Контрольная работа

по дисциплине "Эконометрика"

студента гр. ВФ-108

Звягиной Марии Михайловны

# Раздел I. Практическая часть

**Содержание заданий.**

# Задание 1

1. По исходным данным выполнить корреляционный анализ:

Таблица 9

Основные показатели работы грузовых автомобилей крупных и средних организаций автомобильного транспорта в 2006 году

|  |  |  |
| --- | --- | --- |
|  | Перевезено грузов, тыс. тонн | Расходы, млн, руб |
| Владимирская | 594,6 | 258,3 |
| Брянская | 3178,9 | 656,5 |
| Белгородская | 523,8 | 824,4 |
| Воронежская | 2572,3 | 220,1 |
| Ивановская | 308,5 | 73,8 |
| Костромская | 580,5 | 82,7 |
| Рязанская | 203,7 | 65,4 |
| Смоленская | 389,3 | 86,6 |
| Тульская | 225,8 | 36,5 |
| Ярославская | 693,4 | 279,9 |

Основной задачей *корреляционного анализа* является - выявление связи между случайными переменными и оценка её тесноты. Показателем тесноты линейной связи является *коэффициент корреляции* *r*.

* 1. Построить корреляционное поле и предложить гипотезу о связи исследуемых факторов

Для трактовки линейной связи между переменной X ("Перевезено грузов") и Y ("Расходы") при помощи встроенных возможностей Microsoft Excel построим поле корреляции заданной выборки наблюдений (диаграмма 1).

корреляционный регрессионный анализ

Характер расположения точек на диаграмме позволяет сделать предварительный вывод о том, что связь между переменными прямая, т.е. увеличение одной из переменных ведет увеличению условной (групповой) средней другой.

Связь между переменными в диапазоне  достаточно тесная, однако в диапазоне  имеются точки выброса, т.е. точки, находящиеся на достаточно отдаленном расстоянии от общего массива точек. Им соответствуют данные по Брянской, Белгородской и Воронежской областям.

Диаграмма 1.



Сделаем предположения, что:

1. данные по Брянской области являются точкой выброса;
2. данные по Белгородской области являются точкой выброса;
3. данные по Воронежской области являются точкой выброса;
4. данные по Брянской и Белгородской областям являются точками выброса;
5. данные по Брянской и Воронежской областям являются точками выброса;
6. данные по Белгородской и Воронежской областям являются точками выброса
7. данные по Брянской, Белгородской и Воронежской областям являются точками выброса.
   1. Определить коэффициенты корреляции

Для заданного массива переменных коэффициент корреляции *r* = 0,454 (рассчитан при помощи функции Microsoft Excel КОРРЕЛ).

Коэффициент корреляции *r* > 0, следовательно, корреляционная связь между переменными прямая, что подтверждает предварительный вывод, сделанный в п.1.1.

Коэффициент корреляции *r* принял значение на отрезке [-1; 1], следовательно, мы можем оценить тесноту связи случайных величин, заданных массивами, при помощи шкалы Чеддока:

|  |  |  |
| --- | --- | --- |
| Теснота связи | Значение коэффициента корреляции при наличии: | |
| прямой связи | обратной связи |
| Слабая | 0,1 - 0,3 | (-0,1) - (-0,3) |
| Умеренная | 0,3 - 0,5 | (-0,3) - (-0,5) |
| Заметная | 0,5 - 0,7 | (-0,5) - (-0,7) |
| Высокая | 0,7 - 0,9 | (-0,7) - (-0,9) |
| Весьма высокая | 0,9 - 0,99 | (-0,9) - (-0,99) |

Коэффициент корреляции *r* принадлежит интервалу (0,3; 0,5), следовательно, связь между переменными умеренная.

