**ВАРИАНТ 5**

Изучается зависимость средней ожидаемой продолжительности жизни от нескольких факторов по данным за 1995 г., представленным в табл. 5.

**Таблица 5**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Страна** | ***Y*** | ***X*1** | ***X*2** | ***X*3** | ***X*4** |
| Мозамбик | 47 | 3,0 | 2,6 | 2,4 | 113 |
| Бурунди | 49 | 2,3 | 2,6 | 2,7 | 98 |
| …………………………………………………………………………………….. | | | | | |
| Швейцария | 78 | 95,9 | 1,0 | 0,8 | 6 |

*Принятые в таблице обозначения:*

* *Y* — средняя ожидаемая продолжительность жизни при рождении, лет;
* *X*1— ВВП в паритетах покупательной способности;
* *X*2— цепныетемпы прироста населения, %;
* *X*3— цепныетемпы прироста рабочей силы, %;
* *Х*4 — коэффициент младенческой смертности, %*.*

**Требуется:**

1. Составить матрицу парных коэффициентов корреляции между всеми исследуемыми переменными и выявить коллинеарные факторы.
2. Построить уравнение регрессии, не содержащее коллинеарных факторов. Проверить статистическую значимость уравнения и его коэффициентов.
3. Построить уравнение регрессии, содержащее только статистически значимые и информативные факторы. Проверить статистическую значимость уравнения и его коэффициентов.

Пункты 4 — 6 относятся к уравнению регрессии, построенному при выполнении пункта 3.

1. Оценить качество и точность уравнения регрессии.
2. Дать экономическую интерпретацию коэффициентов уравнения регрессии и сравнительную оценку силы влияния факторов на результативную переменную *Y*.
3. Рассчитать прогнозное значение результативной переменной *Y*, если прогнозные значения факторов составят 75 % от своих максимальных значений. Построить доверительный интервал прогноза фактического значения *Y* c надежностью 80 %.

**Решение.** Для решения задачи используется табличный процессор EXCEL.

1.С помощью надстройки «*Анализ данных*… *Корреляция*» строим матрицу парных коэффициентов корреляции между всеми исследуемыми переменными (меню «*Сервис*» → «*Анализ данных*…» → «*Корреляция*»). На **рис. 1**изображена панель корреляционного анализа с заполненными полями[[1]](#footnote-1). Результаты корреляционного анализа приведены в **прил. 2** и перенесены в **табл. 1**.

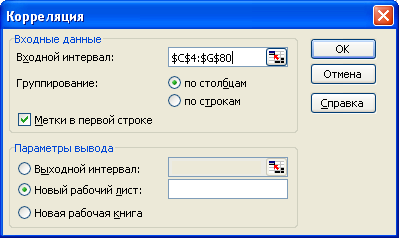


рис. 1. Панель корреляционного анализа

**Таблица 1**

**Матрица парных коэффициентов корреляции**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Y* | *X1* | *X2* | *X3* | *X4* |
| Y | 1 |  |  |  |  |
| X1 | 0,780235 | 1 |  |  |  |
| X2 | -0,72516 | -0,62251 | 1 |  |  |
| X3 | -0,53397 | -0,65771 | **0,874008** | 1 |  |
| X4 | -0,96876 | -0,74333 | 0,736073 | 0,55373 | 1 |

Анализ ***межфакторных*** коэффициентов корреляции показывает, что значение 0,8 превышает по абсолютной величине коэффициент корреляции между парой факторов *Х*2–*Х*3 (выделен жирным шрифтом). Факторы *Х*2–*Х*3 таким образом, признаются коллинеарными.

2. Как было показано в пункте 1, факторы *Х*2–*Х*3 являются коллинеарными, а это означает, что они фактически дублируют друг друга, и их одновременное включение в модель приведет к неправильной интерпретации соответствующих коэффициентов регрессии. Видно, что фактор *Х*2 имеет больший по модулю коэффициент корреляции с результатом *Y*, чем фактор *Х*3: *ry*,*x*2=0,72516; *ry*,*x*3=0,53397; |ry,x2|>|ry,x3| (см. **табл. 1**). Это свидетельствует о более сильном влиянии фактора *Х*2 на изменение *Y*. Фактор *Х*3, таким образом, исключается из рассмотрения.

Для построения уравнения регрессии значения используемых переменных (*Y*, *X*1, *X*2, *X*4) скопируем на чистый рабочий лист (**прил. 3)**. Уравнение регрессии строим с помощью надстройки «*Анализ данных… Регрессия*» (меню «*Сервис»* → «*Анализ данных…*» → «*Регрессия*»). Панель регрессионного анализа с заполненными полями изображена на **рис. 2**.

