

Оценка тесноты связи между признаками

Курс «Маркетинговые исследования»

Рассматриваемые вопросы

- Задачи маркетинговых исследований, решаемые с помощью оценок тесноты связи между признаками. Подходы к оценке тесноты связи между признаками
- Анализ парных взаимосвязей
- Альтернативная вариация признаков
- Вариация качественных, но не альтернативных признаков
- Вариация количественного и качественного признаков
- Вариация количественных признаков
- Многомерный анализ взаимосвязей
 - Множественный корреляционно-регрессионный анализ
 - Дискриминантный анализ
 - Факторный анализ

Зависимые и независимые переменные

■ Зависимые

переменные

(сегмент рынка, объем продаж или отношение к той или иной торговой марке).

■ Независимые переменные

■ *социально- демографические переменные:*

- социальные;
- этнические;
- демографические;

■ *психографические переменные* (характеризующие жизненный стиль).

■ *переменные, связанные с управлением маркетингом:*

- продукт, его потребительские свойства, дизайн, упаковка, качество;
- цена на продукт;
- система распределения;
- система стимулирования сбыта;

Показатели тесноты связи между признаками.

Результативные показатели	Шкалы измерения признака	Показатели тесноты связи результативных показателей с признаками*		
		альтернативными	качественными	количественными
Альтернативные	наименований	A, k	P, C	Z
Качественные	наименований	P, C	P, C	ω
Количественные	порядковые	Z, ω	R, W	R, W
	интервальные и относительные	Z, ω	ω	η, r

- * A, k - коэффициенты ассоциации и контингенции, соответственно
 P, C - коэффициенты взаимной сопряженности Пирсона и Чупрова
 R, W - ранговые коэффициенты Спирмена и Кендела
- Z, ω - бисерийный и дисерийный коэффициенты
 η - эмпирическое корреляционное отношение,
 r - коэффициент корреляции

Альтернативная вариация признаков

Признак 1 (причина)	Признак 2 (следствие)	
	Первое значение признака 2	Второе значение признака 2
Первое значение признака 1	a	b
Второе значение признака 1	c	d

Коэффициенты тесноты связи рассчитываются по формулам:

Коэффициент ассоциации (A)

$$A = \sqrt{\frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}}$$

Контингенции (k)

$$k = \sqrt{\frac{a \cdot d + b \cdot c}{\sqrt{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)}}}$$

Например, при исследовании эффективности рекламных мероприятий по продвижению шампуней отечественного производства фирмой-производителем шампуня изучался вопрос о связи между полом и восприятием рекламы. В магазине было опрошено 150 покупателей шампуня. Те респонденты, которые купили шампунь под воздействием рекламы, были отнесены к неслучайным покупателям; остальные — к случайным. Можно предположить, что воздействие рекламы на покупателей-женщин сильнее, чем на мужчин. Для измерения этой связи составим таблицу сопряженности между признаками

Пол респондентов	Покупатели		Итого
	Случайные	Неслучайные	
Мужской	36 (a)	17 (b)	53
Женский	42 (c)	87 (d)	129
Итого	78	104	182

$$A := \frac{36 \cdot 87 - 42 \cdot 17}{36 \cdot 87 + 42 \cdot 17} = 0.629$$

$$k := \frac{36 \cdot 87 - 42 \cdot 17}{\sqrt{53 \cdot 129 \cdot 78 \cdot 104}} = 0.325$$

При анализе работы фирмы, оказывающей услуги как компаниям, так и индивидуальным заказчикам, возникла необходимость оценить связь между типом клиента и качеством обслуживания

Тип клиента	Уровень обслуживания		Итого
	Устраивает	Не устраивает	
Компании	150	18	168
Индивидуальные заказчики	252	113	365
Итого	402	131	533

Вариация качественных, но не альтернативных признаков (коэффициент взаимной сопряженности Пирсона)

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$\chi^2 = n \cdot \left[\sum_{ij} \frac{(n_{ij})^2}{\sum_i n_i \cdot \sum_j n_j} - 1 \right]$$

- Пример расчета коэффициента взаимной сопряженности Пирсона для исследования взаимосвязи между предпочитаемой маркой растворимого кофе и типом потребителя. Опросу подвергались люди, которые пьют растворимый кофе.
- Группировка респондентов производилась следующим образом:
 - если человек пьет кофе несколько раз в день (более 2 чашек в день), то он относился к группе *активных* потребителей;
 - если человек пьет кофе 1-2 чашки каждый день (или почти каждый день), то он относился к группе *умеренных* потребителей;
 - если респондент выпивает кофе реже, чем 2-3 раза в неделю, то он включался в группу *слабых* потребителей,
 - если реже, чем 1 раз в неделю, то к группе *случайных* потребителей.

