

Решение работы в Statistica выполнено на сайте www.matburo.ru
Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
©МатБюро. Решение задач по математике, экономике, программированию

Лабораторная работа в Statistica Множественная регрессия

(В файле представлена только часть решения)

Задание

Цели работы: анализ связи между несколькими независимыми переменными и зависимой переменной, построение математической модели и определение ее статистической надежности, формулирование представления о возможных путях экономического развития, что позволяет спрогнозировать ту или иную ситуацию, предвидеть будущие значения экономических показателей.

Задачи:

1. Используя исходные данные повести их общую описательную статистику, оценить показатели вариации каждого признака.
2. По имеющимся n -наблюдениям за совместным изменением параметров x и y необходимо определить степень аналитической зависимости этих величин, с помощью корреляционного анализа. Провести анализ мультиколлинеарности факторов. Проанализировать линейные коэффициенты парной и частной корреляции. Сравнить полученные результаты с результатами, полученные в программной среде Statistica 7.
3. С помощью пакета "Анализ данных" построить уравнение множественной регрессии в стандартизованной и естественной форме. Обосновать отбор факторов в модели линейной множественной регрессии и построить модель регрессии только с информативными факторами. Оценить качество полученной модели регрессии. Выполнить матричные вычисления метода ридж-регрессии для стандартизованной модели.

Решение работы в Statistica выполнено на сайте www.matburo.ru
Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
©МатБюро. Решение задач по математике, экономике, программированию

4. С помощью матричных вычислений провести расчеты точечного прогноза результативного показателя и построить доверительные интервалы прогноза. Проанализировать изменение ширины доверительного интервала при варьировании величины факторных признаков в диапазоне от их минимальных до максимальных значений.

5. Для уравнения множественной регрессии с информативными факторами провести анализ ряда остатков с помощью пакетов "Анализ данных". Построить графики остатков, выполнить тесты Гольдфельда-Квандта для проверки гомоскедастичности остатков.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Исходные данные

В качестве исходных данных возьмем следующие показатели о социально-экономическом положении регионов Сибирского ФО.:

Субъекты	Рождаемость (на 1000 женщин фертильного возраста)	Ожидаемая продолжительность жизни при рождении, лет	Число разводов на 1000 браков	Уровень безработицы, %	Среднедушевой доход населения, руб.	Жилплощадь на душу населения, кв. метров	Заболееваемость на 1000 населения	Аборты (на 100 родов)	Коэффициент демографической нагрузки на 1000 населения	Величина прожиточного минимума (руб./чел.)
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Республика Алтай	88,5	66,8	473	11,6	14278	18,9	885,2	51	734	6505
Республика Бурятия	68,2	66,79	537	7,9	17119	20,2	654,8	64	666	6766
Республика Тыва	97	61,09	295	18,4	11933	13,2	622,4	57	729	6462
Республика Хакасия	62,9	67,64	572	7,9	15991	21,6	834,1	68	683	6405
Алтайский край	56,2	69,11	576	6,2	13629	22,4	1087,5	49	709	6257
Забайкальский край	63,1	66,24	595	10,6	17336	20	737,5	63	651	6784
Красноярский край	55,3	68,42	467	5,5	22138	22,9	825,8	75	626	7715
Иркутская область	61,6	66,32	684	7,8	17720	22,1	920,6	68	676	6557
Кемеровская область	54,8	66,76	583	7,1	18386	22,7	818,5	78	691	5698
Новосибирская област	54,9	69,72	538	5,6	20637	22,2	722,2	81	645	6989
Омская область	58,1	69,25	516	6,9	19469	23,2	960,5	43	647	5773
Томская область	51	70,07	554	8,4	17876	22,2	708,7	70	609	7077

Необходимо построить уравнение множественной регрессионной зависимости суммарного коэффициента рождаемости (на 1 женщину фертильного возраста) от факторов (все остальные переменные).

Решение работы в Statistica выполнено на сайте www.matburo.ru
Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
©МатБюро. Решение задач по математике, экономике, программированию

Описательная статистика

Первым этапом исследования будет ознакомление с данными путем общего обобщения. Это необходимо для получения первых выводов и применения стратегического решения. Для этого используем описательную статистику, которая позволит обобщить первичные данные. Все расчеты описательных статистик сводятся к группировке данных по их значениям, построению распределения их частот, выявлению центральных тенденций распределения и, наконец, к оценке разброса данных по отношению к найденной центральной тенденции.

