

Анализ и представление данных психологического исследования

Лекция 3. Корреляционный анализ

- ***Корреляционный анализ*** - метод, позволяющий обнаружить зависимость между несколькими случайными величинами.
- ***Коэффициент корреляции*** – это показатель характера взаимного стохастического влияния изменения двух случайных величин. Может принимать значения от -1 до $+1$.

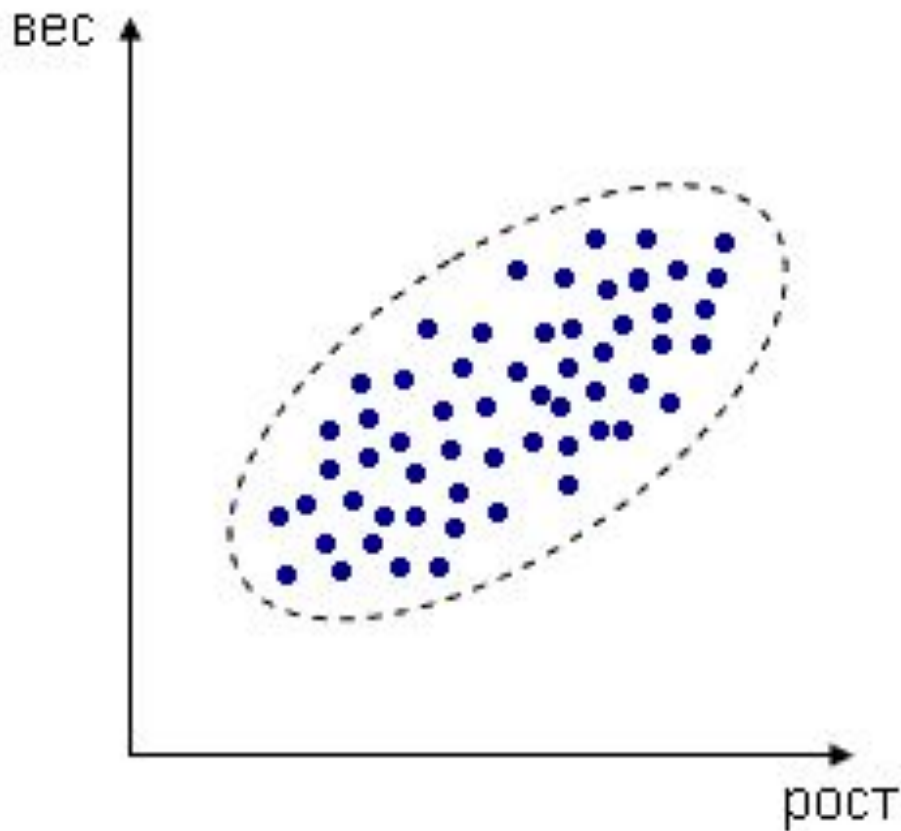
Смысл корреляционного анализа

- *Смысл корреляционного анализа* – знание о возможности строить прогноз относительно одной переменной на основании знания о другой.