Рассчитаем коэффициенты корреляции, исключая данные по субъектам РФ согласно выдвинутым предположениям:

|  |
| --- |
| *r* = 0,116 |
| *r* = 0,821 |
| *r* = 0,578 |
| *r* = 0,511 |
| *r* = 0,455 |
| *r* = 0,949 |
| *r* = 0,824 |

Анализ полученных коэффициентов показывает, что предположение 5 верно, т.е. данные по Брянской и Белгородской областям являются точками выброса (исключение точек, соответствующих указанным субъектам РФ, из корреляционного поля не повлекло за собой значительного изменения коэффициента корреляции). Все остальные предположения считаем неверными. Кроме того, отмечается значительное увеличение тесноты связи между переменными при исключении из корреляционного поля точек, соответствующих данным по Белгородской и Воронежской областям (предположение 6), и её значительное уменьшение при исключении данных по Брянской области.

* 1. Оценить статистическую значимость вычисленных коэффициентов корреляции

Оценку статистической значимости коэффициентов корреляции будем проводить при помощи t-критерия Стьюдента на уровне значимости α= 0,05.

|  |  |  |
| --- | --- | --- |
| Парный двухвыборочный t-тест для средних | | |
| *r* = 0,454 |  |  |
|  | *Переменная 1* | *Переменная 2* |
| Среднее | 927,08 | 258,42 |
| Дисперсия | 1101362,746 | 73524,47289 |
| Наблюдения | 10 | 10 |
| Корреляция Пирсона | 0,454062283 |  |
| Гипотетическая разность средних | 0 |  |
| df | 9 |  |
| t-статистика | 2, 208751921 |  |
| P (T<=t) одностороннее | 0,027278104 |  |
| t критическое одностороннее | 1,833112923 |  |
| P (T<=t) двухстороннее | 0,054556208 |  |
| t критическое двухстороннее | 2,262157158 |  |

Расчетное значение критерия Стьюдента *tр* = 2,21 меньше критического *tКРИТ*= 2,306 (взято из таблицы t-распределений Стьюдента при числе степеней свободы *n*-2= 8 и величине погрешности α= 0,05), из чего делаем вывод о незначимости коэффициента корреляции.

Так как исключение данных по Брянской и Белгородской областям согласно ранее проведенному анализу не значительно влияет на коэффициент корреляции, то при нахождении t-критерия Стьюдента для выборки исходных данных при предположении 5 получим практически аналогичный результат.

|  |  |  |
| --- | --- | --- |
| Парный двухвыборочный t-тест для средних | | |
| *r* = 0,455 |  |  |
|  | *Переменная 1* | *Переменная 2* |
| Среднее | 696,0125 | 137,9125 |
| Дисперсия | 607399,8755 | 9534,678393 |
| Наблюдения | 8 | 8 |
| Корреляция Пирсона | 0,510547416 |  |
| Гипотетическая разность средних | 0 |  |
| df | 7 |  |
| t-статистика | 2,149664636 |  |
| P (T<=t) одностороннее | 0,034323806 |  |
| t критическое одностороннее | 1,894578604 |  |
| P (T<=t) двухстороннее | 0,068647613 |  |
| t критическое двухстороннее | 2,364624251 |  |

Расчетное значение критерия Стьюдента *tр* = 2,15 меньше критического *tКРИТ*= 2,45 (взято из таблицы t-распределений Стьюдента при числе степеней свободы *n*-2= 6 и величине погрешности α= 0,05). Коэффициент корреляции незначим.

* 1. Сделать итоговые выводы.

Между показателями работы грузовых автомобилей крупных и средних организаций автомобильного транспорта в 2006 году существует умеренная статистическая взаимосвязь. Для проведения анализа данные по Брянской и Белгородской областям можно не учитывать.

# Задание 2

1. По исходным данным выполнить регрессионный анализ:
   1. Рассчитать параметры уравнения линейной парной регрессии;

Линейное уравнение парной регрессии имеет вид:

,

где  - оценка условного математического ожидания *y*;

*b*0, *b*1 - эмпирические коэффициенты регрессии, подлежащие определению.