Предпочитаемая марка кофе	Активные потребители	Умеренные потребители	Слабые потребители	Случайные потребители	Всего
Ambassador	43	53	26	14	136
Jacobs	107	94	34	18	253
Nescafe	93	102	37	32	264
Carte Noire	134	98	58	28	318
Maxwell House	124	142	11	16	293
King cup	14	19	78	93	204
Другая марка	62	72	86	61	281
Нет предпочтений	28	46	65	78	217
Всего	605	626	395	340	1966

$$\chi^2 = 1966 \cdot \left(\frac{43^2}{605 \cdot 136} + \frac{53^2}{626 \cdot 136} + \frac{26^2}{395 \cdot 136} + \dots + \frac{78^2}{340 \cdot 217} - 1 \right) = 495,537$$

$$C := \sqrt{\frac{495.537}{1966 + 495.537}} = 0.449$$

Форма собственности	Оценка уровня жизни				Итого
	Вполне удовлетворен	Скорее удовлетворен	Скорее не удовлетворен	Совсем не удовлетворен	
Государственная	31	35	35	35	136
Муниципальная	17	13	14	9	53
Смешанная российская	4	2	1	1	8
Частная	8	5	4	3	20
Итого	60	55	54	48	217 ₁₁

Вариация количественного и качественного признаков (дисерийные коэффициенты)

$$\omega_i = \frac{Y_i - Y}{\sigma_y \cdot \sqrt{\frac{\sum f}{\sum f_i} - 1}}$$

Y - среднее значение признака;

Y_i - среднее значение признака в i -ой группе;

σ_y - среднее квадратическое отклонение;

f - частота.

Тип покупателей	Уровень доходов на одного члена семьи (руб.)					Всего
	До 1000 (750)	1000-1500 (1250)	1500-2000 (1750)	2000-3000 (2500)	Свыше 3000 (3500)	
Случайные покупатели (f 1)	21	18	16	7	3	65
Лояльные покупатели (f 2)	23	31	39	39	43	175
Всего (f)	44	49	55	46	46	240

$$Y_1 = \frac{21 \cdot 750 + 18 \cdot 1250 + 16 \cdot 1750 + 7 \cdot 2500 + 3 \cdot 3500}{65} = 1450$$

$$Y_2 = \frac{23 \cdot 750 + 31 \cdot 1250 + 39 \cdot 1750 + 39 \cdot 2500 + 43 \cdot 3500}{175} = 2127.143$$

$$Y := \frac{44 \cdot 750 + 49 \cdot 1250 + 55 \cdot 1750 + 46 \cdot 2500 + 46 \cdot 3500}{240} = 1943.75$$

$$\sigma_y := \sqrt{\frac{\sum_i (Y_i - Y)^2 \cdot f}{\sum_i f}}$$

$$\sqrt{\frac{(750 - 1943.75)^2 \cdot 44 + (1250 - 1943.75)^2 \cdot 49 + (1750 - 1943.75)^2 \cdot 55 + (2500 - 1943.75)^2 \cdot 46 + (3500 - 1943.75)^2 \cdot 46}{240}} = 944.26$$

$$\omega_1 = \frac{1450 - 1943.75}{944.26 \cdot \sqrt{\frac{240}{65} - 1}} = -0.319$$

$$\omega_2 = \frac{2127.143 - 1943.75}{944.26 \cdot \sqrt{\frac{240}{175} - 1}} = 0.319$$

Оценка качества услуг	Надежность банка (в баллах)				Итого
	До 3	3-3.5	3.5-4	Свыше 4	
Отличное	21	25	32	47	125
Хорошее	18	14	23	25	80
Удовлет.	30	35	20	8	93
Неудовл.	25	16	9	4	54
Итого	94	90	84	84	352

