в уравнение после фактора  не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения  после , то результат расчета частного -критерия для  будет иным. , т.е. вероятность его случайного формирования меньше принятого стандарта . Следовательно, значение частного -критерия для дополнительно включенного фактора  не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора  является существенным. Фактор  должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора .

1. Общий вывод состоит в том, что множественная модель с факторами  и  с  содержит неинформативный фактор . Если исключить фактор , то можно ограничиться уравнением парной регрессии:

 , 