Основные показатели описательной статистики:

- Среднее значение (среднее арифметическое, медиана, мода)
- Усредненное значение
- Разброс (диапазон разброса данных)
- Дисперсия
- Стандартное (среднеквадратическое) отклонение
- Квартили
- Доверительный интервал

Провести расчет описательной статистики можно в различных программных средах. Ниже будет приведена описательная статистика, выполненная в MS Excel.

Реализация в MS Excel. Вкладка *Данные* → *Анализ данных* → *Описательная статистика* → *Входной интервал* - выбираем диапазон исходных данных → Выбираем «*Итоговая статистика*».

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?pl=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Описательная статистика

Входные данные
 Входной интервал: \$B\$2:\$K\$14
 Группирование: по столбцам по строкам
 Метки в первой строке

Параметры вывода
 Выходной интервал: \$A\$16
 Новый рабочий лист
 Новая рабочая книга
 Итоговая статистика
 Уровень надежности: 95 %

OK
Отмена
Справка

Получаем.

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Среднее	64,3	67,35083333	532,5	8,6583	17209,333	20,9667	814,82	63,92	672,166667	6582,333333
Стандартная ошибка	4,102272098	0,6930209	27,2312	1,0305	843,85593	0,80069	38,991	3,437	11,3068988	159,6128523
Медиана	59,85	67,22	546	7,85	17528	22,15	822,15	66	671	6531
Мода	#Н/Д	#Н/Д	#Н/Д	7,9	#Н/Д	22,2	#Н/Д	68	#Н/Д	#Н/Д
Стандартное отклонение	14,2106874	2,400694817	94,3316	3,5697	2923,2027	2,77369	135,07	11,9	39,1682463	552,9151395
Дисперсия выборки	201,9436364	5,763335606	8898,45	12,743	8545113,9	7,69333	18243	141,7	1534,15152	305715,1515
Эксцесс	1,94119489	3,791957155	3,46407	5,026	-0,209793	5,86153	-0,06	-0,8	-0,77669287	0,668932816
Асимметричность	1,658704514	-1,588618957	-1,2851	2,0902	-0,201718	-2,28364	0,4974	-0,35	0,1485446	0,25290977
Интервал	46	8,98	389	12,9	10205	10	465,1	38	125	2017
Минимум	51	61,09	295	5,5	11933	13,2	622,4	43	609	5698
Максимум	97	70,07	684	18,4	22138	23,2	1087,5	81	734	7715
Сумма	771,6	808,21	6390	103,9	206512	251,6	9777,8	767	8066	78988
Счет	12	12	12	12	12	12	12	12	12	12

Далее рассчитываем вариативность каждого показателя.

Это коэффициент вариации - мера относительного разброса случайной величины; показывает, какую долю среднего значения этой величины составляет её средний разброс. Коэффициент вариации равен отношению стандартного отклонения к среднему значению.

При значении коэффициента вариации менее 10% выборка показателя слабо вариативна, от 10% до 20% - средне вариативна, свыше 20% - высоко вариативна, с такими переменными трудно работать, так как их тяжело предсказать.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Полученные коэффициенты вариации

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Среднее	64,3	67,35083333	532,5	8,6583	17209,333	20,9667	814,82	63,92	672,166667	6582,333333
Стандартная ошибка	4,102272098	0,6930209	27,2312	1,0305	843,85593	0,80069	38,991	3,437	11,3068988	159,6128523
Медиана	59,85	67,22	546	7,85	17528	22,15	822,15	66	671	6531
Мода	#Н/Д	#Н/Д	#Н/Д	7,9	#Н/Д	22,2	#Н/Д	68	#Н/Д	#Н/Д
Стандартное отклонение	14,2106874	2,400694817	94,3316	3,5697	2923,2027	2,77369	135,07	11,9	39,1682463	552,9151395
Дисперсия выборки	201,9436364	5,763335606	8898,45	12,743	8545113,9	7,69333	18243	141,7	1534,15152	305715,1515
Эксцесс	1,94119489	3,791957155	3,46407	5,026	-0,209793	5,86153	-0,06	-0,8	-0,77669287	0,668932816
Асимметричность	1,658704514	-1,588618957	-1,2851	2,0902	-0,201718	-2,28364	0,4974	-0,35	0,1485446	0,25290977
Интервал	46	8,98	389	12,9	10205	10	465,1	38	125	2017
Минимум	51	61,09	295	5,5	11933	13,2	622,4	43	609	5698
Максимум	97	70,07	684	18,4	22138	23,2	1087,5	81	734	7715
Сумма	771,6	808,21	6390	103,9	206512	251,6	9777,8	767	8066	78988
Счет	12	12	12	12	12	12	12	12	12	12
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Вариация	22%	4%	18%	41%	17%	13%	17%	19%	6%	8%

Вариативность зависимой переменной (22%) – достаточно большая.