Эмпирические коэффициенты регрессии *b*0, *b*1 будем определять с помощью инструмента Регрессия MS Excel.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ВЫВОД ИТОГОВ | | | | | |
|  | |  | |  | |
| *Регрессионная статистика* | | | |  | |
| Множественный R | | | | 0,454062283 | |
| R-квадрат | | | | 0, 206172557 | |
| Нормированный R-квадрат | | | | 0,106944127 | |
| Стандартная ошибка | | | | 991,7552465 | |
| Наблюдения | | | | 10 | |
|  | |  | |  | |
| Дисперсионный анализ | | | | | | | | |
|  | *df* | | *SS* | | *MS* | | *F* | *Значимость F* |
| Регрессия | 1 | | 2043636,965 | | 2043636,965 | | 2,078 | 0,187 |
| Остаток | 8 | | 7868627,751 | | 983578,469 | |  |  |
| Итого | 9 | | 9912264,716 | |  | |  |  |
|  |  | |  | |  | |  |  |
|  | *Коэффициенты* | | *Стандартная ошибка* | |
| Y-пересечение | 472,939 | | 444,546 | |
| Переменная X 1 | 1,757 | | 1,219 | |

Таким образом, эмпирические коэффициенты регрессии соответственно равны *b*0 = 472,94, *b*1 = 1,76.

Тогда уравнение парной линейной регрессии, связывающей объемы перевозимых грузовыми автомобилями крупных и средних организаций автомобильного транспорта в 2006 году, *y* с величиной расходов на перевозку *x*, имеет вид:



* 1. Дать с помощью общего (среднего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом

Оценим тесноту статистической связи между расходами на перевозки, производимые грузовыми автомобилями крупных и средних организаций в 2006 году, *x* и их объемами *y*. Эта оценку производится с помощью коэффициента корреляции *rxy*.

Величина этого коэффициента рассчитана в п.1.2 и равна *r* = 0,454. Как говорилось выше, связь между переменными умеренная прямая.

Параметр R-квадрат представляет собой квадрат коэффициента корреляции *rxy*2 и называется коэффициентом детерминации. Величина данного коэффициента характеризует долю дисперсии зависимой переменной *y*, объясненную регрессией (объясняющей переменной *x*).

Соответственно величина 1 - *rxy*2 характеризует долю дисперсии переменной *y*, вызванную влиянием всех остальных, неучтенных в эконометрической модели объясняющих переменных.

Таким образом, доля всех неучтенных в полученной эконометрической модели объясняющих переменных приблизительно составляет: 1 - 0, 206 = 0,794 или 79,4%. Степень связи объясняющей переменной *x* с зависимой переменной *y* определяется при помощи коэффициента эластичности, который для модели парной линейной регрессии определяется в виде:

.

Тогда



Следовательно, при изменении величины расходов на грузоперевозки на 1% их объем изменяется на 0,49%.

* 1. Оценить качество уравнения с помощью средней ошибки аппроксимации.

Средняя ошибка аппроксимации оценивается по зависимости:



Для этого исходную таблицу дополняем двумя колонками, в которых определяем значения, рассчитанные с использованием зависимости и значения разности .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Перевезено грузов, тыс. тонн** | **Расходы, млн, руб** |  |  |
| Владимирская | 594,6 | 258,3 | 926,869 | 0,559 |
| Брянская | 3178,9 | 656,5 | 1626,656 | 0,488 |
| Белгородская | 523,8 | 824,4 | 1921,720 | 2,669 |
| Воронежская | 2572,3 | 220,1 | 859,737 | 0,666 |
| Ивановская | 308,5 | 73,8 | 602,633 | 0,953 |
| Костромская | 580,5 | 82,7 | 618,274 | 0,065 |
| Рязанская | 203,7 | 65,4 | 587,871 | 1,886 |
| Смоленская | 389,3 | 86,6 | 625,128 | 0,606 |
| Тульская | 225,8 | 36,5 | 537,083 | 1,379 |
| Ярославская | 693,4 | 279,9 | 964,828 | 0,391 |
|  |  |  |  | сумма = 9,662 |

Средняя ошибка аппроксимации составляет:



Практически полагают, что значение средней ошибки аппроксимации не должно превышать 12-15% для грубого приближения регрессии к реальной зависимости. В нашем случае ошибка чрезмерна велика.