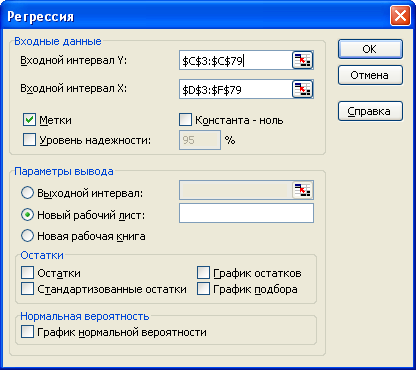
Результаты регрессионного анализа приведены в **прил. 4** и перенесены в **табл. 2**. Уравнение регрессии имеет вид (см. «*Коэффициенты»* в **табл. 2**):

ŷ = 75.44 + 0.0447 · x1 - 0.0453 · x2 - 0.24 · x4

Уравнение регрессии признается статистически значимым, так как вероятность его случайного формирования в том виде, в котором оно получено, составляет 1.04571·10-45 (см. *«Значимость F»* в**табл. 2**), что существенно ниже принятого уровня значимости α=0,05.

Вероятность случайного формирования коэффициентов при факторе *Х*1 ниже принятого уровня значимости α=0,05 (см. «*P-Значение»* в **табл. 2**), что свидетельствует о статистической значимости коэффициентов и существенном влиянии этих факторов на изменение годовой прибыли *Y*.

Вероятность случайного формирования коэффициентов при факторах *Х*2 и *Х*4 превышает принятый уровень значимости α=0,05 (см. «*P-Значение»* в**табл. 2**), и эти коэффициенты не признаются статистически значимыми.



**рис. 2. Панель регрессионного анализа модели *Y*(*X*1,*X*2,*X*4)**

**Таблица 2**

**Результаты регрессионного анализа модели *Y*(*X*1, *X*2, *X*4)**

|  |  |
| --- | --- |
| Регрессионная статистика | |
| Множественный R | 0,97292594 |
| R-квадрат | 0,946584884 |
| Нормированный R-квадрат | 0,944359254 |
| Стандартная ошибка | 2,267611945 |
| Наблюдения | 76 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ | | | | | |  |  |  |
|  | df | SS | MS | F | Значимость F |  |  |  |
| Регрессия | 3 | 6560,929292 | 2186,98 | 425,31101 | 1,04571E-45 |  |  |  |
| Остаток | 72 | 370,2286032 | 5,14206 |  |  |  |  |  |
| Итого | 75 | 6931,157895 |  |  |  |  |  |  |
|  |  | Уравнение регрессии | |  |  |  |  |  |
|  | Коэффициенты | Стандартная ошибка | t-статистика | P-Значение | Нижние 95% | Верхние 95% | Нижние 95,0% | Верхние 95,0% |
| Y-пересечение | 75,43927547 | 0,998411562 | 75,5593 | 2,545E-70 | 73,44897843 | 77,4295725 | 73,44897843 | 77,42957252 |
| X1 | 0,044670594 | 0,01380341 | 3,2362 | 0,0018316 | 0,017154 | 0,07218719 | 0,017154 | 0,072187188 |
| X2 | -0,045296701 | 0,421363275 | -0,1075 | 0,914691 | -0,885269026 | 0,79467562 | -0,885269026 | 0,794675624 |
| X4 | -0,239566687 | 0,013204423 | -18,1429 | 1,438E-28 | -0,265889223 | -0,2132442 | -0,265889223 | -0,213244151 |

3.По результатам проверки статистической значимости коэффициентов уравнения регрессии, проведенной в предыдущем пункте, строим новую регрессионную модель, содержащую только информативные факторы, к которым относятся:

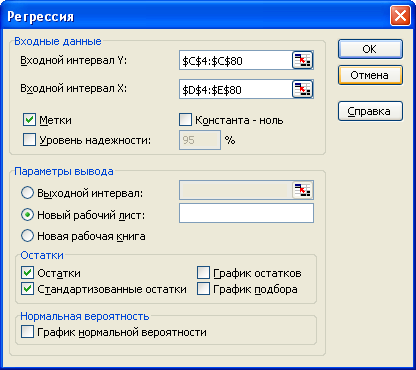
* факторы, коэффициенты при которых статистически значимы;
* факторы, у коэффициентов которых *t*‑статистика превышает по модулю единицу (другими словами, абсолютная величина коэффициента больше его стандартной ошибки).

К первой группе относится фактор *Х*1 ко второй — фактор *X*4. Фактор *X*2 исключается из рассмотрения как неинформативный, и окончательно регрессионная модель будет содержать факторы *X*1, *X*4.

Для построения уравнения регрессии скопируем на чистый рабочий лист значения используемых переменных (**прил. 5)** и проведем регрессионный анализ (**рис. 3**). Его результаты приведены в **прил. 6** и перенесены в **табл. 3**. Уравнение регрессии имеет вид:

ŷ = 75.38278 + 0.044918 · x1 - 0.24031 · x4

(см. «*Коэффициенты»* в**табл.3**).



