

Дополнение к практическому заданию №1
«Корреляционный анализ»

Подготовила:
Студентка факультета ИТ
2 курса 1 группы
Кашперко Василиса Сергеевна

Минск, 2022

Ход работы

Формируем выборку стран, в которой должны быть страны, упомянутые в обеих сводках из статистического приложения). Выборка объемом 12 экземпляров включает 3 страны из Европы, 3 страны из Азии, 3 страны из Африки, 2 страны из Америки (Северной + Южной), и 1 страна из Австралии и Океании. Внутри каждой континентальной локации страны выбираются случайным образом.

Страна	Ряды измерений индексов		Ранговые числа		Разность рангов	
	Индекс(счастья)	Индекс(безработицы)	X _i	Y _i	d=x _i -y _i	d ²
Норвегия	7,39	3,30	1	9	-8	64
Новая Зеландия	7,28	4,10	2	3	-1	1
Германия	7,16	3,00	3	10	-7	49
Великобритания	7,06	3,90	4	5	-1	1
США	6,95	3,70	5	6	-1	1
Сингапур	6,38	4,10	6	4	2	4
Мексика	6,32	3,40	7	8	-1	1
Япония	5,94	2,30	8	11	-3	9
Гонконг	5,48	3,60	9	7	2	4
Нигер	5,07	8,10	10	2	8	64
Южная Африка	4,96	28,20	11	1	10	100
Чад	4,36	1,90	12	12	0	0
Σ			78	78		298

Сумма квадратов разности рангов $\sum d^2$:

$$\sum d^2 = 64+1+49+1+1+4+1+9+4+64+100+64=298$$

Коэффициент ранговой корреляции R_s по формуле Спирмена:

$$R_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}, \text{ где } n - \text{это количество пар значений}$$
$$R_s = 1 - \frac{6 \cdot 298}{12(12^2-1)} = -0,042$$

Вывод:

На основании полученного результата выявляем связь между изучаемыми признаками стратификации:

Так как коэффициент имеет отрицательный знак (-), то связь отрицательная.

По абсолютному значению коэффициента (от 0 до 1) оцениваем количественную меру связи:

R_s принадлежит отрезку $[0,09; 0,19]$ – статическая взаимосвязь очень слабая.

По итогу: на основании расчетного R_s делаем вывод о том, что между исследуемыми признаками стратификации существует очень слабая отрицательная связь. В нашем случае, связь между страновым индексом счастья и страновым индексом безработицы является очень слабой отрицательной корреляцией, если растет один индекс, другой будет уменьшаться, но темп прироста будет слабо заметен (чем меньше безработицы в стране, тем больше в ней счастливых людей). Мы установили эмпирическую закономерность, а теоретическая социология попытается ее объяснить.

Проводим проверку достоверности выявленной связи (корреляции) осуществляется сравнением R_s и $R_{s \text{ крит}}$.

Выбираем значение $p = 0,05$.

На основании того, что $R_s > R_{s \text{ крит}}$, наличие обнаруженной связи считается *достоверным* при $p = 0,05$.

На основании того, что $R_s < R_{s \text{ крит}}$, наличие обнаруженной связи считается *недостоверным* при $p = 0,05$.

Таблица критических значений $R_{s \text{ крит}}$

Число коррелируемых пар n	$p=0,05$	$p=0,01$	Число коррелируемых пар n	$p=0,05$	$p=0,01$
4	1,000	—	14	0,456	0,645
5	0,900	1,000	16	0,425	0,601
6	0,829	0,943	18	0,399	0,564
7	0,714	0,893	20	0,377	0,534
8	0,643	0,833	22	0,359	0,508
9	0,600	0,783	24	0,343	0,485
10	0,564	0,746	26	0,329	0,465
12	0,506	0,712	28	0,317	0,448
			30	0,306	0,432

В нашем примере количество коррелируемых пар – 12, следовательно, 8 строка / 2 столбец.

$R_s(0,042) < R_{s \text{ крит}}(0,506)$. Следовательно, обнаруженная **корреляция НЕ достоверна**))))))

Чего? :(

Тогда придется посчитать что-нибудь еще:

Ход работы

Страна	Ряды измерений индексов		Ранговые числа		Разность рангов	
	Индекс _(верх. закона)	Индекс _(ур. образования)	X _i	Y _i	d=x _i -y _i	d ²
Норвегия	0,89	0,93	1	2	-1	1
Германия	0,84	0,94	2	1	1	1
Новая Зеландия	0,83	0,93	3	4	-1	1
Велико-британия	0,80	0,93	4	3	1	1
Сингапур	0,79	0,84	5	8	-3	9
Япония	0,78	0,85	6	7	-1	1
Гонконг	0,76	0,88	7	6	1	1
США	0,72	0,90	8	5	3	9
Южная Африка	0,59	0,72	9	9	0	0
Чад	0,46	0,29	10	11	-1	1
Нигер	0,45	0,25	11	12	1	1
Мексика	0,44	0,70	12	10	2	2
Σ			78	78		28

Сумма квадратов разности рангов Σd^2 :

$$\Sigma d^2 = 28.$$

Коэффициент ранговой корреляции R_s по формуле Спирмена:

$$R_s = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}, \text{ где } n - \text{это количество пар значений}$$

$$R_s = 1 - \frac{6 \cdot 28}{12(12^2-1)} = 0,9021$$

Вывод:

На основании полученного результата выявляем связь между изучаемыми признаками стратификации:

Так как коэффициент имеет положительный знак (+), то связь положительная.

По абсолютному значению коэффициента (от 0 до 1) оцениваем количественную меру связи:

R_s принадлежит отрезку $[0,7; 0,99]$ – статическая взаимосвязь сильная.

По итогу: на основании расчетного R_s делается вывод о том, что между исследуемыми признаками стратификации существует сильная положительная связь. В нашем случае, связь между страновым индексом верховенства закона и страновым индексом уровня образования является сильной положительной корреляцией: если растет один индекс, будет расти и другой и темп прироста будет примерно одинаковым (чем больше в стране преобладает верховенство закона, тем выше в ней уровень образования). Мы установили эмпирическую закономерность, а теоретическая социология попытается ее объяснить.

Проводим проверку достоверности выявленной связи (корреляции) осуществляется сравнением R_s и $R_{s \text{ крит}}$.

Выбираем значение $p = 0,05$.

На основании того, что $R_s > R_{s \text{ крит}}$, наличие обнаруженной связи считается *достоверным* при $p = 0,05$.

На основании того, что $R_s < R_{s \text{ крит}}$, наличие обнаруженной связи считается *недостоверным* при $p = 0,05$.

Таблица критических значений $R_{s \text{ крит}}$

Число коррелируемых пар n	$p=0,05$	$p=0,01$	Число коррелируемых пар n	$p=0,05$	$p=0,01$
4	1,000	—	14	0,456	0,645
5	0,900	1,000	16	0,425	0,601
6	0,829	0,943	18	0,399	0,564
7	0,714	0,893	20	0,377	0,534
8	0,643	0,833	22	0,359	0,508
9	0,600	0,783	24	0,343	0,485
10	0,564	0,746	26	0,329	0,465
12	0,506	0,712	28	0,317	0,448
			30	0,306	0,432

В нашем примере количество коррелируемых пар – 12, следовательно, 8 строка / 2 столбец.

$R_s (0,902) < R_{s \text{ крит}} (0,506)$. Следовательно, обнаруженная **корреляция достоверна!**

Ура :)