Вариация количественных признаков (коэффициенты корреляции и регрессии)

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

Влияние рекламной кампании на объем продаж

№ магазина	Динамика объема продаж (Y), %	Относительные затраты на рекламу (X), %				
1	112,3	116,7				
2	97,7	95,2				
3	106,4	103,4				
4	95,9	85,3				
...						
19	94,8	97,9				
20	88,3	93,1				
Σ	1958,87	1991,5				
В среднем	97,9435	99,575				

Рассчитаем коэффициенты корреляции и регрессии

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

$$Y = a + b \cdot X,$$

$$\begin{cases} \sum y = n \cdot a + b \sum x \\ \sum xy = a \sum x + b \sum x^2 \end{cases}$$

Влияние рекламной кампании на объем продаж

№ магазина	Динамика объема продаж (Y), %	Относительные затраты на рекламу (X), %	Y^2	X^2	$Y \cdot X$	Y_x
1	112,3	116,7	12611,29	13618,89	13105,41	112,143
2	97,7	95,2	9545,29	9063,04	9301,04	94,3159
3	106,4	103,4	11320,96	10691,56	11001,76	101,1151
4	95,9	85,3	9196,81	7276,09	8180,27	86,10716
...						
19	94,8	97,9	8987,04	9584,41	9280,92	96,55465
20	88,3	93,1	7796,89	8667,61	8220,73	92,57465
Σ	1958,87	1991,5	194430,9349	201434,35	197650,38	1958,87
В среднем	97,9435	99,575	-	-	9882,519	-

$$\begin{cases} 1958,87 = 20 a + 1991,5 b \\ 197650,38 = 1991,5 a + 195896,59 b \end{cases}$$

$$\text{Отсюда, } a = 15,37934; b = 0,82917;$$
$$Y = 15,37934 + 0,82917X$$

Уравнение регрессии позволяет определить, как изменится результативный показатель при изменении факторного, т.е. при увеличении относительного уровня затрат на рекламу на 1 % объем продаж увеличится на 0,829 %

$$\sigma_x = \sqrt{\frac{(116.7 - 99.575)^2 + (95.2 - 99.575)^2 + \dots + \dots + (93.1 - 99.575)^2}{20}} = 12.83650$$

$$\sigma_y = \sqrt{\frac{(112.3 - 97.9435)^2 + (97.7 - 97.9435)^2 + \dots + \dots + (88.3 - 97.9435)^2}{20}} = 11.63559$$

Коэффициент корреляции между изучаемыми признаками равен:

$$r := \frac{9882.519 - 97.9435 \cdot 99.575}{12.83650 \cdot 11.63559} = 0.869$$

Проверка существенности уравнения регрессии

Вариация признака	Степени свободы	Сумма квадратов отклонений	Среднее значение (гр. 3 : гр. 2)	F-статистика
Объясненная	1	$\sum (Y_x - Y)^2$	$\sum (Y_x - Y)^2 = S_1$	$F_{\text{расч.}} = S_1 / S_2$
Необъясненная	$N - 2$	$\sum (Y_i - Y_x)^2$	$\sum (Y_i - Y_x)^2 : (n - 2) = S_2$	
Общая	$N - 1$	$\sum (Y_i - Y)^2$		

Вариация признака	Степени свободы df	Сумма квадратов отклонений (Sums of Squares)	Среднее значение (Mean Squares)	F - статистика	Уровень значимости (p-level)
Объясненная (Regress.)	1	1942,534	1942,534	55,5168	,000000
Необъясненная (Residual)	18	629,814	34,990		
Общая (Total)	19	2572,351			

На долю объясненной вариации приходится 75,516 % $(1942,534:2572,351 \cdot 100 \%)$;
 Необъясненная вариация составляет 24,484 % $(629,814: 2572,351 \cdot 100 \%)$
 $100 \% = 75,516 \% + 24,484 \%$