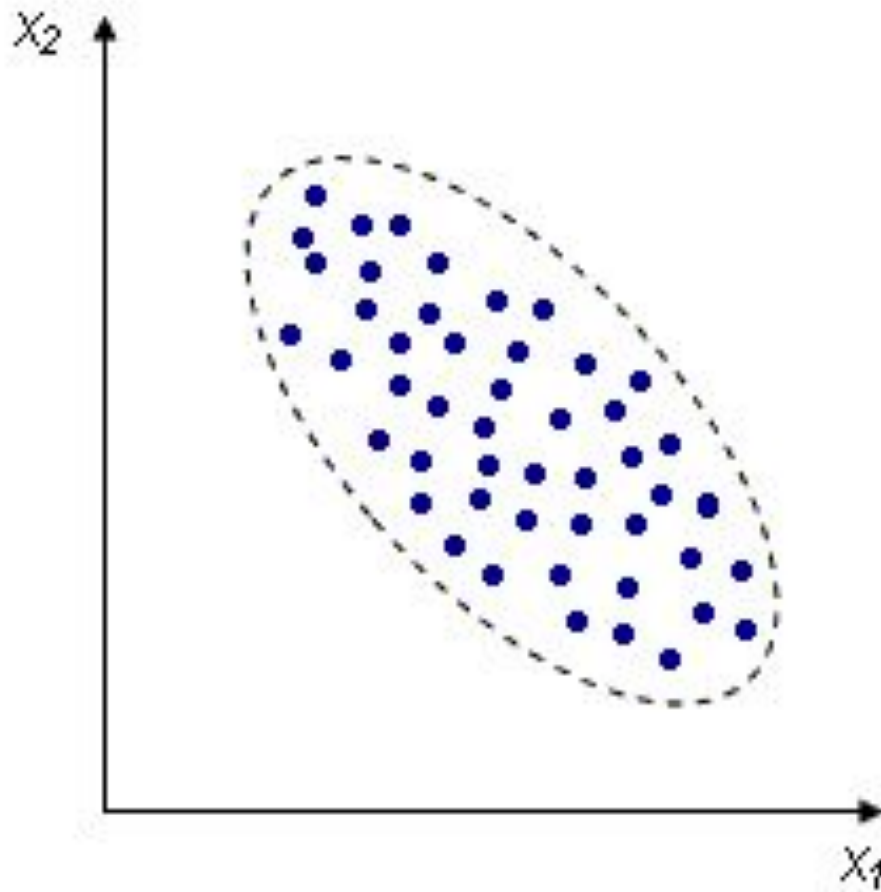
Интерпретация коэффициента корреляции

Значение	Интерпретация
$0 < r \leq 0,2$	Очень слабая корреляция
$0,2 < r \leq 0,5$	Слабая корреляция
$0,5 < r \leq 0,7$	Средняя корреляция
$0,7 < r \leq 0,9$	Сильная корреляция
$0,9 < r \leq 1$	Очень сильная корреляция

