

Регрессионные модели с переменной структурой

Модели регрессии с фиктивными переменными сдвига

ФУНКЦИЯ СПРОСА

$$y = a + bx + \varepsilon$$

y = количество потребляемого кофе

x = цена

Модели регрессии с фиктивными переменными сдвига

Мужчины

$$y_1 = a_1 + b_1 \cdot x_1 + \varepsilon_1$$

Женщины

$$y_2 = a_2 + b_2 \cdot x_2 + \varepsilon_2$$

$$\bar{y}_1$$

$$\bar{y}_2$$

различны

но сила влияния x на y может быть одинаковой, то есть

$$b \approx b_1 \approx b_2$$

Модели регрессии с фиктивными переменными сдвига

$$y = a + bx + \delta \cdot z + \varepsilon$$

фиктивная
переменная

$$z = \begin{cases} 1 & \text{— мужской пол,} \\ 0 & \text{— женский пол.} \end{cases}$$

Модели регрессии с фиктивными переменными сдвига

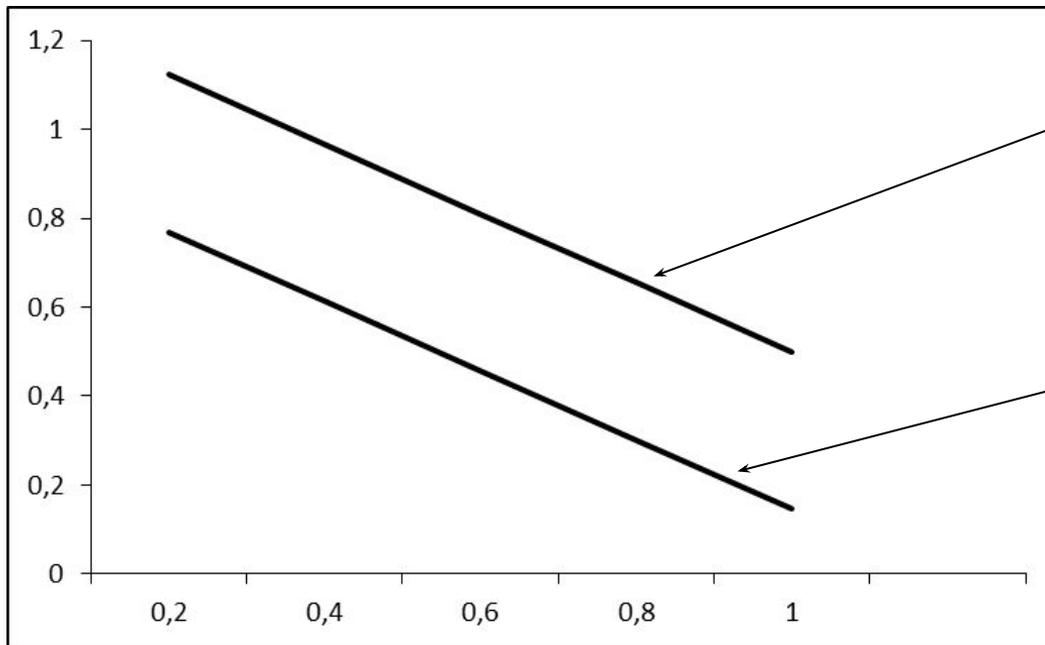
Мужчины $\hat{y} = (a + \delta) + bx$

Женщины $\hat{y} = a + bx$

Различие потребления вызваны
различием свободных членов уравнения
регрессии

Модели регрессии с фиктивными переменными сдвига

$$y = a + b \cdot x + \delta \cdot z + \varepsilon$$



$$\hat{y} = a + b \cdot x + \delta \cdot z$$

$$\hat{y} = a + b \cdot x$$

Модели регрессии с фиктивными переменными сдвига

$$z_1 = \begin{cases} 1 & \text{— проживание в северном регионе,} \\ 0 & \text{— в остальных регионах;} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{— проживание в южном регионе,} \\ 0 & \text{— в остальных регионах.} \end{cases}$$