-
- **POP_CHNG X1 Population change (1960-1970) –
Изменение численности населения**
- **N_EMPLD X2 No. of persons employed in agriculture –
Численность занятых в с/х**
- **PT_POOR X3 Percent of families below poverty level –
процент семей ниже уровня бедности**
- **TAX_RATE X4 Residential and farm property tax rate –
налоговая ставка на собственность (жилую и
фермерскую)**
- **PT_PHONE X5 Percent residences with telephones –
процент резидентов с телефоном**
- **PT_RURAL X6 Percent rural population – процент
сельского населения**
- **AGE X7 Median age - медиана возрастного состава
населения**

Итоги регрессии для зависимой переменной: PT_POOR (Poverty) R= ,88215345 R2= ,77819471 Скоррект. R2= ,72033245 F(6,23)=13,449 p<,00000 Станд. ошибка оценки: 3,3986						
N=30	БЕТА	Ст.Ош. БЕТА	B	Ст.Ош. B	t(23)	p-знач.
Св. член			31,26604	13,26511	2,35701	0,027309
POP_CHNG	-0,630788	0,129413	-0,39234	0,08049	-4,87421	0,000064
N_EMPLD	0,238314	0,140987	0,00075	0,00044	1,69033	0,104476
TAX_RATE	0,038799	0,100611	1,23012	3,18985	0,38563	0,703311
PT_PHONE	-0,129627	0,203294	-0,08325	0,13056	-0,63763	0,530012
PT_RURAL	0,618746	0,231173	0,16554	0,06185	2,67655	0,013476
AGE	-0,188205	0,114652	-0,41926	0,25541	-1,64153	0,114292

Итоги регрессии для зависимой переменной: PT_POOR (Poverty)						
R= ,88134020 R2= ,77676055 Скоррект. R2= ,73025233						
F(5,24)=16,702 p<,00000 Станд. ошибка оценки: 3,3378						
N=30	БЕТА	Ст.Ош. БЕТА	B	Ст.Ош. B	t(24)	p-знач.
Св.член			32,30393	12,75677	2,53230	0,018287
POP_CHNG	-0,623453	0,125717	-0,38778	0,07820	-4,95917	0,000046
N_EMPLD	0,244869	0,137454	0,00077	0,00043	1,78146	0,087498
PT_PHONE	-0,135855	0,199025	-0,08725	0,12781	-0,68260	0,501396
PT_RURAL	0,619107	0,227034	0,16563	0,06074	2,72694	0,011756
AGE	-0,187156	0,112568	-0,41693	0,25077	-1,66260	0,109400

$$Y=25,32-0,42X1+0,001X2+0,2X6-0,48X7$$

Итоги регрессии для зависимой переменной: PT_POOR (Poverty) R= ,87887800 R2= ,77242653 Скоррект. R2= ,73601478 F(4,25)=21,214 p<,000000 Станд. ошибка оценки: 3,3019						
N=30	БЕТА	Ст.Ош. БЕТА	В	Ст.Ош. В	t(25)	p-знач.
Св.член			25,32965	7,556230	3,35215	0,002553
POP_CHNG	-0,677589	0,096498	-0,42145	0,060021	-7,02179	0,000000
N_EMPLD	0,272765	0,129829	0,00086	0,000409	2,10095	0,045893
PT_RURAL	0,746716	0,127440	0,19977	0,034095	5,85935	0,000004
AGE	-0,214441	0,104103	-0,47771	0,231910	-2,05990	0,049963

Дисперсионный анализ; ЗП: PT_POOR (Poverty)					
Эффект	Сумма квадр.	сс	Средн. квадр.	F	p-знач.
Регресс.	925,156	4	231,2890	21,21366	0,000000
Остатки	272,571	25	10,9028		
Итого	1197,727				

Переменная	Предск.значения для (Poverty) перемен.: PT_POOR		
	В-Вес	Значение	В-Вес * знач.
POP_CHNG	-0,421454	30,000	-12,6436
N_EMPLD	0,000860	5000,000	4,2998
PT_RURAL	0,199772	100,000	19,9772
AGE	-0,477710	40,000	-19,1084
Св.член			25,3296
Предсказанные			17,8546
-95,0%ИС			10,2608
+95,0%ИС			25,4484

Если привлечь 30% населения в регион для развития с/х, чтобы численность занятых была на уровне 5000, населения относится к сельскому со средним возрастом 40 лет, то процент бедных семей будет находится в интервале от 10,26 % до 25,45%, (в среднем 17%)