**рис. 3. Панель регрессионного анализа модели *Y*(*X*1, *X*4)**

**Таблица 3**

**Результаты регрессионного анализа модели *Y*(*X*1, *X*4)**

|  |  |
| --- | --- |
| Регрессионная статистика | |
| Множественный R | 0,972922 |
| R-квадрат | 0,946576 |
| Нормированный R-квадрат | 0,945113 |
| Стандартная ошибка | 2,252208 |
| Наблюдения | 76 |
|  |  |
| Дисперсионный анализ | | | | | |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 2 | 6560,87 | 3280,435 | 646,7175 | 3,65E-47 |
| Остаток | 73 | 370,288 | 5,072439 |  |  |
| Итого | 75 | 6931,158 |  |  |  |
| Уравнение регрессии | | | | | | |  |
|  | Коэффициенты | Стандартная ошибка | t-статистика | P-Значение |  |
| Y-пересечение | 75,38278 | 0,843142 | 89,40701 | 2,44E-76 |  |
| X1 | 0,044918 | 0,013518 | 3,322694 | 0,001395 |  |
| X4 | -0,24031 | 0,011185 | -21,4848 | 2,74E-33 |  |

Уравнение регрессии статистически значимо: вероятность его случайного формирования ниже допустимого уровня значимости α=0,05 (см. «*Значимость F»* в**табл.3**).

Статистически значимым признается и коэффициент при факторе *Х*1 вероятность его случайного формирования ниже допустимого уровня значимости α=0,05 (см. «*P-Значение»* в **табл. 3**). Это свидетельствует о существенном влиянии ВВП в паритетах покупательной способности *X*1 на изменение годовой прибыли *Y*.

Коэффициент при факторе *Х*4 (годовой коэффициент младенческой смертности) не является статистически значимым. Однако этот фактор все же можно считать информативным, так как *t*‑статистика его коэффициента превышает по модулю единицу, хотя к дальнейшим выводам относительно фактора *Х*4 следует относиться с некоторой долей осторожности.

4.Оценим качество и точность последнего уравнения регрессии, используя некоторые статистические характеристики, полученные в ходе регрессионного анализа (см*.* «*Регрессионную статистику*» в **табл. 3**):

* множественный коэффициент детерминации

n

∑ (ŷi - y)2

R2= \_i=1\_\_\_\_\_\_\_\_\_\_\_\_ =0.946576

n

∑(ŷi - y)2

i=1

*R2=*показывает, что регрессионная модель объясняет 94,7 % вариации средней ожидаемой продолжительности жизни при рождении *Y*, причем эта вариация обусловлена изменением включенных в модель регрессии факторов *X*1, *X*4;

* стандартная ошибка регрессии



показывает, что предсказанные уравнением регрессии значения средней ожидаемой продолжительности жизни при рождении *Y* отличаются от фактических значений в среднем на 2,252208 лет.

Средняя относительная ошибка аппроксимации определяется по приближенной формуле:

Sрег

Eотн≈0,8 · — · 100%=0.8 · 2.252208/66.9 · 100%≈2.7

− y

где тыс. руб. — среднее значение продолжительности жизни (определено с помощью встроенной функции «*СРЗНАЧ*»; **прил. 1**).



*Е*отн показывает, что предсказанные уравнением регрессии значения годовой прибыли *Y* отличаются от фактических значений в среднем на 2,7 %. Модель имеет высокую точность (при — точность модели высокая, при — хорошая, при — удовлетворительная, при — неудовлетворительная).



5.Для экономической интерпретации коэффициентов уравнения регрессии сведем в таблицу средние значения и стандартные отклонения переменных в исходных данных (**табл. 4***)*. Средние значения были определены с помощью встроенной функции «*СРЗНАЧ*», стандартные отклонения — с помощью встроенной функции «*СТАНДОТКЛОН*» (см. **прил. 1**).

**Таблица 4**

**Средние значения и стандартные отклонения используемых переменных**

|  |  |  |  |
| --- | --- | --- | --- |
| **Переменная** | ***Y*** | ***X*1** | ***X4*** |
| **Среднее** | **66,9** | **29,75** | **40,9** |
| **Стандартное отклонение** | **9,6** | **28,76** | **34,8** |

***1) Фактор X1 (*ВВП в паритетах покупательной способности*)***

Значение коэффициента *b*1=0,044918 показывает, что рост ВВП в паритетах покупательной способности на 1 %. приводит к повышению средней ожидаемой продолжительности жизни при рождении на 0,044918 лет.

