ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Экономика и менеджмент в машиностроении»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к практической работе
по дисциплинам «Статистика: теория статистики»,
«Статистика: социально–экономическая статистика»
и «Теоретические основы прогнозирования»

«Выявление множественной линейной корреляции»

Авторы
Борисова Л.В.,
Сербулова Н.М.,
Борисова Д.В.

Ростов-на-Дону, 2015

Аннотация

Методические указания предназначены для проведения практических работ по дисциплинам «Статистика: теория статистики», «Статистика: социально–экономическая статистика» и «Теоретические основы прогнозирования» со студентами, обучающимися по направлениям 38.03.01, 38.03.02 и другим направлениям.

Авторы

д.т.н., профессор
Борисова Л.В.,

к.т.н., доцент
Сербулова Н.М.,

ассистент
Борисова Д.В.

Оглавление

ВВЕДЕНИЕ

При изучении сложных явлений необходимо учитывать более двух случайных факторов. Правильное представление о природе связи между этими факторами можно получить только в том случае, если подвергнуть исследованию сразу все рассматриваемые случайные факторы. Совместное изучение трех и более случайных факторов позволит исследователю установить более или менее обоснованные предположения о причинных зависимостях между изучаемыми явлениями. Простой формой множественной связи является линейная зависимость между тремя признаками. Случайные факторы обозначаются как X_1, X_2 и X_3. Парный коэффициенты корреляции между X_1 и X_2 обозначается как r_{12}, соответственно между X_1 и X_3 · r_{13}, между X_2 и X_3 · r_{23}. В качестве меры тесноты линейной связи трех признаков используют множественные коэф-фициенты корреляции, обозначаемые R_{123}, R_{213}, R_{312} и частные коэффициенты корреляции, обозначаемые $r_{12.3}$, $r_{13.2}$, $r_{23.1}$.

1. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

Множественный коэффициент корреляции $R_{1.23}$ трех факторов · это показатель тесноты линейной связи между одним из факторов (индекс перед точкой) и совокупностью двух других факторов (индексы после точки).

Значения коэффициента R всегда находятся в пределах от 0 до 1. При приближении R к единице степень линейной связи трех признаков увеличивается.

Между коэффициентом множественной корреляции, например R_{213}, и двумя коэффициентами парной корреляции r_{12} и r_{23} существует соотношение: каждый из парных коэффициентов не может превышать по абсолютной величине R_{213}.

Формулы для вычисления множественных коэффициентов корреляции при известных значениях коэффициентов парной корреляции r_{12}, r_{13} и r_{23} имеют вид:

$$R_{1\cdot 23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}};$$

$$R_{2\cdot 13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}};$$

$$R_{3\cdot 12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}}$$

Квадрат коэффициента множественной корреляции R^2 называется *коэффициентом множественной детерминации.* Он показывает долю вариации зависимой переменной под воздействием изучаемых факторов.

Значимость множественной корреляции оценивается по F-критерию:

$$F = \frac{R^2}{1 - R^2}\left(\frac{n-1}{k-1}\right),$$

где n – объем выборки; k – число факторов. В нашем случае $k = 3$.

Нулевая гипотеза о равенстве множественного коэффициента корреляции в совокупности нулю (H_0:R=0) принимается, если $F_ф < F_t$, и отвергается, если $F_ф \geq F_т$.

Теоретическое значение F-критерия определяется для $\nu_1 = k - 1$ и $\nu_2 = n - k$ степеней свободы и принятого уровня значимости α (прил. 1).

Пример вычисления коэффициента множественной корреляции. При изучении взаимосвязи между факторами были получены коэффициенты парной корреляции (n =15): r_{12}=0,6; r_{13} = 0,3; r_{23} = - 0,2.

Необходимо выяснить зависимость признака X_2 от признака X_1 и X_3, т. е. рассчитать коэффициент множественной корреляции:

$$R_{2\cdot13} = \sqrt{\frac{(0{,}6)^2 + (0{,}3)^2 - 2\times0{,}6\times0{,}3\times(-0{,}2)}{1-(0{,}2)^2}} = 0{,}74;$$

$$F = \frac{(0{,}74)^2}{1-(0{,}74)^2}\left(\frac{15-3}{2}\right) = 7{,}33.$$

Табличное значение F-критерия при ν_1 = 2 и ν_2 = 15 – 3 = 12 степенях свободы при α = 0,05 $F_{0,05}$ = 3,89 и при α = 0,01 $F_{0,01}$ = 6,93.

Таким образом, взаимосвязь между признаками $R_{2.13}$ = 0,74 значима на 1%-м уровне значимости $F_ф > F_{0,01}$.