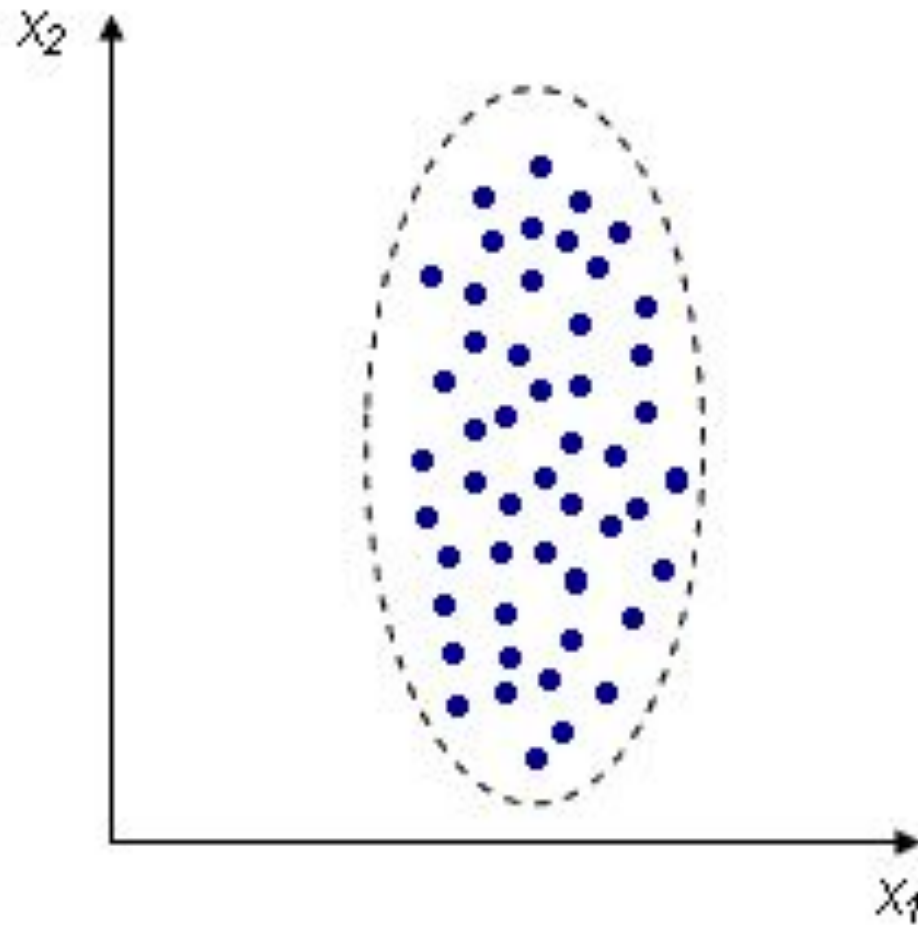
Положительная корреляция



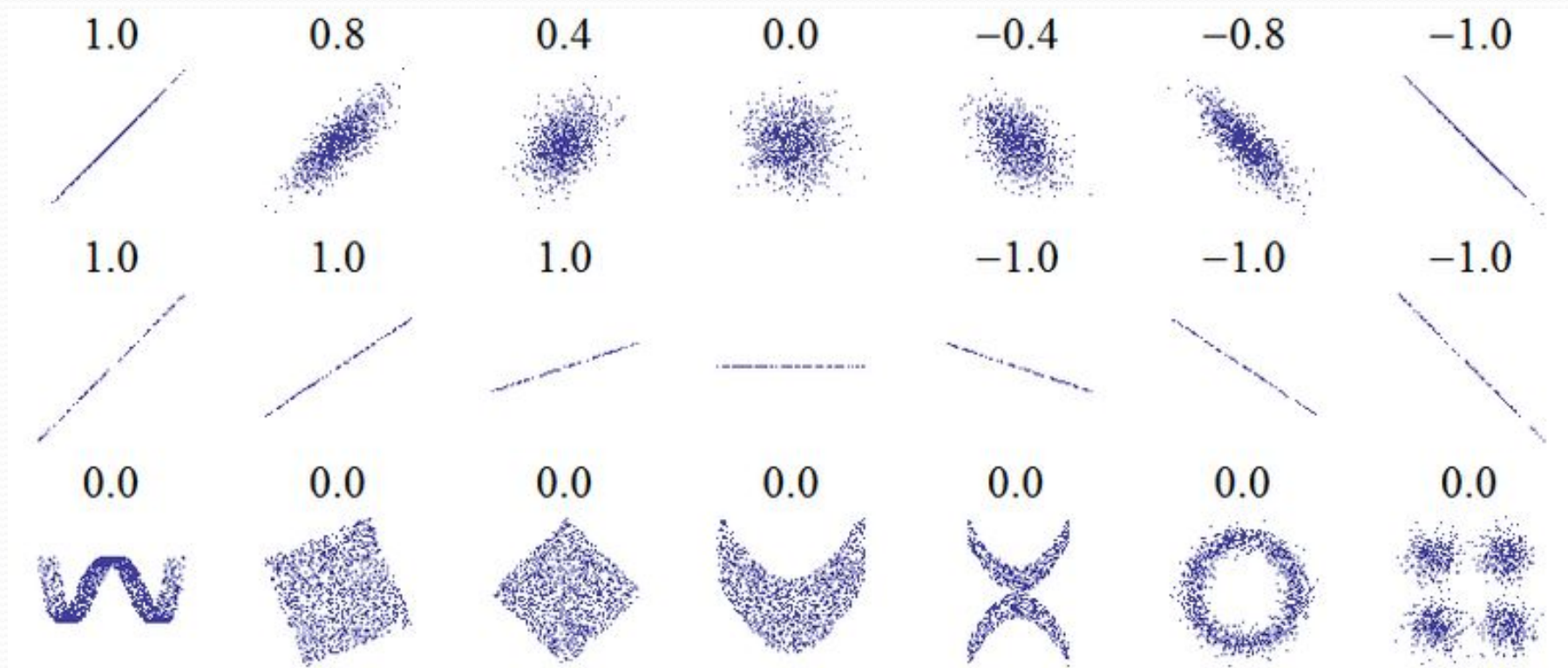
Отрицательная корреляция



Отсутствие корреляции



Графики распределений пар (x,y) с соответствующими коэффициентами корреляций



Ограничения корреляционного анализа

- Минимальное количества случаев для применения должно составлять от 25 до 100 пар наблюдений
- Отражает только **линейную связь**
- **Не даёт** основания утверждать, какая из переменных предшествует (является причиной изменений), или что переменные вообще **причинно связаны** между собой
- Не учитывает, что разные градации X могут по-разному «влиять» на Y