$$\hat{y} = a + bx + \delta \cdot z + \delta_1 \cdot z_1 + \delta_2 \cdot z_2$$

N	Потреб	Цена	Пол	Северн.	Южн.	N	Потреб	Цена	Пол	Северн	Южн.
	л. (кг)	(тыс. руб)		регион	регион		л. (кг)	(тыс. руб)		регион	регион
	Y	X	Z	Z_1	Z_2		Y	X	Z	Z_1	Z_2
1	0,2	1	1	0	0	16	0,6	0,6	0	1	0
2	0,4	1	0	0	0	17	0,6	0,5	0	1	0
3	0,4	0,8	1	0	0	18	0,65	0,5	1	1	0
4	0,6	0,8	0	0	0	19	0,6	0,3	1	1	0
5	0,6	0,6	1	0	0	20	0,7	0,3	0	1	0
6	0,8	0,6	0	0	0	21	0,5	1	1	0	1
7	0,75	0,5	0	0	0	22	0,6	1	0	0	1
8	0,9	0,5	1	0	0	23	0,7	0,8	1	0	1
9	0,9	0,3	1	0	0	24	0,9	0,8	0	0	1
10	1,1	0,3	0	0	0	25	0,9	0,6	1	0	1
11	0,2	1	1	1	0	26	1,1	0,6	0	0	1
12	0,45	1	0	1	0	27	1	0,5	0	0	1
13	0,45	0,8	1	1	0	28	1,2	0,5	1	0	1
14	0,6	0,8	0	1	0	29	1,2	0,3	1	0	1

Модели регрессии с фиктивными переменными сдвига

$$\hat{y} = 1,26 - 0,84x - 0,11z - 0,13z_1 + 0,29z_2$$

$$R^2 = 0,88$$

$$F = 47,951 > F_{табл} = 2,76$$

	Значения коэффициентов	Стан. ошибка	t-стат.	P-Значение
a	1,26	0,07	19,26	1,64E-16
b	-0,84	0,08	-10,30	1,77E-10
δ	-0,11	0,04	-2,88	0,008
δ_1	-0,13	0,05	-2,70	0,012
δ_2	0,29	0,05	5,91	3,63E-0,6

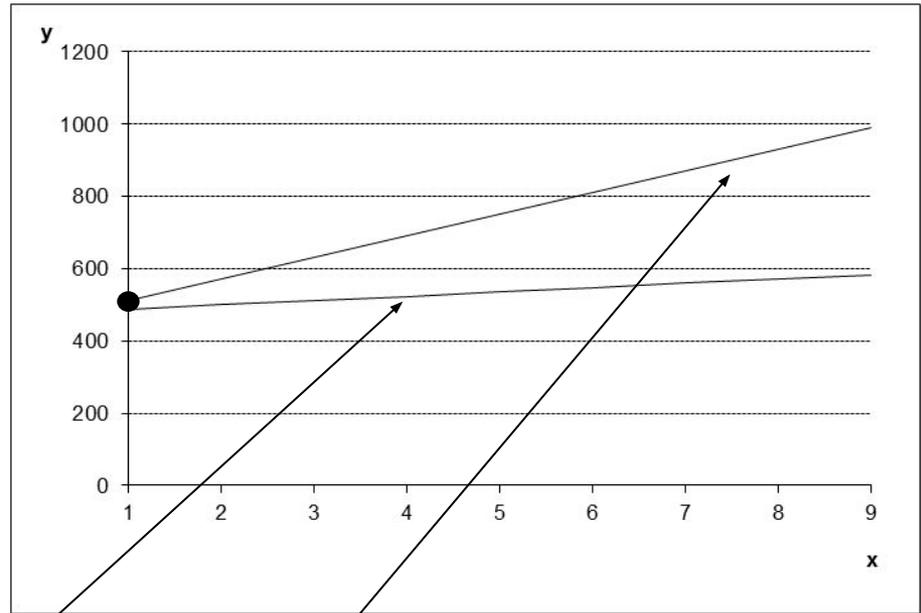
Тип категории	Уравнение
Женщины (северные регионы)	$\hat{Y} = 1,13 - 0,84x$
Мужчины (северные регионы)	$\hat{Y} = 1,02 - 0,84x$
Женщины (центральные регионы)	$\hat{Y} = 1,26 - 0,84x$
Мужчины (центральные регионы)	$\hat{Y} = 1,15 - 0,84x$
Женщины (южные регионы)	$\hat{Y} = 1,55 - 0,84x$
Мужчины (южные регионы)	$\hat{Y} = 1,44 - 0,84x$