Судя по коэффициенту множественной детерминации R^2 = $(0{,}74)^2$ = 0,55, вариация признака X_2 на 55% связана с действием изучаемых факторов, а 45% вариации ($1-R^2$) не может быть объяснено влиянием этих переменных.

2. ЧАСТНАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

Частный коэффициент корреляции - это показатель, измеряющий степень сопряженности двух признаков.

Математическая статистика позволяет установить корреляцию между двумя признаками при постоянном значении третьего, не ставя специального эксперимента, а используя парные коэффициенты корреляции r_{12}, r_{13}, r_{23}.

Частные коэффициенты корреляции рассчитывают по формулам:

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}};\quad r_{13\cdot2}=\frac{r_{13}-r_{12}\cdot r_{23}}{\sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)}};\qquad(1)$$

$$r_{23\cdot1}=\frac{r_{23}-r_{12}\cdot r_{13}}{\sqrt{\left(1-r_{12}^2\right)\left(1-r_{13}^2\right)}};$$

Цифры перед точкой указывают, между какими признаками изучается зависимость, а цифра после точки - влияние какого признака исключается (элиминируется). Ошибку и критерий значимости частной корреляции определяют по тем же формулам, что и парной корреляции:

$$S_{r_{12\cdot3}}=\sqrt{\frac{1-r_{12.3}^2}{n-2}};\quad t=\frac{r}{s_r}.\qquad(2)$$

Теоретическое значение t-критерия определяется для $v = n - 2$ степеней свободы и принятого уровня значимости α (прил. 2).

Нулевая гипотеза о равенстве частного коэффициента корреляции в совокупности нулю (H_0: $r = 0$) принимается, если $t_ф < t_т$, и отвергается, если $t_ф \geq t_т$.

Частные коэффициенты могут принимать значения, заключенные между -1 и +1. Частные *коэффициенты детерминации* находят путем возведения в квадрат частных коэффициентов корреляции:

$D_{12.3} = r^2_{123}$; $d_{13.2} = r^2_{132}$; $d_{231} = r^2_{231}$.

Определение степени частного воздействия отдельных факторов на результативный признак при исключении (элиминировании) связи его с другими признаками, искажающими эту корреляцию, часто представляет большой интерес. Иногда бывает, что при постоянном значении элиминируемого признака нельзя подметить его статистического влияния на изменчивость других признаков. Чтобы уяснить технику расчета частного коэффициента корреляции, рассмотрим пример. Имеются три параметра *X*, *Y* и *Z*. Для объема выборки n = 180 определены парные коэффициенты корреляции

R_{xy} = 0,799; r_{xz} = 0,57; r_{yz} = 0,507.

По формуле (2) определим частные коэффициенты корреляции:

$$r_{xy\cdot z} = \frac{0{,}799 - 0{,}570 \times 0{,}507}{\sqrt{\left(1-0{,}570^2\right)\left(1-0{,}507^2\right)}} = 0{,}720;$$

$$r_{xz\cdot y} = \frac{0{,}570 - 0{,}799 \times 0{,}507}{\sqrt{\left(1-0{,}799^2\right)\left(1-0{,}507^2\right)}} = 0{,}318;$$

$$r_{yz\cdot x} = \frac{0{,}507 - 0{,}799 \times 0{,}570}{\sqrt{\left(1-0{,}799^2\right)\left(1-0{,}570^2\right)}} = 0{,}105.$$

Частный коэффициент корреляции между параметром *X* и *Y* с постоянным значением параметра *Z* (r_{xyz} = 0,720) показывает, что лишь незначительная часть взаимосвязи этих признаков в общей корреляции (r_{xy} = 0,799) обусловлена влиянием третьего признака (*Z*). Аналогичное заключение необходимо сделать и в отношении частного коэффициента корреляции между параметром *X* и параметром *Z* с постоянным значением параметра *Y* (r_{xzy} = 0,318 и r_{xz} = 0,57). Напротив, частный коэффициент корреляции между параметрами *Y* и *Z* с постоянным значением параметра *X* r_{yzx} = 0,105 значительно отличается от общего коэффициента корреляции r_{yz} = 0,507. Из этого видно, что если подобрать объекты с одинаковым значением параметра *X*, то связь между признаками *Y* и *Z* у них будет очень слабой, так как значительная часть в этой взаимосвязи обусловлена варьированием параметра *X*.

При некоторых обстоятельствах частный коэффициент корреляции может оказаться противоположным по знаку парному.

Например, при изучении взаимосвязи между признаками *X*, *Y* и *Z* · были получены парные коэффициенты корреляции (при n = 100): r_{xy} = 0,6; r_{xz}= 0,9; r_{yz} = 0,4.