Коэффициенты корреляции и шкалы измерения

Тип шкалы		Мера связи
Переменная X	Переменная Y	
Номинальная	Номинальная	Коэфф. ассоциации Пирсона ϕ , Критерий χ^2
Номинальная	Порядковая	Рангово-бисериальный коэффициент R_{rb}
Номинальная	Интервальная или отношений	Бисериальный коэффициент R_{bis}
Порядковая	Порядковая, интервальная или отношений	Коэфф. корреляции Спирмена r_s Коэфф. корреляции Кендалла τ
Интервальная или отношений	Интервальная или отношений	Коэфф. корреляции Пирсона r_{xy}

Корреляции в номинальной шкале

Дихотомические шкалы

Высшее образование	Нравится ли вам поэзия		Всего
	ДА	НЕТ	
Наличие	10 (<i>a</i>)	10 (<i>b</i>)	20
Отсутствие	13 (<i>c</i>)	7 (<i>d</i>)	20
Всего	23	17	40

Коэффициент ассоциации Пирсона:

$$\varphi_{эмп} = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Корреляции в номинальной шкале

Недихотомические шкалы

Пол	Будут учиться	Не определились	Не будут учиться	Всего
Юноши	35	5	85	125
Девушки	22	10	99	131
Всего	57	15	184	256

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_o – наблюдаемая частота в ячейке
 f_e – ожидаемая частота в ячейке

$$f_e = \frac{\text{общая частота по строке} \times \text{общая частота по столбцу}}{\text{количество случаев в выборке}}$$

- Таблицы сопряжённости:

Analyze – Descriptive Statistics – Crosstabs

- Ожидаемые частоты (произведение сумм соответствующей строки и столбца, делённое на общую сумму частот): *Cells – Counts (Expected)*
- Остатки (разность наблюдаемой и ожидаемой частот): *Cells – Residuals (Unstandardized)*
- Процентные доли: *Cells – Percentages*

Корреляции в порядковой шкале

Analyze – Correlate – Bivariate

- Коэффициент ранговой корреляции Спирмена (r_s)

$$r_s = 1 - \frac{6 \sum(D^2)}{n(n^2 - 1)}$$

D – разность рангов в паре переменных,
 n – число случаев

- Коэффициент ранговой корреляции Кендала (τ)

$$\tau = \frac{2S}{n(n - 1)}$$

$$S = P - Q$$

P – суммарное число наблюдений, следующих за текущими наблюдениями с **большим** значением рангов,

Q – суммарное число наблюдений, следующих за текущими наблюдениями с **меньшим** значением рангов