Модели регрессии с фиктивными переменными наклона

$$\hat{y} = a + b_1 \cdot x, \text{ при } z = 0$$

$$\hat{y} = a + b_2 \cdot x, \text{ при } z = 1$$

$$b_1 \neq b_2$$



$$\hat{y} = a + b \cdot x + \delta \cdot x \cdot z$$

$$\text{при } z = 0 \quad \hat{y} = a + b \cdot x,$$

$$\text{при } z = 1 \quad \hat{y} = a + (b + \delta) \cdot x$$

Станция назначения	Расстояние х, км	Тип вагона z	Стоимость проезда у, руб.		Станция назначения	Расстояние х, км	Тип вагона z	Стоимость проезда у, руб.
Мга	42	0	98,1		Мга	42	1	313,4
Волхов	114	0	124,9		Волхов	114	1	384,2
Будогощь	140	0	133,4		Будогощь	140	1	404,4
Приозерск	141	0	137,4		Приозерск	141	1	414,6
Луга	147	0	137,4		Луга	147	1	414,6
М. Вишера	162	0	145,5		М. Вишера	162	1	435,5
Тихвин	192	0	157,6		Тихвин	192	1	465,8
Пикалево	230	0	169,7		Пикалево	230	1	495,3
Подпорожье	272	0	189,8		Подпорожье	272	1	547,3
Псков	284	0	189,8		Псков	284	1	547,3
Бологое	319	0	227,9		Бологое	319	1	643,3
Петрозаводск	394	0	229,2		Петрозаводск	394	1	648,4
Осташков	431	0	249,3		Осташков	431	1	697,4
Вел. Луки	588	0	308,9		Вел. Луки	588	1	850,5
Вологда	592	0	463,4		Вологда	592	1	1275,2
Ярославль	691	0	349,6		Ярославль	691	1	950,7

$$\hat{y} = a + b \cdot x + \delta \cdot x \cdot z$$

$$\text{при } z = 0 \quad \hat{y} = a + b \cdot x,$$

$$\text{при } z = 1 \quad \hat{y} = a + (b + \delta) \cdot x$$

Стоимость проезда в разных типах вагонов различна, но зависит от расстояния

$z = 0$ – плацкартный вагон

$z = 1$ – купейный вагон

Оценка по МНК

$$\hat{y} = 112,21 + 0,44 \cdot x + 1,15 \cdot x \cdot z$$

Общий вид модели регрессии с фиктивными переменными

$$z = \begin{cases} 0 & \text{— мужчины,} \\ 1 & \text{— женщины.} \end{cases} \quad (z \cdot x)$$

$$\hat{y} = a + b \cdot x + \delta \cdot x \cdot z + \gamma \cdot z$$

$$\text{при } z = 0 \quad \hat{y} = a + b \cdot x,$$

$$\text{при } z = 1 \quad \hat{y} = (a + \gamma) + (b + \delta) \cdot x$$

N набл.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y	9	6	10	7	12	8,5	13	9	15	9	18	9,5	20	12	22	15	25	16
x	2	2	3	3	4	4	5	5	7	7	8	8	10	10	12	12	15	15
z	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$z \cdot x$	0	2	0	3	0	4	0	5	0	7	0	8	0	10	0	12	0	15