Вариативность объясняющих переменных разная – от 4% до 41%.

Чем меньше вариативность, тем лучше подходит данная переменная в качестве объясняющей, поэтому будем стараться использовать в регрессии именно переменные x_1 (ожидаемая продолжительность жизни), x_8 (коэффициент демографической нагрузки).

Переменная x_3 имеет большую вариативность и хуже подходит в качестве объясняющей переменной.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Парная и частная корреляция

Корреляция – это статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Различают парную, частную и множественную корреляцию.

- Парная корреляция – это связь между двумя признаками (результативным и факторным или между двумя факторными).
- Частная корреляция – это связь между двумя признаками (результативным и факторным или между двумя факторными) при фиксированном значении других факторных признаков.
- Множественная корреляция – это связь между результативным и двумя или более факторными признаками, включенными в исследование.

Рассчитаем коэффициенты парной корреляции.

А	В	С	Д	Е	Ф	Г	Н	И	Ж
Субъекты	Рождаемость (на 1000 женщин фертильного возраста)	Ожидаемая продолжительность жизни при рождении лет	Число разводов на 1000 браков	Уровень безработицы, %	Среднедушевой доход населения, руб.	Жилая площадь на душу населения, кв. метров	Заболелость на 1000 населения	Аборты (на 100 родов)	Коэффициент демографической нагрузки на 1000 населения
	Y	X1	X2	X3	X4	X5	X6	X7	X8
Республика Алтай	88,5	66,8	473	11,6	14278	18,9	885,2	51	734
Республика Бурятия	68,2								
Республика Тыва	97								
Республика Хакасия	62,9								
Алтайский край	56,2								
Забайкальский край	63,1								
Красноярский край	55,3								
Иркутская область	61,6								
Кемеровская область	54,8								
Новосибирская область	54,9								
Омская область	58,1								
Томская область	51								

Корреляция

Входные данные
 Входной интервал:

Группирование:
 по столбцам
 по строкам

Метки в первой строке

Параметры вывода
 Выходной интервал:
 Новый рабочий дист.:

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Получаем.

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y	1									
X1	-0,813	1								
X2	-0,692	0,530	1							
X3	0,885	-0,867	-0,676	1						
X4	-0,712	0,598	0,343	-0,718	1					
X5	-0,916	0,879	0,737	-0,963	0,710	1				
X6	-0,303	0,419	0,420	-0,463	-0,031	0,528	1			
X7	-0,418	0,137	0,246	-0,323	0,552	0,261	-0,462	1		
X8	0,738	-0,622	-0,321	0,557	-0,813	-0,589	0,203	-0,435	1	
X9	-0,111	0,163	-0,124	-0,108	0,375	0,007	-0,391	0,451	-0,512	1

Цветом выделены высокие коэффициенты корреляции (по модулю более 0,7).

Исходя из значений коэффициентов парной корреляции можно предположить, что наилучшую тесную связь «рождаемость, у» имеет с «ожидаемой продолжительностью жизни, x_1 », «уровнем безработицы, x_2 », «средним доходом, x_3 », «удельной жилплощадью, x_4 » и «коэффициентом демографической нагрузки, x_8 ».

В то же время имеет место высокая связь между этими факторами (x_1x_3 , x_1x_5 , x_3x_4 , x_3x_5 , x_4x_5 , x_4x_8), поэтому следует рассмотреть коэффициенты частной корреляции, которые дают более точную характеристику тесноты связи анализируемых признаков.