n – число случаев

Корреляции в порядковой шкале

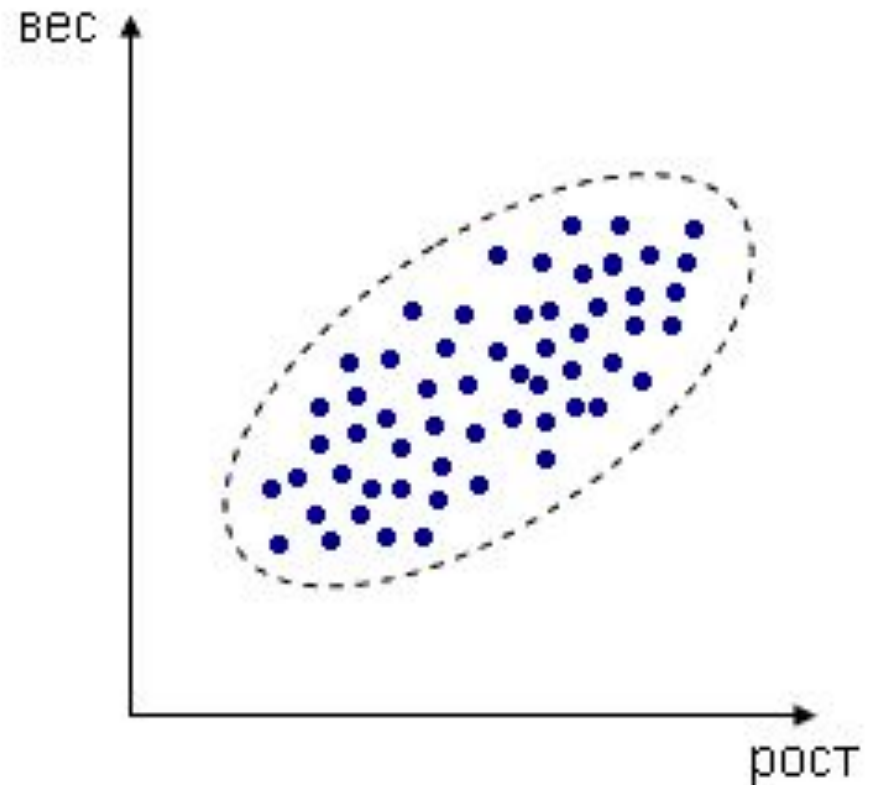
Пример

Проверим социальный стереотип:
красивая женщина – глупая женщина

Переменные:

Красота – ?

Ум – ?



Корреляции в интервальной шкале

- Коэффициент корреляции Пирсона
Analyze – Correlate – Bivariate

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{(n - 1) s_x s_y}$$

Толкование корреляций

- Время, затрачиваемое ребенком на ролевые игры, положительно коррелирует с уровнем произвольности
- Пол ребенка коррелирует с предпочтением точных или гуманитарных предметов
- Удовлетворенность работой отрицательно коррелирует с намерением уволиться

Нет оснований утверждать, что одна переменная является причиной другой!

Условия вывода о причинно-следственной связи

- Изменение независимой переменной происходит до изменения зависимой или одновременно с ним.
- Независимая и зависимая переменные связаны между собой неслучайно.
- Существует разумное объяснение наблюдаемого отношения, а также отсутствует правдоподобное альтернативное объяснение.

Только при выполнении всех условий можно сделать причинно-следственный вывод!

Представление результатов корреляционного анализа (1)

Table 1. Means, SDs, Internal consistency (Cronbach's alpha on the diagonal), and Correlations of the variables used in the study (n=776^a)

Measures	# items	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Age	-	38.1	12.8	-	-0.02	-0.17***	-0.11**	-0.13***	0.30***	0.28***	0.26***	-0.28***	-0.16***	0.26***	0.11**	0.22***
2 Gender ^b	-	1.1	0.2	-	-	-0.30***	-0.05	-0.08*	-0.05	-0.03	-0.02	-0.01	0.08*	-0.03	0.07†	-0.05
3 Section ^c	-	1.7	0.4	-	-	-	0.07†	0.20***	-0.02	-0.07†	-0.09*	0.16***	0.15***	-0.10**	-0.06	0.11**
4 Work excessively	5	10.4	3.2	-	-	-	0.70	0.49***	-0.04	0.04	0.08*	0.52***	0.38***	-0.28***	-0.12***	0.01
5 Work compulsively	5	9.8	3.1	-	-	-	-	0.70	0.21***	0.19***	0.24***	0.39***	0.31***	-0.12**	-0.11**	-0.08*
6 Vigor	3	10.8	4.3	-	-	-	-	-	0.84	0.82***	0.75***	-0.29***	-0.23***	0.41***	0.17***	0.17***
7 Dedication	3	11.5	4.0	-	-	-	-	-	-	0.78	0.81***	-0.27***	-0.17***	0.42***	0.20***	0.24***
8 Absorption	3	10.1	4.1	-	-	-	-	-	-	-	0.81	-0.18***	-0.09*	0.38***	0.13***	0.15***
9 Psychological distress	15	30.2	9.6	-	-	-	-	-	-	-	-	0.93	0.68***	-0.50***	-0.31***	-0.16***
10 Physical complaints	11	19.6	6.5	-	-	-	-	-	-	-	-	-	0.88	-0.36***	-0.21***	-0.09*
11 Job satisfaction	1	2.6	0.8	-	-	-	-	-	-	-	-	-	-	-	0.33***	0.14***
12 Family satisfaction	1	3.0	0.8	-	-	-	-	-	-	-	-	-	-	-	-	0.12***
13 Job performance	1	6.6	2.0	-	-	-	-	-	-	-	-	-	-	-	-	-

† $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

^aThe numbers did not add up to the total number of the participants because of occasional missing data; ^bMen=1, Women=2; ^cIndirect=1, Direct=2.