Частные коэффициенты корреляции при исключении влияния третьего признака:

$$r_{xy\cdot z} = \frac{0,6 - 0,9 \times 0,4}{\sqrt{\left(1 - 0,9^2\right)\left(1 - 0,4^2\right)}} = 0,60;$$

$$\mathrm{r}_{\mathrm{xz}\cdot\mathrm{y}} = \frac{0,9 - 0,6 \times 0,4}{\sqrt{\left(1 - 0,6^2\right)\left(1 - 0,4^2\right)}} = 0,90;$$

$$r_{yz\cdot x} = \frac{0,4 - 0,6 \times 0,9}{\sqrt{\left(1 - 0,6^2\right)\left(1 - 0,9^2\right)}} = -0,40.$$

Из примера видно, что значения парного коэффициента и частного коэффициента корреляции разнятся в знаке.

Метод частной корреляции дает возможность вычислить частный коэффициент корреляции второго порядка. Этот коэффициент указывает на взаимосвязь между первым и вторым признаком при постоянном значении третьего и четвертого. Определение частного коэффициента второго порядка ведут на основе частных коэффициентов первого порядка по формуле:

$$r_{12\cdot 34} = \frac{r_{12\cdot 4} - r_{13\cdot 4} \cdot r_{23\cdot 4}}{\sqrt{\left(1 - r_{13\cdot 4}^2\right)\left(1 - r_{23\cdot 4}^2\right)}},$$

где $r_{12.4}$, r_{134}, r_{234} — частные коэффициенты первого порядка, значение которых определяют по формуле частного коэффициента, используя коэффициенты парной корреляции r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}.

3. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

В соответствии с вариантом определить:

– множественные коэффициенты корреляции, оценить их значимость;

– множественные коэффициенты детерминации и дать трактовку результатов;

– частные коэффициенты корреляции, оценить их значимость;

– частные коэффициенты детерминации и дать трактовку результатов.

В таблице приняты обозначения: уровень значимости – α; объем выборки – n.

№ варианта	n	r_{12}	r_{13}	r_{23}	α
1	15	0,7	0,55	0,64	0,05
2	75	0,7	0,55	0,64	0,01
3	10	0,79	0,59	0,51	0,1
4	100	0,79	0,59	0,51	0,01
5	24	0,51	0,68	0,9	0,1
6	7	0,51	0,68	0,9	0,05
7	45	–0,82	–0,43	0,5	0,1
8	10	–0,82	–0,43	0,5	0,01
9	20	0,7	0,6	0,8	0,05
10	70	0,7	0,6	0,8	0,1
11	7	0,68	0,59	0,9	0,01
12	69	0,68	0,59	0,9	0,05
13	5	0,85	–0,65	–0,7	0,1
14	45	0,85	–0,65	–0,7	0,05
15	12	0,7	0,55	0,9	0,05
16	140	0,7	0,55	0,9	0,1
17	8	0,58	0,4	0,8	0,1
18	27	0,58	0,4	0,8	0,01
19	10	0,69	0,8	0,8	0,5
20	50	0,69	0,8	0,8	0,1