$$\hat{y} = 6,69 + 1,27 \cdot x - 2,08 \cdot z - 0,5 \cdot (z \cdot x)$$

$$R^2 = 0,98$$

$$F = 283,92 > F_{табл} = 3,33$$

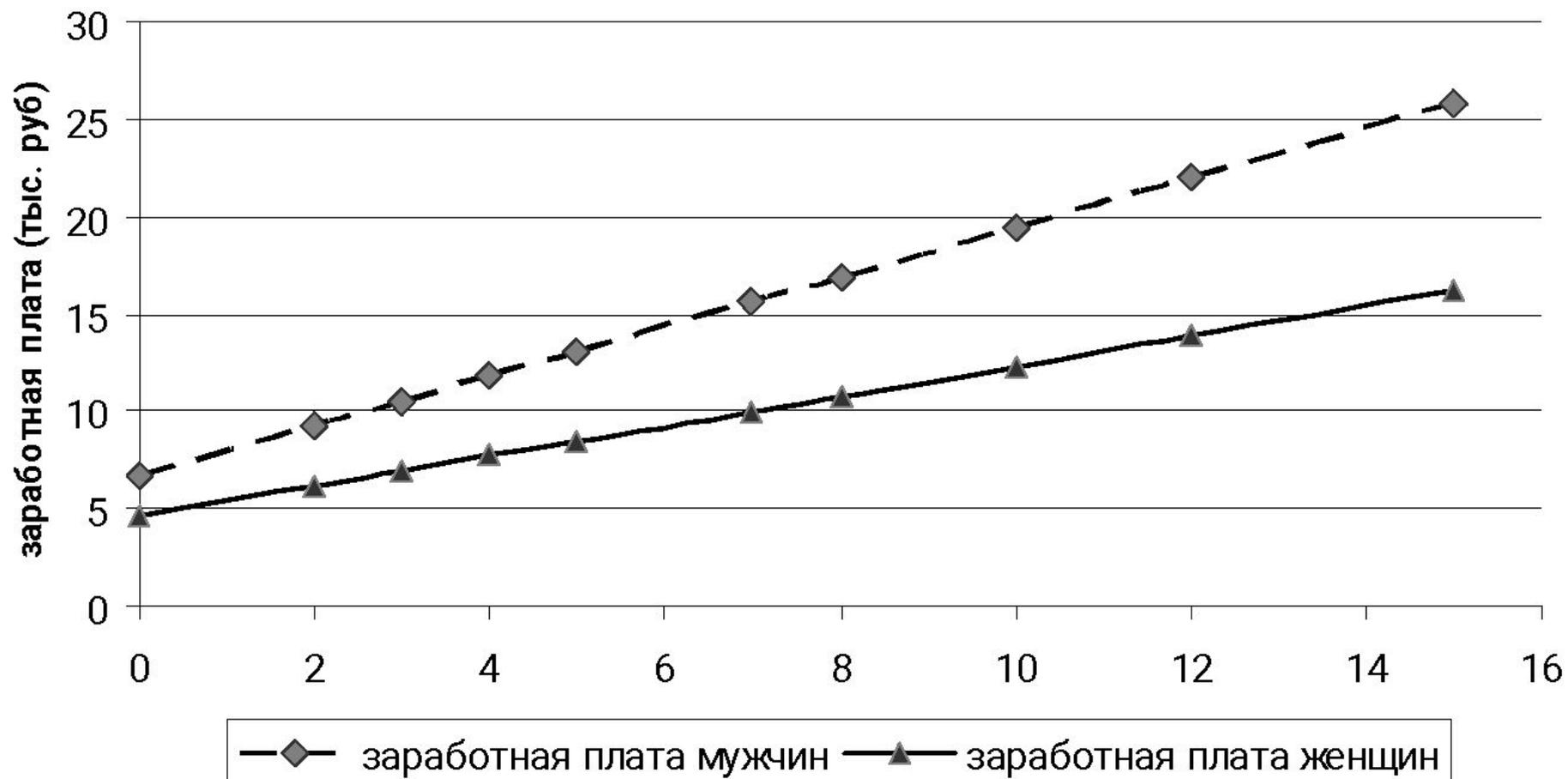
	Значения коэффициентов	Ст.ошибка	t-ст.	P-Значение
<i>a</i>	6,69	0,51	13,01	3,3E-09
<i>b</i>	1,27	0,06	20,76	6,48E-12
<i>δ</i>	-2,08	0,73	-2,86	0,01262
<i>γ</i>	-0,50	0,09	-5,83	4,36E-05

Мужчины

$$\hat{y} = 6,69 + 1,27 \cdot x$$

Женщины

$$\hat{y} = 4,61 + 0,77 \cdot x$$



Тест ЧОУ

Пусть имеется две подвыборки: одна объемом n_1 , а другая объемом n_2 .

1. По каждой подвыборке строятся линейные регрессионные модели с m переменными:

$$y_1 = b_{1_0} + \sum_{j=1}^m b_{1_j} \cdot x_j + \varepsilon_1 \quad y_2 = b_{2_0} + \sum_{j=1}^m b_{2_j} \cdot x_j + \varepsilon_2$$

Для каждого уравнения рассчитывается остаточная сумма квадратов

$$SS_1 \quad SS_2$$

2. Строится линейная регрессия по объединенной выборке

$$y = b_0 + \sum_{j=1}^m b_j \cdot x_j + \varepsilon$$

Вычисляется ее сумма квадратов остатков

$$SS$$

3. Формулируется нулевая гипотеза

$$H_0 : b_{1_j} = b_{2_j}, \quad j = \overline{0, m}, \quad \leftrightarrow \quad SS = SS_1 + SS_2$$

4. Для проверки гипотезы вычисляется фактическое значение F -статистики по формуле

$$F = \frac{(SS - (SS_1 + SS_2))}{(SS_1 + SS_2)} \cdot \frac{n - 2 \cdot m - 2}{m + 1}$$

$$\hat{y}_1 = 6,689 + 1,269 \cdot x \quad R^2 = 0,99; \quad SS_1 = 2,94$$

$$\hat{y}_2 = 4,61 + 0,765 \cdot x \quad R^2 = 0,95; \quad SS_2 = 5,019$$

$$\hat{y} = 5,649 + 1,018 \cdot x \quad R^2 = 0,639; \quad SS = 177,52$$

$$F = \frac{(177,52 - 2,94 - 5,019) \cdot 14}{(2,94 + 5,019) \cdot 2} = 80,4$$

$$F_{табл}(\alpha = 0,05; 2, 14) = 3,74$$