***Лабораторная работа №1.***

***Определение корреляционной взаимосвязи переменных***

***Что такое корреляция, регрессионная зависимость, доверительный интервал определение***

**Корреляция** — это показатель, отражающий взаимосвязь между объектами или событиями.

**Корреля́ция, или корреляцио́нная зави́симость** — статистическая взаимосвязь двух или более случайных величин, при этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

Регрессия -  это **зависимость математического ожидания случайной величины от одной или нескольких других случайных величин.**

Для исследования силы связи между переменными широко применяется ***корреляционный анализ***, позволяющий, совместно с ***регрессионным анализом***, решать задачи прогнозирования, планирования и анализа хозяйственной деятельности экономических систем (предприятий, фирм, отраслей и т.д.).

В случае лишь одной независимой переменой X в качестве меры связи между результативным признаком Y и независимой переменной X служит ***коэффициент корреляции***. Он оценивается по выборке объема n связанных пар наблюдений (xi, yi).

Коэффициент корреляции характеризует степень линейной зависимости  и , проявляющейся в том, что при возрастании одной случайной величины другая проявляет тенденцию тоже возрастать (либо убывать). В первом случае  (положительная корреляция), во втором  (отрицательная корреляция). Для любых двух случайных величин . Если зависимость отсутствует, то коэффициент корреляции равен нулю.

В случае ***нескольких*** переменных необходимо последовательно вычислять коэффициенты корреляции по нескольким рядам числовых данных. Полученные коэффициенты сводят в таблицы, называемые ***корреляционными матрицами***. ***Корреляционная матрица*** представляет собой квадратную матрицу, на пересечении строки и столбца которой находится коэффициент корреляции между соответствующими переменными.

***А2.*** В случае ***нелинейной парной регрессии*** а также для ***многофакторных линейных корреляционных*** моделей для оценки тесноты связи X и Y вместо ***коэффициента*** корреляции используется ***индекс*** корреляции, рассчитываемый по формуле:

, (2)

где– экспериментальные значения,  - теоретические значения, рассчитанные по уравнению регрессии, - усредненное значение экспериментальных результатов.

***B. Множественная корреляция***

В пакете ***Анализ данных*** MS Excel имеется инструмент ***Корреляция***, позволяющий автоматизировать процесс расчета коэффициента корреляции (для двух случайных величин) и корреляционной матрицы (для многомерной выборки). Рассмотрим его действие на ряде примеров. Для начала проверьте правильность расчета коэффициента корреляции в примере 1.

***Пример 2.***

Имеются ежемесячные данные наблюдений за состоянием погоды и посещаемостью музея и парка, приведенные в таблице

|  |  |  |
| --- | --- | --- |
| Число ясных дней | Количество посетителей  музея | Количество посетителей парка |
| 8 | 495 | 132 |
| 14 | 503 | 348 |
| 20 | 380 | 643 |
| 25 | 305 | 865 |
| 20 | 348 | 743 |
| 15 | 465 | 541 |

Необходимо определить, существует ли взаимосвязь между состоянием погоды и посещаемостью музеев и парков.

***Решение***

В меню ***Сервис*** выберем пункт ***Анализ данных*** и выберем ***Корреляция***. В окне диалога укажем входной интервал (численные значения трех столбцов), выберем расположение данных ***по столбцам***, укажем выходной диапазон. На выходе получим следующую корреляционную матрицу

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Столбец 1* | *Столбец 2* | *Столбец 3* |
| Столбец 1 | 1 |  |  |
| Столбец 2 | -0,9218543 | 1 |  |
| Столбец 3 | 0,9745756 | -0,9193752 | 1 |

Видно, что корреляция между состоянием погоды и посещаемостью музея равна -0,92, а между состоянием погоды и посещаемостью парка 0,97. Таким образом, выявлены сильная степень обратной линейной зависимости между посещаемостью музея и количеством солнечных дней (), практически линейная (очень сильная прямая связь) между посещаемостью парка и состоянием погоды ().

***Самостоятельное задание 1.***

С целью анализа взаимосвязи показателей эффективности производства продукции: производительности труда (), фондоотдачи () и материалоемкости производства () была отобрана группа из десяти однотипных предприятий. Получены следующие данные

|  |  |  |  |
| --- | --- | --- | --- |
| № предприятия |  |  |  |
| 1 | 6,0 | 2,0 | 25 |
| 2 | 4,9 | 0,8 | 30 |
| 3 | 7,0 | 2,7 | 20 |
| 4 | 6,7 | 3,0 | 21 |
| 5 | 5,8 | 1,0 | 28 |
| 6 | 6,1 | 2,1 | 26 |
| 7 | 5,0 | 0,9 | 30 |
| 8 | 6,9 | 2,6 | 22 |
| 9 | 6,8 | 3,0 | 20 |
| 10 | 5,9 | 1,1 | 29 |

Требуется рассчитать корреляционную матрицу системы.

Проанализируйте полученные результаты.

***Самостоятельное задание 2.***

Определите, имеется ли взаимосвязь между годовым уровнем инфляции (%), ставкой рефинансирования (%) и курсом доллара (руб./%) по следующим данным ежегодных наблюдений:

|  |  |  |
| --- | --- | --- |
| Уровень инфляции | Ставка рефинансирования | Курс $ |
| 84 | 85 | 441 |
| 45 | 55 | 980 |
| 56 | 65 | 1400 |
| 34 | 40 | 1960 |
| 23 | 28 | 2030 |

**Множественная линейная регрессия**

***Тема:***

Модели простой и множественной линейной регрессии

***Цели:***

1. Научиться строить модели множественной регрессии.
2. Рассмотреть способы оценивания статистической значимости параметров модели.
3. Рассмотреть способы оценивания адекватности модели.