Воспользуемся результатами исследования, проведенного в п.1, т. е исключим из рассматриваемой выборки данные по Брянской и Белгородской областям.

В этом случае уравнение парной регрессии примет вид:

.

Доля неучтенных в полученной эконометрической модели объясняющих переменных составит: 1 - 0,260 = 0,74 или 74%.

Коэффициент эластичности составит:

,

а средняя ошибка аппроксимации:



Исключение точек выброса из рассматриваемой выборки снизило ошибку аппроксимации, однако её значение превышает допустимое значение.

* 1. Оценить статистическую надежность результатов регрессионного моделирования с помощью критерия Стъюдента и F-критерия Фишера.

Проведем более строгую оценку статистической надежности моделирования с помощью F-критерия Фишера.

Для этого проверим нулевую гипотезу *H*0 о статистической незначимости полученного уравнения регрессии по условию: если при заданном уровне значимости α = 0,05 теоретическое (расчетное) значение F-критерия (*F*) больше его критического значения (*FКРИТ*), то нулевая гипотеза отвергается и полученное уравнение регрессии принимается значимым.

Расчетное значение *F*, определенное с помощью инструмента Регрессия MS Excel, составило *F* = 2,078.

Критическое значение *FКРИТ* определим при помощи статистической функции FРАСПОБР. Входными параметрами функции является уровень значимости (вероятность) и число степеней свободы 1 и 2. Для модели парной регрессии число степеней свободы соответственно равно 1 (одна объясняющая переменная) и *n -* 2 = 10 - 2 = 8.

*FКРИТ* = 5,318.

Расчетное значение *F* = 2,078 меньше критического *FКРИТ* = 5,318, поэтому нулевая гипотеза *H*0 о статистической незначимости уравнения регрессии  принимается, что подтверждает вывод, сделанный в п.2.3.

При расчете критериев Фишера для сокращенной выборки (исключая данные по Брянской и Белгородской областям) получаем аналогичный результат.

*F* = 2,115 < *FКРИТ* = 5,987.

* 1. Сделать итоговые выводы.

1. Уравнение парной линейной регрессии, связывающее объемы перевозимых грузовыми автомобилями крупных и средних организаций автомобильного транспорта в 2006 году, *y* с величиной расходов на перевозку *x*, имеет вид:



При этом доля всех неучтенных в полученной эконометрической модели объясняющих переменных приблизительно составляет 79,4%, т.е. учтенными остаются лишь 20,6 % параметров.

Величина коэффициента эластичности говорит о том, что при изменении величины расходов на грузоперевозки на 1% их объем должен измениться на 0,49%.

Расчет средней ошибки аппроксимации (*А* = 96,62 %), а также анализ при помощи критерия Фишера показал, что полученное уравнение регрессии не соответствует реальной зависимости (в силу большой доли неучтенных в зависимости параметров).

2. Уравнение парной линейной регрессии для выборки исходных данных, исключающей данные по Брянской и Белгородской областям, которые по результатам выполнения задания 1 признаны точками выброса, имеет вид:



При этом доля всех неучтенных в полученной эконометрической модели объясняющих переменных приблизительно составляет 74%.

Величина коэффициента эластичности говорит о том, что при изменении величины расходов на грузоперевозки на 1% их объем должен измениться на 0,81%.

Расчет средней ошибки аппроксимации (*А* = 56,25 %), а также анализ при помощи критерия Фишера показал, что полученное уравнение регрессии также не соответствует реальной зависимости (в силу большой доли неучтенных в зависимости параметров).

Результаты регрессионного моделирования не надежны.