

Данные анализируемых переменных

Analyzed variable data

Год	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1994	0,4	-9.6	22,2	61	81	291	2143	44	18	23	6
1995	0,5	-7.6	22,2	55	67	264	1890	43	20	22	5
1996	0,6	-8.8	22,3	56	67	270	1916	35	17	19	5
1997	0,6	-8.1	22,4	52	70	265	1905	38	17	24	7
1998	0,6	-8.7	22,4	61	77	315	2203	42	13	20	11
1999	0,5	-8.9	22,3	69	99	367	2645	54	16	36	17
2000	0,5	-9.1	22,5	65	88	316	2305	48	18	20	12
2001	0,8	-8.7	22,4	66	83	347	2402	41	20	22	12
2002	0,9	-7.6	22,4	69	93	355	2543	48	29	24	13
2003	1,1	-8.0	22,4	69	93	360	2557	49	29	23	11
2004	1,3	-9.8	22,5	70	97	367	2627	54	30	26	13
2005	1,6	-7.7	22,7	74	100	368	2672	58	49	31	12
2006	2,1	-8.8	22,6	71	97	345	2544	51	45	30	12
2007	1,8	-7.3	22,7	68	90	306	2314	46	47	24	10
2008	1,8	-7.2	22,7	71	92	319	2399	48	50	25	10
2009	2,2	-7.7	22,7	72	96	317	2430	48	49	26	11
2010	3	-7.7	22,7	77	105	341	2626	60	64	28	11
2011	2,8	-7.3	22,8	77	103	337	2593	67	62	28	11
2012	2,8	-7.8	22,8	77	103	337	2594	62	60	29	11
2013	3,2	-7.4	22,8	79	106	334	2618	60	60	29	11

Источники: 1. Заболеваемость взрослого населения Республики Саха (Якутия): Статистические материалы. ГУ ЯРМИАЦ МЗ РС (Я). Якутск. 2003-2017 гг.

2. Государственный комитет РФ по статистике, Государственный комитет Республики Саха (Якутия) по статистике. Уровень жизни населения Республики Саха (Якутия) за 1992 – 1997 гг., Якутск, 1999. – С 43.

3. Государственный комитет РФ по статистике, Государственный комитет Республики Саха (Якутия) по статистике. Уровень жизни населения Республики Саха (Якутия) за 1990, 1995 – 2000 гг., Якутск, 2001. – С 55.

4. Федеральная служба государственной статистики (Росстат). Потребление продуктов питания в домашних хозяйствах, 2008-2014 гг.
http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1140095125312 доступ 30 ноября 2014 г.

5. Федеральная служба государственной статистики. Территориальный орган Федеральной службы государственной статистики по РС(Я). Социальное положение и уровень жизни населения Республики Саха (Якутия) / Стат. сб., Якутск, 2014. – С. 396-397.

Уравнение множественной регрессии

Проверка гипотез относительно коэффициентов уравнения регрессии (проверка значимости параметров множественного уравнения регрессии).

Число $v = n - m - 1$ называется числом степеней свободы. Считается, что при оценивании

множественной линейной регрессии для обеспечения статистической надежности требуется, чтобы число наблюдений, по крайней мере, в 3 раза превосходило число оцениваемых параметров.

y – впервые в жизни выявленная заболеваемость сахарным диабетом 2 типа (СД2) на 1000 населения;

x_1 – средняя температура жилищ в отопительный сезон в °С;

x_2 – калорийность рациона, в ккал;

x_3 – потребление овощей и бахчевых, в г.;

x_4 – потребление фруктов и ягод, в г.

x_5 – потребление растительного масла и других жиров, в г.

1) t -статистика

$$T_{\text{табл}}(n-m-1; \alpha/2) = (16; 0.025) = 2.473$$

$$t_i = \frac{b_i}{S_{b_i}}$$

$$t_0 = \frac{0.999}{0.307} = 3.255 > 2.473$$

Статистическая значимость коэффициента регрессии b_0 подтверждается.

$$t_1 = \frac{-0.0444}{0.0789} = 0.563 < 2.473$$

Статистическая значимость коэффициента регрессии b_1 не подтверждается.

$$t_2 = \frac{-0.000421}{0.00128} = 0.33 < 2.473$$

Статистическая значимость коэффициента регрессии b_2 не подтверждается.

$$t_3 = \frac{0.00629}{0.026} = 0.242 < 2.473$$

Статистическая значимость коэффициента регрессии b_3 не подтверждается.

$$t_4 = \frac{0.0516}{0.0101} = 5.101 > 2.473$$

Статистическая значимость коэффициента регрессии b_4 подтверждается.

$$t_5 = \frac{0.0296}{0.0835} = 0.354 < 2.473$$

Статистическая значимость коэффициента регрессии b_5 не подтверждается.

Доверительный интервал для коэффициентов уравнения регрессии.

