**Министерство образования и науки Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**Псковский государственный университет**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Лабораторная работа**

по дисциплине

**« Технологии обработки информации »**

Выполнил студент

группы 1022-03

Ковалевский Р.А.

Проверил: Полетаева О.А.

Дата проверки:\_\_\_\_\_\_\_\_

**Псков**

**2018**

**Задание 1:**

**Набрать таблицу 1 и по ней заполнить таблицу 2, используя функции статистики.**

|  |  |
| --- | --- |
| Вариант 3 | |
| Прожиточный минимум | Заработная плата |
| в рублях | в рублях |
| 206 | 369 |
| 216 | 410 |
| 229 | 441 |
| 208 | 383 |
| 235 | 443 |
| 279 | 526 |
| 177 | 335 |
| 232 | 436 |
| 192 | 369 |
| 229 | 439 |
| 200 | 380 |
| 303 | 583 |
| 182 | 360 |
| 221 | 459 |
| 242 | 472 |

***Таблица 2***

|  |  |  |  |
| --- | --- | --- | --- |
| *Прожиточный уровень* |  | *Заработная плата* |  |
|  | |  | |
| Среднее | 223,4 | Среднее | 427 |
| Стандартная ошибка | 8,741145179 | Стандартная ошибка | 17,17944508 |
| Медиана | 221 | Медиана | 436 |
| Мода | 229 | Мода | 369 |
| Стандартное отклонение | 33,85430971 | Стандартное отклонение | 66,5357047 |
| Дисперсия выборки | 1146,114286 | Дисперсия выборки | 4427 |
| Эксцесс | 1,153523614 | Эксцесс | 0,810188883 |
| Асимметричность | 0,979757122 | Асимметричность | 0,905535964 |
| Интервал | 126 | Интервал | 248 |
| Минимум | 177 | Минимум | 335 |
| Максимум | 303 | Максимум | 583 |
| Сумма | 3351 | Сумма | 6405 |
| Счет | 15 | Счет | 15 |
| Уровень надежности(95,0%) | 18,74789182 | Уровень надежности(95,0%) | 36,84624512 |

**Стандартной отклонение**= 33,85430971(ПМ) 66,5357047(ЗП) квадратный корень из дисперсии выборки - мера того, насколько широко разбросаны точки данных относительно их среднего.

**Дисперсия выборки** = 1146,114286 - среднее арифметическое квадратов отклонений значений от их среднего.

**Эксцесс** показывает "остроту пика" распределения, характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. *Э*ксцесс положительный (1,153523614 (ПМ) 0,810188883(ЗП)) обозначает относительно остроконечное распределение (пик заострен). Если он был бы отрицательный эксцесс обозначал относительно сглаженное распределение (пик закруглен).

**Асимметрия =** 0,979757122(ПМ) 0,905535964 (ЗП) показывает отклонение распределения от симметричного. Если распределение имеет длинный правый хвост, асимметрия положительна; если длинный левый хвост - отрицательна.

Корреляционный анализ применяется для количественной оценки взаимосвязи двух наборов данных, представленных в безразмерном виде.

На данном графике большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция) - наличие прямой линейной связи.

Таким образом, можем установить зависимость между переменными x и y.

ЗАДАНИЕ 2.

**1.Построить диаграмму рассеивания по таблице 1.**

Диаграмма рассеивания при­меняется для выявления зависимости одной переменной величины. На данном графике прямая зависимость.

**2. Рассчитать коэффициент корреляции Пирсона между двумя массивами (x и y) при помощи функции MS Excel PEARSON (массив1;массив2).**

**Коэффициенты корреляции** показывают высокую зависимость между Заработной платой и прожиточным минимумом.

Показывает, на сколько единиц увеличивается значение, принимаемое ПМ,  если ЗП изменится на единицу своего изменения. В нашем случае знак коэффициента регрессии положительный и равен 0,979, что близко к 1 следовательно, связь является сильной и положительной.

**4. Используя пакет анализа MS Excel и инструмент анализа "Корреляция", построить корреляционную матрицу по таблице 1**

|  |  |  |
| --- | --- | --- |
| *Коэффициент корреляции* | *Столбец 1* | *Столбец 2* |
| Столбец 1 | 1 |  |
| Столбец 2 | 0,978808343 | 1 |

**5. Используя пакет анализа MS Excel и инструмент анализа "Регрессия", решить задачу регрессионного анализа по таблице 1. Построить графики остатков, подбора и нормального распределения**

**Вывод итогов**

Величина R-квадрат, называемая также мерой определенности, характеризует качество полученной регрессионной прямой. Это качество выражается степенью соответствия между исходными данными и регрессионной моделью (расчетными данными). Мера определенности всегда находится в пределах интервала [0;1].

В большинстве случаев значение R-квадрат находится между этими значениями, называемыми экстремальными, т.е. между нулем и единицей.

Если значение R-квадрата близко к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных. И наоборот, значение R-квадрата, близкое к нулю, означает плохое качество построенной модели.

В нашем примере мера определенности равна 0,97881, что говорит об очень хорошей подгонке регрессионной прямой к исходным данным.

множественный R - коэффициент множественной корреляции R - выражает степень зависимости независимых переменных (X) и зависимой переменной (Y).

Множественный R равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы.