Составляем полную корреляционную матрицу и матрицу корреляции переменных X.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?pl=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y	1	-0,813	-0,692	0,885	-0,712	-0,916	-0,303	-0,418	0,738	-0,111
X1	-0,813	1	0,530	-0,867	0,598	0,879	0,419	0,137	-0,622	0,163
X2	-0,692	0,530	1	-0,676	0,343	0,737	0,420	0,246	-0,321	-0,124
X3	0,885	-0,867	-0,676	1	-0,718	-0,963	-0,463	-0,323	0,557	-0,108
X4	-0,712	0,598	0,343	-0,718	1	0,710	-0,031	0,552	-0,813	0,375
X5	-0,916	0,879	0,737	-0,963	0,710	1	0,528	0,261	-0,589	0,007
X6	-0,303	0,419	0,420	-0,463	-0,031	0,528	1	-0,462	0,203	-0,391
X7	-0,418	0,137	0,246	-0,323	0,552	0,261	-0,462	1	-0,435	0,451
X8	0,738	-0,622	-0,321	0,557	-0,813	-0,589	0,203	-0,435	1	-0,512
X9	-0,111	0,163	-0,124	-0,108	0,375	0,007	-0,391	0,451	-0,512	1

Матрица корреляции зависимых переменных									
	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1	0,530	-0,867	0,598	0,879	0,419	0,137	-0,622	0,163
X2	0,530	1	-0,676	0,343	0,737	0,420	0,246	-0,321	-0,124
X3	-0,867	-0,676	1	-0,718	-0,963	-0,463	-0,323	0,557	-0,108
X4	0,598	0,343	-0,718	1	0,710	-0,031	0,552	-0,813	0,375
X5	0,879	0,737	-0,963	0,710	1	0,528	0,261	-0,589	0,007
X6	0,419	0,420	-0,463	-0,031	0,528	1	-0,462	0,203	-0,391
X7	0,137	0,246	-0,323	0,552	0,261	-0,462	1	-0,435	0,451
X8	-0,622	-0,321	0,557	-0,813	-0,589	0,203	-0,435	1	-0,512
X9	0,163	-0,124	-0,108	0,375	0,007	-0,391	0,451	-0,512	1

Находим обратную матрицу к корреляционной, ее определитель и матрицу алгебраических дополнений.

Обратная матрица										
	60,165	-3,334	-2,458	-43,793	-29,238	-22,443	29,714	29,562	-63,835	-21,093
	-3,334	11,120	3,558	2,646	6,925	-18,094	3,250	1,647	3,149	-1,794
	-2,458	3,558	4,309	0,292	4,732	-10,372	0,761	-0,983	2,382	-0,034
	-43,793	2,646	0,292	58,245	22,851	52,975	-29,592	-24,522	57,509	21,036
	-29,238	6,925	4,732	22,851	22,906	-0,301	-12,346	-14,618	34,007	9,263
	-22,443	-18,094	-10,372	52,975	-0,301	110,673	-40,617	-25,008	46,929	23,690
	29,714	3,250	0,761	-29,592	-12,346	-40,617	28,054	21,572	-41,315	-15,345
	29,562	1,647	-0,983	-24,522	-14,618	-25,008	21,572	20,219	-36,571	-13,515
	-63,835	3,149	2,382	57,509	34,007	46,929	-41,315	-36,571	80,630	27,456
	-21,093	-1,794	-0,034	21,036	9,263	23,690	-15,345	-13,515	27,456	11,738

Определитель									
	7700611,5								

Матрица алгебраических дополнений										
	463305488	-25673751	-18927864	-337236489	-225152967	-172824176	228812863	227645557	-491571140	-162432614
	-25673751	85629532	27401568	20378331	53328248	-139338683	25025331	12679222	24249010	-13818654
	-18927864	27401568	33185437	2251560,3	36436248	-79872225	5860477,8	-7570301,4	18341413	-260319,8
	-337236489	20378331	2251560,3	448519800	175965177	407936123	-227876497	-188835393	442856992	161988181
	-225152967	53328248	36436248	175965177	176388768	-2314165,3	-95071355	-112570561	261876062	71329463
	-172824176	-139338683	-79872225	407936123	-2314165,3	852251058	-312776006	-192575551	361384524	182430595
	228812863	25025331	5860477,8	-227876497	-95071355	-312776006	216030688	166120023	-318152657	-118166341
	227645557	12679222	-7570301,4	-188835393	-112570561	-192575551	166120023	155697448	-281620604	-104077404
	-491571140	24249010	18341413	442856992	261876062	361384524	-318152657	-281620604	620899585	211426238
	-162432614	-13818654	-260319,8	161988181	71329463	182430595	-118166341	-104077404	211426238	90390409

На их основе рассчитываем коэффициенты частной корреляции и коэффициент множественной корреляции.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Коэффициенты частной корреляции		
x ₁	0,13	
x ₂	0,15	
x ₃	0,74	
x ₄	0,79	
x ₅	0,28	
x ₆	-0,72	
x ₇	-0,85	
x ₈	0,92	
x ₉	0,79	
Коэффициент множественной корреляции		
	0,992	

Коэффициент множественной корреляции

$$R_{y x_1 x_2 \dots x_9} = \sqrt{1 - \frac{\Delta r}{\Delta r_x}} = 0,992$$

Δr - определитель полной матрицы коэффициентов парной корреляции;

Δr_x - определитель матрицы межфакторных коэффициентов парной корреляции.