Определим доверительные интервалы коэффициентов регрессии, которые с надежностью 95% будут следующими:

$$(b_i - t_{\text{табл}} * S_{b_i}; b_i + t_{\text{табл}} * S_{b_i})$$

$$b_0: (0.999 - 2.473 * 0.307; 0.999 + 2.473 * 0.307) = (0.24; 1.759)$$

$$b_1: (-0.0444 - 2.473 * 0.0789; -0.0444 + 2.473 * 0.0789) = (-0.24; 0.151)$$

Поскольку найденный интервал включает 0, то коэффициент b_1 не значим.

$$b_2: (-0.000421 - 2.473 * 0.00128; -0.000421 + 2.473 * 0.00128) = (-0.00358; 0.00273)$$

Поскольку найденный интервал включает 0, то коэффициент b_2 не значим.

$$b_3: (0.00629 - 2.473 * 0.026; 0.00629 + 2.473 * 0.026) = (-0.0579; 0.0705)$$

Поскольку найденный интервал включает 0, то коэффициент b_3 не значим.

$$b_4: (0.0516 - 2.473 * 0.0101; 0.0516 + 2.473 * 0.0101) = (0.0266; 0.0767)$$

$$b_5: (0.0296 - 2.473 * 0.0835; 0.0296 + 2.473 * 0.0835) = (-0.177; 0.236)$$

Поскольку найденный интервал включает 0, то коэффициент b_5 не значим.

6. Проверка общего качества уравнения множественной регрессии.

F-статистика. Критерий Фишера.

$$R^2 = 1 - \frac{s_e^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{3.017}{19.59} = 0.846$$

Проверим гипотезу об общей значимости - гипотезу об одновременном равенстве нулю всех коэффициентов регрессии при объясняющих переменных:

$$H_0: R^2 = 0; \beta_1 = \beta_2 = \dots = \beta_m = 0.$$

$$H_1: R^2 \neq 0.$$

Проверка этой гипотезы осуществляется с помощью F-статистики распределения Фишера (правосторонняя проверка).

Если $F < F_{\text{кр}} = F_{\alpha; n-m-1}$, то нет оснований для отклонения гипотезы H_0 .

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} = \frac{0.846}{1-0.846} \cdot \frac{22-5-1}{5} = 17.575$$

Табличное значение при степенях свободы $k_1 = 5$ и $k_2 = n-m-1 = 22 - 5 - 1 = 16$, $F_{кр}(5;16) = 2.85$. Поскольку фактическое значение $F > F_{кр}$, то коэффициент детерминации статистически значим и уравнение регрессии статистически надежно (т.е. коэффициенты b_i совместно значимы).

Выводы.

В результате расчетов было получено уравнение множественной регрессии: $Y = 0.9993 - 0.04444X_1 - 0.000421X_2 + 0.00629X_3 + 0.05163X_4 + 0.02956X_5$. Статистическая значимость уравнения проверена с помощью коэффициента детерминации и критерия Фишера. Установлено, что в исследуемой ситуации 84.6% общей вариабельности Y объясняется изменением факторов X_j . Установлено также, что параметры b_1, b_2, b_3, b_5 модели статистически не значимы.

Решение было получено и оформлено с помощью сервиса:

[Уравнение множественной регрессии \(https://math.semestr.ru/regress/corel.php\)](https://math.semestr.ru/regress/corel.php) доступ

06.06.2022 г.

Результаты измерения средней температуры (°C) жилого частного дома газовым отоплением с 15 сентября 2021 по 15 мая 2022 г. (г. Якутск)

дни	сент	окт	ноя	дек	янв	фев	мар	апр	май
1		25,5	23,5	25,0	24,0	25,0	25,0	22,0	22,0
2		25,0	25,0	24,0	24,0	24,5	25,0	23,5	25,0
3		25,0	27,0	25,0	24,0	22,5	23,0	24,0	24,0
4		25,0	25,0	23,5	24,0	24,5	24,0	24,0	21,5
5		25,0	25,0	25,0	24,0	22,0	24,5	23,0	24,0
6		26,0	26,0	25,0	24,0	25,0	23,0	22,5	23,5
7		24,0	25,0	24,5	24,0	23,5	25,0	22,0	22,5
8		24,0	23,5	23,5	25,0	24,5	25,0	23,0	25,0
9		24,0	25,0	24,5	25,0	23,0	22,5	22,0	25,0
10		25,0	25,0	25,0	23,0	24,5	22,0	23,5	24,0
11		23,5	24,5	25,0	24,0	25,0	23,5	21,5	25,0
12		23,0	24,0	24,0	25,0	23,0	25,0	23,0	24,0
13		22,5	23,5	24,0	24,0	23,5	24,0	23,0	24,0
14		24,0	25,0	26,0	24,5	24,0	23,0	23,5	25,0
15	25,5	23,5	25,0	25,0	24,0	23,0	23,5	25,0	
16	27,0	23,0	24,0	26,0	24,0	24,0	23,5	24,0	
17	25,0	22,5	25,0	25,0	24,5	25,0	23,5	22,0	
18	25,0	24,0	25,0	25,0	24,0	23,5	23,0	25,0	
19	26,0	24,0	25,5	24,5	24,0	24,0	23,5	25,0	
20	26,0	24,0	23,5	24,5	24,0	24,0	22,0	26,0	
21	24,0	25,0	24,0	24,5	24,0	23,5	23,0	25,0	
22	24,0	25,0	24,0	25,0	25,0	24,0	23,0	22,0	
23	23,0	23,0	25,0	24,0	25,0	23,5	23,5	24,0	
24	25,0	25,0	23,0	26,0	25,0	25,0	23,0	24,0	
25	26,0	25,0	25,0	24,0	25,0	25,0	23,0	22,0	
26	26,5	22,5	24,0	23,0	24,0	23,0	23,0	23,5	
27	23,5	23,0	25,0	23,5	24,0	23,0	23,0	24,0	
28	24,0	23,0	25,0	23,5	24,5	25,0	25,0	25,0	

29	24,0	25,0	23,5	23,0	25,0	25,0	23,5
30	25,0	22,0	23,0	24,5	25,0	21,5	22,0
31		24,0		24,0	24,0	22,5	