В простом линейном регрессионном анализе множественный R равен коэффициенту корреляции Пирсона. Действительно, множественный R в нашем случае равен коэффициенту корреляции Пирсона из предыдущего примера (0,978808343).

|  |  |
| --- | --- |
| ВЫВОД ИТОГОВ |  |
|  |  |
| *Регрессионная статистика* |  |
| Множественный R | 0,978808343 |
| R-квадрат | 0,958065772 |
| Нормированный R-квадрат | 0,954840062 |
| Стандартная ошибка | 14,1394146 |
| Наблюдения | 15 |

Здесь даны коэффициент регрессии **b(Y-пересечение)** и смещение по оси ординат, т.е. константа **a(переменная X1)**.Исходя из расчетов, можем записать уравнение регрессии таким образом:**Y= x\*a+b**

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициентов регрессии (коэффициента b).

Если знак при коэффициенте регрессии - положительный, связь зависимой переменной с независимой будет положительной. В нашем случае

знак коэффициента регрессии положительный, следовательно, связь также является положительной.

Если знак при коэффициенте регрессии - отрицательный, связь зависимой переменной с независимой является отрицательной (обратной).

В представлены результаты вывода остатков.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |
| Y-пересечение | -2,755683801 | 25,20238961 | -0,109342163 | 0,91460092 |
| Переменная X 1 | 1,923704941 | 0,111622938 | 17,23395728 | 2,46243E-10 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| -57,20213639 | 51,69076879 | -57,20213639 | 51,69076879 |
| 1,682558245 | 2,164851637 | 1,682558245 | 2,164851637 |

**Вывод остатка**

При помощи этой части отчета мы можем видеть отклонения каждой точки от построенной линии регрессии. Имея уравнение регрессии, задача прогнозирования сводится к решению уравнения **Y= x\*a+b** с известными значениями x.

|  |  |  |
| --- | --- | --- |
| ВЫВОД ОСТАТКА |  |  |
|  |  |  |
| *Наблюдение* | *Предсказанное Y* | *Остатки* |
| 1 | 393,527534 | -24,52753403 |
| 2 | 412,7645834 | -2,764583437 |
| 3 | 437,7727477 | 3,227252331 |
| 4 | 397,3749439 | -14,37494391 |
| 5 | 449,3149773 | -6,314977315 |
| 6 | 533,9579947 | -7,957994715 |
| 7 | 337,7400907 | -2,740090741 |
| 8 | 443,5438625 | -7,543862492 |
| 9 | 366,5956649 | 2,404335145 |
| 10 | 437,7727477 | 1,227252331 |
| 11 | 381,9853044 | -1,985304383 |
| 12 | 580,1269133 | 2,873086703 |
| 13 | 347,3586154 | 12,64138455 |
| 14 | 422,3831081 | 36,61689186 |
| 15 | 462,7809119 | 9,219088099 |

|  |  |
| --- | --- |
| ВЫВОД ВЕРОЯТНОСТИ |  |
|  |  |
| *Персентиль* | *Y* |
| 3,333333333 | 335 |
| 10 | 360 |
| 16,66666667 | 369 |
| 23,33333333 | 369 |
| 30 | 380 |
| 36,66666667 | 383 |
| 43,33333333 | 410 |
| 50 | 436 |
| 56,66666667 | 439 |
| 63,33333333 | 441 |
| 70 | 443 |
| 76,66666667 | 459 |
| 83,33333333 | 472 |
| 90 | 526 |
| 96,66666667 | 583 |

Таким образом, в результате использования регрессионного анализа в пакете Microsoft Excel мы:

* построили уравнение регрессии;
* установили форму зависимости и направление связи между переменными - положительная линейная регрессия, которая выражается в равномерном росте функции;
* установили направление связи между переменными;
* оценили качество полученной регрессионной прямой;
* смогли увидеть отклонения расчетных данных от данных исходного набора;
* предсказали будущие значения зависимой переменной.

Если функция регрессии определена, интерпретирована и обоснована, и оценка точности регрессионного анализа соответствует требованиям, можно считать, что построенная модель и прогнозные значения обладают достаточной надежностью.

Прогнозные значения, полученные таким способом, являются средними значениями, которые можно ожидать.

**ЗАДАНИЕ 3**

**Результаты прогнозирования зависимой переменной Y на шесть шагов вперед представить** [**в таблице 4**](http://www.intuit.ru/studies/courses/6/6/lecture/172?page=4#table.8.4).

|  |  |
| --- | --- |
| Таблица прогнозирования зависимости переменных X и Y | |
| 206 | 369 |
| 216 | 410 |
| 229 | 441 |
| 208 | 383 |
| 235 | 443 |
| 279 | 526 |
| 177 | 335 |
| 232 | 436 |
| 192 | 369 |
| 229 | 439 |
| 200 | 380 |
| 303 | 583 |
| 182 | 360 |
| 221 | 459 |
| 242 | 472 |
| 230 | 458 |
| 231 | 462 |
| 232 | 466 |
| 233 | 469 |
| 234 | 473 |
| 234 | 477 |

**Среднесрочный прогноз - это прогноз на 3-5% от объема наблюдений, но не более 7-12 шагов вперед; также под этим типом прогноза понимают прогноз на один или половину сезонного цикла.**