Представление результатов корреляционного анализа (2)

Table 1. Descriptive statistics and correlations

	1	2	3	4	5	6	7	\bar{X}	SD	α
1. NHS org' support								2.81	0.69	.76
2. Trust org' support	0.64***							2.81	0.83	.86
3. NHS identification	0.26***	0.20***						3.47	0.84	.93
4. Trust identification	0.37***	0.54***	0.55***					3.02	0.82	.94
5. NHS involvement	0.26***	0.22***	0.61***	0.39***				3.80	0.67	.79
6. Trust involvement	0.26***	0.36***	0.42***	0.54***	0.66***			3.63	0.73	.75
7. NHS intent to leave	-0.33***	-0.30***	-0.50***	-0.38***	-0.36***	-0.27***		2.32	1.00	.95
8. Trust intent to leave	-0.39***	-0.58***	-0.28***	-0.57***	-0.22***	-0.31***	0.59***	2.50	1.02	.95

$N = 736$. * $p < .05$, ** $p < .01$, *** $p < .001$.

Представление результатов корреляционного анализа (3)

Table 1
Mean, Standard Deviation, Internal Consistency (Dutch/Japanese Cronbach's Alpha on the Diagonal),
and Zero-Order Correlations of the Study Variables in the Dutch ($N = 7,594^a$; Below the Diagonal)
and Japanese ($N = 3,311^b$; Above the Diagonal) Samples

Variable	Dutch		Japanese		1	2	3	4	5	6	7	8
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>								
1. Working excessively	2.53	0.60	2.31	0.67	.78/.73	.46***	.46***	.29***	.04	-.14***	-.03	-.13***
2. Working compulsively	2.01	0.63	1.94	0.58	.44***	.78/.68	.36***	.25***	.02	.08***	.15***	.09***
3. Exhaustion	1.28	0.87	3.86	1.50	.27***	.43***	.85/.92	.63***	-.05*	-.38***	-.27***	-.30***
4. Cynicism	1.16	0.93	2.32	1.42	.05*	.30***	.53***	.77/.86	-.13***	-.44***	-.42***	-.35***
5. Professional efficacy	4.24	0.80	2.41	1.14	.11***	-.25***	-.34***	-.45***	.81/.85	.47***	.48***	.38***
6. Vigor	3.09	1.14	1.96	1.25	-.01	-.10***	-.51***	-.53***	.65***	.85/.86	.81***	.73***
7. Dedication	3.40	1.93	2.40	1.24	.02*	-.11***	-.37***	-.64***	.69***	.79***	.87/.86	.69***
8. Absorption	3.38	1.73	1.49	1.33	.03*	.01	-.22***	-.37***	.52**	.71**	.69***	.81/.83

a. $N = 1,406$ for the burnout scales.

b. $N = 2,024$ for the burnout scales.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Представление результатов корреляционного анализа (4)

Table 2

Correlations between measures within Times 1 and Time 2.

	1	2	3	4	5	6	7	8	9	10
<i>Perceived culture</i>										
1. Human relations	–	.78	.72	.65	.14	.16	.10	.20	.57	.28
2. Open systems	.83	–	.61	.78	.17	.18	.09	.20	.54	.34
3. Internal process	.67	.63	–	.65	.09	.10	.13	.12	.38	.20
4. Rational goal	.66	.80	.65	–	.15	.14	.01	.14	.40	.23
<i>Preferred culture</i>										
5. Human relations	.13	.16	.10	.16	–	.84	.57	.67	.15	.05
6. Open systems	.15	.16	.06	.12	.83	–	.51	.73	.14	.02
7. Internal process	.10	.11	.02	.01	.63	.59	–	.54	.10	.06
8. Rational goal	.19	.16	.07	.09	.65	.76	.61	–	.27	.10
9. Affective commitment	.58	.55	.38	.41	.10	.15	.12	.23	–	.54
10. Intention to stay	.32	.33	.22	.27	.09	.09	.16	.16	.55	–

Note. Correlations below the diagonal are for Time 1 (Ns range from 673 to 685) and correlations above the diagonal are for Time 2 (Ns range from 620 to 625). Correlations below the diagonal greater than .07 are significant at $p < .05$, and correlations greater than .09 are significant at $p < .01$. Correlations above the diagonal greater than .07 are significant at $p < .05$, and correlations greater than .10 are significant at $p < .01$.