***Ход работы***

***Построение модели множественной линейной регрессии в Ехсеl***

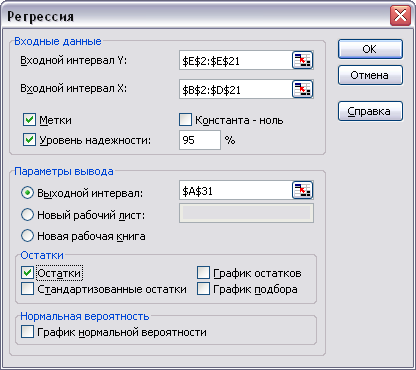
1) В Ехсеl постройте следующую таблицу:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 1 | 12 | 2 | 8 | 139 |
| 2 | 17 | 5 | 12 | 182 |
| 3 | 14 | 6 | 11 | 164 |
| 4 | 13 | 4 | 9 | 150 |
| 5 | 16 | 3 | 12 | 176 |
| 6 | 15 | 2 | 9 | 168 |
| 7 | 13 | 6 | 10 | 173 |
| 8 | 11 | 5 | 13 | 145 |
| 9 | 15 | 4 | 10 | 175 |
| 10 | 13 | 6 | 11 | 157 |
| 11 | 12 | 5 | 14 | 142 |
| 12 | 15 | 3 | 14 | 151 |
| 13 | 13 | 2 | 8 | 148 |
| 14 | 16 | 5 | 11 | 186 |
| 15 | 17 | 5 | 10 | 201 |
| 16 | 15 | 4 | 13 | 169 |
| 17 | 11 | 5 | 12 | 160 |
| 18 | 14 | 4 | 12 | 151 |
| 19 | 13 | 2 | 14 | 129 |
| 20 | 15 | 3 | 11 | 163 |

2) Выполните команду меню *Сервис -Надстройки* и установите флажок напротив надстройки *Пакет анализа.*

3) Выполните команду меню *Сервис-Анализ данных* и выберите инструмент *Регрессия.*

4) Если Ваша таблица начинается в ячейке Аl, то заполните диалог следующим образом:



5) С помощью функции СТЪЮДРАСПОБР( 1 *-γ* , *Т-М-1)* рассчитайте критические значения распределения Стьюдента  для уровней *γ* = 0,90, 0,95,0,99, где *М -–* это число независимых переменных, *Т* – количество точек.

1. Проверьте статистическую значимость параметров модели для уровней *γ* = 0,90, 0,95,0,99 , для этого используйте расчетные значения распределения Стьюдента из графы *t-статιιстика.*
2. Постройте доверительные интервалы для всех параметров модели

вида 

для уровней *γ* = 0,90, 0,95,0,99, где *j = 0…M* Значения несмещенной дисперсии для параметров модели  находятся в графе *Стандартная ошибка.*

1. На основе исправленного коэффициента детерминации из графы *Нормированный R-квадрат* дайте предварительную оценку адекватности построенной модели.
2. С помощью функции FРАСПОБР(1*-γ* , M+1, *Т-М*-1) рассчитайте критические распределения Фишера  для уровней *γ* = 0,90, 0,95,0,99
3. Проверьте гипотезу о статистической значимости коэффициента детерминации для уровней *γ* = 0,90,0,95,0.99, используя для этого расчетное значение распределения Фишера из графы *F.*

***Самостоятельная работа***

Бюджетное обследование случайно выбранных семей дало следующие результаты (в тыс.у.е.)

|  |  |  |  |
| --- | --- | --- | --- |
| **Семья** | **Накопления** | **Доходы** | **Имущество** |
|  | **у** | **х1** | **х2** |
| 1 | 1,3 | 11 | 20 |
| 2 | 2,3 | 19 | 14 |
| 3 | 1,8 | 13 | 12 |
| 4 | 1,4 | 14 | 8 |
| 5 | 1,1 | 11 | 10 |
| 6 | 1,2 | 17 | 6 |
| 7 | 2,7 | 23 | 16 |
| 8 | 1,9 | 11 | 15 |
| 9 | 1,5 | 13 | 8 |
| 10 | 2,1 | 20 | 17 |
| 11 | 1,7 | 15 | 12 |

**Задания:**

1. Оцените регрессию у на х1 и х2..
2. Спрогнозируйте накопление семьи, имеющей доход 15 тыс.руб. и имущество стоимостью 18 тыс.руб.
3. Если предположить, что доход семьи возрос на 5 тыс.руб., в то время как стоимость имущества не изменилась. Оцените рост накоплений.
4. Оцените, как возрастут накопления семьи, если ее доход вырос на 3 тыс.руб., а стоимость имущества на 5 тыс.руб.
5. Определите, как изменятся накопления, если доход увеличится на 10%.
6. Определите силу связи накопления от дохода и имущества.
7. Оцените, насколько тесно связаны между собой доход и накопления, а также доход и имущество.
8. Определите, какой из факторов в большей степени объясняет изменение результативного показателя.
9. Проверьте, насколько точно построенная модель выражает изучаемую закономерность.