ПРИЛОЖЕНИЕ 1

Значения критерия *F* при 5%–м уровне значимости

Степени свободы для меньшей дисперсии (знаменателя) ν_2	Степени свободы для большей дисперсии (числителя) ν_1													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	161	200	216	225	230	234	237	239	241	242	244	249	252	253
2	18,5	9,00	19,16	19,25	19,30	19,3	19,36	19,37	19,38	19,39	19,41	19,45	19,47	19,49
3	1	9,55	9,28	9,12	9,01	3	8,88	8,84	8,81	8,78	8,74	8,64	8,58	8,56
4	10,1	6,94	6,59	6,39	6,26	8,94	6,09	6,04	6,00	5,96	5,91	5,77	5,70	5,66
5	3	5,79	5,41	5,19	5,05	6,16	4,88	4,82	4,78	4,74	4,68	4,53	4,44	4,40
6	7,71	5,14	4,76	4,53	4,39	4,95	4,21	4,15	4,10	4,06	4,00	3,84	3,75	3,71
7	6,61	4,74	4,35	4,12	3,97	4,27	3,79	3,73	3,68	3,63	3,57	3,41	3,32	3,28
8	5,99	4,46	4,07	3,84	3,69	3,87	3,50	3,44	3,39	3,34	3,28	3,12	3,03	2,98
9	5,59	4,26	3,86	3,63	3,48	3,58	3,29	3,23	3,18	3,13	3,07	2,90	2,80	2,76
10	5,32	4,10	3,71	3,48	3,33	3,37	3,14	3,07	3,02	2,97	2,91	2,74	2,64	2,59
11	5,12	3,98	3,59	3,36	3,20	3,22	3,01	2,95	2,90	2,86	2,79	2,61	2,50	2,45
12	4,96	3,88	3,49	3,26	3,11	3,09	2,92	2,85	2,80	2,76	2,69	2,50	2,40	2,35
13	4,84	3,80	3,41	3,18	3,02	3,00	2,84	2,77	2,72	2,67	2,60	2,42	2,32	2,26
14	4,75	3,74	3,34	3,11	2,96	2,92	2,77	2,70	2,65	2,60	2,53	2,35	2,24	2,19
15	4,64	3,60	3,29	3,06	2,90	2,85	2,70	2,64	2,59	2,55	2,48	2,29	2,18	2,12
16	4,60	3,63	3,24	3,01	2,85	2,79	2,66	2,59	2,54	2,49	2,42	2,24	2,13	2,07
17	4,54	3,59	3,20	2,96	2,81	2,74	2,62	2,55	2,50	2,45	2,38	2,19	2,08	2,02
18	4,49	3,55	3,16	2,93	2,77	2,70	2,58	2,51	2,46	2,41	2,34	2,15	2,04	1,98
19	4,45	3,52	3,13	2,90	2,74	2,66	2,55	2,48	2,43	2,38	2,31	2,11	2,00	1,94
20	4,41	3,49	3,10	2,87	2,71	2,63	2,52	2,45	2,40	2,35	2,28	2,08	1,96	1,90
21	4,38	3,47	3,07	2,84	2,68	2,60	2,49	2,42	2,37	2,32	2,25	2,05	1,93	1,87
22	4,35	3,44	3,05	2,82	2,66	2,57	2,47	2,40	2,35	2,30	2,23	2,03	1,91	1,84
23	4,32	3,42	3,03	2,80	2,64	2,55	2,45	2,38	2,32	2,28	2,20	2,00	1,88	1,82
24	4,30	3,40	3,01	2,78	2,62	2,53	2,43	2,36	2,30	2,26	2,18	1,98	1,86	1,80
25	4,28	3,38	2,99	2,76	2,60	2,51	2,41	2,34	2,27	2,24	2,16	1,96	1,84	1,77
26	4,26	3,37	2,98	2,74	2,59	2,49	2,39	2,32	2,25	2,22	2,15	1,95	1,82	1,76
28	4,24	3,34	2.95	2,71	2,56	2,47	2,36	2,29	2,24	2,19	2,12	1,91	1,78	1,72
30	4,22	3,32	2 92	2,69	2,53	2,44	2,34	2,27	2,21	2,12	2,09	1,89	1,76	1,69
40	4,20	3,23	2 84	2,61	2,45	2,42	2,25	2,18	2,12	2,07	2,00	1,79	1,66	1,59
50	4,17	3,18	2,79	2,56	2,40	2,34	2,20	2,13	2,07	2,02	1,95	1,74	1,60	1,52
100	4,08	3,09	2,70	2,46	2,30	2,29	2,10	2,03	1,97	1,92	1,85	1,63	1,48	1,39
	4,03					2,19								
	3,94													

ПРИЛОЖЕНИЕ 2

Значения критерия Стьюдента (для одностороннего критерия) на 10, 5, 1 и 0,1%–ном уровне значимости

Число степеней свободы ν	Уровень значимости			
	0,1	0,05	0,01	0,001
2	6,31	12,71	63,68	636,61
3	2,92	4,30	9,92	31,60
4	2,35	3,18	5,84	12,92
5	2,13	2,78	4,60	8,61
6	2,02	2,57	4,03	6,87
7	1,94	2,45	3,71	5,96
8	1,89	2,36	3,50	5,41
9	1,86	2,31	3,36	5,04
10	1,83	2,26	3,25	4,78
11	1,81	2,23	3,17	4,59
12	1,80	2,20	3,11	4,44
13	1,78	2,18	3,05	4,32
14	1,77	2,16	3,01	4,22
15	1,76	2,14	2,98	4,14
16	1,75	2,13	2,95	4,07
17	1,75	2,12	2,92	4,02
18	1,74	2,11	2,90	3,97
19	1,73	2,10	2,88	3,92
20	1,73	2,09	2,86	3,88
21	1,72	2,09	2,85	3,85
22	1,72	2,08	2,83	3,82
23	1,72	2,07	2,82	3,79
24	1,71	2,07	2,81	3,77
25	1,71	2,06	2,80	3,75
26	1,71	2,06	2,79	3,73
27	1,71	2,06	2,78	3,71
28	1,70	2,05	2,77	3,69
29	1,70	2,05	2,76	3,67
30	1,70	2,05	2,75	3,66
40	1,70	2,04	2,75	3,65
60	1,68	2,02	2,70	3,55
120	1,67	2,00	2,66	3,46
∞	1,66	1,98	2,62	3,37
	1,64	1,96	2,58	3,29