Далее определим коэффициенты корреляции в пакете STATISTICA 7.

Заносим данные.

Data: Данные (10v by 12c)										
	1 Y	2 X1	3 X2	4 X3	5 X4	6 X5	7 X5	8 X7	9 X8	10 X9
1	88,5	66,8	473	11,6	14278	18,9	885,2	51	734	6505
2	68,2	66,79	537	7,9	17119	20,2	654,8	64	666	6766
3	97	61,09	295	18,4	11933	13,2	622,4	57	729	6462
4	62,9	67,64	572	7,9	15991	21,6	834,1	68	683	6405
5	56,2	69,11	576	6,2	13629	22,4	1087,5	49	709	6257
6	63,1	66,24	595	10,6	17336	20	737,5	63	651	6784
7	55,3	68,42	467	5,5	22138	22,9	825,8	75	626	7715
8	61,6	66,32	684	7,8	17720	22,1	920,6	68	676	6557
9	54,8	66,76	583	7,1	18386	22,7	818,5	78	691	5698
10	54,9	69,72	538	5,6	20637	22,2	722,2	81	645	6989
11	58,1	69,25	516	6,9	19469	23,2	960,5	43	647	5773
12	51	70,07	554	8,4	17876	22,2	708,7	70	609	7077

Корреляционная матрица.

Решение работы в Statistica выполнено на сайте www.matburo.ru
 Переходите на сайт, смотрите больше примеров или закажите свою работу
https://www.matburo.ru/ex_ec.php?p1=ecstatis
 ©МатБюро. Решение задач по математике, экономике, программированию

Correlations (Данные)										
Marked correlations are significant at p < ,05000										
N=12 (Casewise deletion of missing data)										
Variable	Y	X1	X2	X3	X4	X5	X5	X7	X8	X9
Y	1,000000	-0,812757	-0,692128	0,885405	-0,711863	-0,915826	-0,302959	-0,418400	0,738304	-0,110573
X1	-0,812757	1,000000	0,529825	-0,867171	0,598240	0,878736	0,419434	0,136592	-0,621982	0,162758
X2	-0,692128	0,529825	1,000000	-0,675702	0,343211	0,736629	0,420116	0,246139	-0,320696	-0,123762
X3	0,885405	-0,867171	-0,675702	1,000000	-0,718306	-0,962572	-0,463433	-0,323330	0,557336	-0,107739
X4	-0,711863	0,598240	0,343211	-0,718306	1,000000	0,710006	-0,030557	0,551980	-0,812920	0,375459
X5	-0,915826	0,878736	0,736629	-0,962572	0,710006	1,000000	0,528324	0,260910	-0,589295	0,006665
X5	-0,302959	0,419434	0,420116	-0,463433	-0,030557	0,528324	1,000000	-0,462418	0,202993	-0,390708
X7	-0,418400	0,136592	0,246139	-0,323330	0,551980	0,260910	-0,462418	1,000000	-0,434936	0,451233
X8	0,738304	-0,621982	-0,320696	0,557336	-0,812920	-0,589295	0,202993	-0,434936	1,000000	-0,511685
X9	-0,110573	0,162758	-0,123762	-0,107739	0,375459	0,006665	-0,390708	0,451233	-0,511685	1,000000

Как видим корреляционная матрица совпадает с матрицей Excel.

Коэффициенты парной корреляции показывают к включению в модель регрессии переменные: x_1 , x_3 , x_4 , x_5 , x_8 .

При этом между ними также обнаруживается корреляционная зависимость, что приводит к мультиколлинеарности.

Потому смотрим частные коэффициенты корреляции.

Согласно частным коэффициентам, x_1 и x_5 – первые на очереди исключения из регрессии.

Коэффициенты частной корреляции	
x_1	0,13
x_2	0,15
x_3	0,74
x_4	0,79
x_5	0,28
x_6	-0,72
x_7	-0,85
x_8	0,92
x_9	0,79

Обязательные переменные к включению в модель регрессии переменные: x_3 , x_4 , x_8 .

(решение представлено не полностью)