Средний коэффициент эластичности фактора*X*1 имеет значение

x1 29.75

Е1= b1 · ― = 0.044918 · \_\_\_\_ ≈ 0.01997

y 66.9

Он показывает, что при увеличении ВВП в паритетах покупательской способности на 1 % годовая прибыль увеличивается в среднем на 0,01997 %.

***2) Фактор X4 (*коэффициент младенческой смертности*)***

Значение коэффициента *b*4=(-0,24031) показывает, что рост коэффициента младенческой смертности на 1 %. приводит к уменьшению средней ожидаемой продолжительности жизни при рождении в среднем на -0,24031 лет.

Средний коэффициент эластичности фактора*X*4 имеет значение

x4 40.9

Е4 = b4 · ― = - 0.24031 · \_\_\_\_ ≈ 0.1469

y 66.9

Он показывает, что при увеличении коэффициента младенческой смертности на 1 % средняя ожидаемая продолжительность жизни увеличивается в среднем на 0,1469 %.

Средний коэффициент эластичности для фиктивных переменных лишен смысла, поэтому не рассчитывается.

Сравним между собой силу влияния факторов, включенных в регрессионную модель, на годовую прибыль, для чего определим их бета–коэффициенты:

Sx1 28.76

B1 = b1 · ― = 0.044918 · \_\_\_\_ ≈ 0.1346;

Sy 9.6

Sx4 3 4.8

B4 = b4 · ― - 0.24031 · \_\_\_\_ ≈ - 0.8711

Sy 9.6

Сравнивая по абсолютной величине значения бета–коэффициентов, можно сделать вывод о том, что на изменение средней ожидаемой продолжительности жизни при рождении *Y* сильнее всего влияет ВВП в паритетах покупательской способности *Х*1, далее по степени влияния следует коэффициент младенческой смертности *Х*4.

Определим дельта–коэффициенты факторов:

ry,x1 0.780235

Δ1 = B1 · \_\_\_ = 0.1346 · \_\_\_\_\_\_\_ ≈ 0.11094;

R2 0.946585

ry,x4 - 0.96876

Δ4 = B4 · \_\_\_ = - 0.8711 · \_\_\_\_\_\_\_ ≈ 0.8915;

R2 0.946585

где *ry*,*x1*=0,780235; *ry*,*x*4=(–0,96876); — коэффициенты корреляции между парами переменных *Y*–*X*1и *Y*–*X*4 соответственно (см. **табл. 1**); *R*2=0,946585 — множественный коэффициент детерминации (см. **табл. 3**).

Сумма дельта–коэффициентов факторов, включенных в модель, должна быть равна единице. Небольшое неравенство может быть вызвано погрешностями промежуточных округлений.

Таким образом, в суммарном влиянии на среднюю ожидаемую продолжительность жизни при рождении *Y* всех факторов, включенных в модель, доля влияния ВВП в паритетах покупательной способности *X*1 составляет 11,094 %, коэффициента младенческой смертности *Х*4 — 89,15 %.

6.Рассчитаем прогнозное значение годовой прибыли, если прогнозные значения факторов составят 75 % от своих максимальных значений в исходных данных. Максимальные значения факторов были определены с помощью встроенной функции «*МАКС*» (см. **прил. 1**). Прогнозные значения рассчитываются только для количественных факторов *X*1 и *X*4:

* фактор *Х*1: х01=0,75\*х1max=0.75\*100=75;
* фактор *Х*4: x04=0.75\*x4max=0.75\*124=93.

Среднее прогнозируемое значение (точечный прогноз) годовой прибыли государственной компании (*x*06=0) составляет:



Для частной компании (*x*06=1) этот показатель равен



Стандартная ошибка прогноза фактического значения годовой прибыли *y*0рассчитывается по формуле



Так как фиктивная переменная *Х*6 может принимать два значения — 0 или 1, то *Sy*0 определяется для обоих случаев:

* для государственных компаний (*x*06=0):



* для частных компаний (*x*06=1):



Построим интервальный прогноз фактического значения годовой прибыли *y*0 с доверительной вероятностью γ=0,8. Доверительный интервал имеет вид:

,



где *t*таб=1,321 — табличное значение *t*-критерия Стьюдента при уровне значимости и числе степеней свободы (*p*=4 — число факторов в модели) (см. ***Справочные таблицы***).



Для государственных компаний:

тыс. руб.



Таким образом, с вероятностью 80 % годовая прибыль государственных компаний при заданных значениях факторов будет находиться в интервале от 272,4 до 945,4 тыс. руб.

Для частных компаний:

тыс. руб.



С вероятностью 80 % годовая прибыль частных компаний будет находиться в интервале от 499,1 до 1173,7 тыс. руб.

1. Для копирования снимка окна в буфер обмена данных WINDOWS используется комбинация клавиш Alt+Print Screen (на некоторых клавиатурах — Alt+PrtSc). [↑](#footnote-ref-1)