

符号逆転現象

γ-GTAP	飲酒量(回/月)	喫煙の有無	ギャンブル嗜好
38	8	0	0
42	12	0	0
102	11	1	1
70	10	0	0
110	26	1	1
58	21	0	0
82	13	1	0
70	24	0	0
62	10	1	1
58	15	1	1
70	9	1	1
38	7	0	0
46	9	0	0
94	15	1	1
122	30	1	1
34	4	0	0
38	6	0	0
62	17	0	0
54	10	0	0
90	23	0	0

この2列の相関が
0.90と良すぎる

重回帰式

γ - GPT

= 2.14 × 飲酒量 + 17.8 × 喫煙 - 1.36 × ギャンブル + 26.4

重回帰分析結果

Excelの「データ分析」の「回帰」を用いて、
重回帰分析実行

回帰統計	
重相関 R	0.884
重決定 R2	0.781
補正 R2	0.741
標準誤差	13.154
観測数	20

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	3	9899.71	3299.90	19.07	0.00
残差	16	2768.29	173.02		
合計	19	12668			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	26.4	6.62	3.99	0.00	12.37	40.44	12.37	40.44
飲酒量(回/月)	2.14	0.43	4.96	0.00	1.23	3.05	1.23	3.05
喫煙の有無	27.8	13.69	2.03	0.06	-1.25	56.80	-1.25	56.80
ギャンブル嗜好	-1.36	14.15	-0.10	0.92	-31.35	28.63	-31.35	28.63

相関係数

	γ-GTAP	飲酒量(回/月)	喫煙の有無	ギャンブル嗜好
γ-GTAP	1.00			
飲酒量(回/月)	0.73	1.00		
喫煙の有無	0.67	0.25	1.00	
ギャンブル嗜好	0.62	0.27	0.90	1.00

正負の符号が逆転

飲酒量、喫煙及びギャンブルの相関係数は全て正

y_i	No.		
	1	2	3
y -GTAP	飲酒量(回/月)	喫煙の有無	ギャンブル嗜好
1	38	8	0
2	42	12	0
3	102	11	1
4	70	10	0
5	110	26	1
6	58	21	0
7	82	13	1
8	70	24	0
9	62	10	1
10	58	15	1
11	70	9	1
12	38	7	0
13	46	9	0
14	94	15	1
15	122	30	1
16	34	4	0
17	38	6	0
18	62	17	0
19	54	10	0
20	90	23	0

Excelの「データ分析」の「回帰」を用いて、
モデルNo.1~7まで、総渡りで重回帰分析実行

モデルNo.1

回帰統計				
重相関 R	0.728			
重決定 R2	0.529			
修正 R2	0.503			
標準誤差	18.201			
観測数	20			

分散分析表					
	自由度	変動	分散	測された分散	有意 F
回帰	1	6705.054	6705.054	20.240	0.000
残差	18	5962.946	331.275		
合計	19	12668			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	30.784	9.020	3.413	0.003	11.834	49.735	11.834	49.735
飲酒量(回/月)	2.587	0.575	4.499	0.000	1.379	3.795	1.379	3.795

残差出力			
観測値	予測値: y -GTAP	残差	残差平方和
1	51.479	-13.479	5962.95
2	61.826	-19.826	
3	59.240	42.760	
4	56.653	13.347	
5	98.042	11.958	
6	85.108	-27.108	
7	64.413	17.587	
8	92.868	-22.868	
9	56.653	5.347	
10	69.587	-11.587	
11	54.066	15.934	
12	48.892	-10.892	
13	54.066	-8.066	
14	69.587	24.413	
15	108.389	13.611	
16	41.132	-7.132	
17	46.305	-8.305	
18	74.760	-12.760	
19	56.653	-2.653	
20	90.281	-0.281	

$$y_{1i} = a_1x_{1i} + b_1 = 2.59x_{1i} + 30.8$$

モデルNo.	組み合わせNo.	飲酒量	喫煙	ギャンブル	切片	残差平方和	AIC
1	1	2.59			30.8	5962.9	156.7
2	2		34.2		53.3	7064.7	160.1
3	3			32.8	55.5	7788.7	162.1
4	12	2.14	26.6		25.5	2769.9	143.4
5	13	2.15		24.2	28.4	3480.3	147.9
6	23		28.7	6.29	53.3	7030.1	162.0
7	123	2.14	27.8	1.36	26.4	2768.3	145.4

No.4適合度大

モデル

- $y = a_1x_1 + b_1$
- $y = a_2x_2 + b_2$
- $y = a_3x_3 + b_3$
- $y = a_1x_1 + a_2x_2 + b_4$
- $y = a_1x_1 + a_3x_3 + b_5$
- $y = a_2x_2 + a_3x_3 + b_6$
- $y = a_1x_1 + a_2x_2 + a_3x_3 + b_7$

ここでは
 x_1 : 飲酒量
 x_2 : 喫煙
 x_3 : ギャンブル
 a_n : 係数
 b_n : 切片

S_e : 残渣の平方和

$$S_e = \sum_{i=1}^{20} (y_{j,i} - y_i)^2$$

$j: 1 \sim 7$

$$y_{7i} = a_1x_{1i} + a_2x_{2i} + a_3x_{3i} + b_7$$

$$= 2.14x_{1i} + 27.8x_{2i} - 1.36x_{3i} + 26.4$$

モデルNo.7

回帰統計				
重相関 R	0.884			
重決定 R2	0.781			
修正 R2	0.741			
標準誤差	13.154			
観測数	20			

分散分析表					
	自由度	変動	分散	測された分散	有意 F
回帰	3	9899.71	3299.90	19.07	0.00
残差	16	2768.29	173.02		
合計	19	12668			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	26.41	6.82	3.89	0.00	12.37	40.44	12.37	40.44
飲酒量(回/月)	2.14	0.43	4.96	0.00	1.23	3.05	1.23	3.05
喫煙の有無	27.78	13.69	2.03	0.06	-1.25	56.80	-1.25	56.80
ギャンブル嗜好	-1.36	14.15	-0.10	0.92	-31.35	28.63	-31.35	28.63

残差出力			
観測値	予測値: y -GTAP	残差	残差平方和
1	43.53	-5.53	2768.29
2	52.09	-10.09	
3	76.36	25.64	
4	47.81	22.19	
5	108.46	1.54	
6	71.34	-13.34	
7	82.00	0.00	
8	77.76	-7.76	
9	74.22	-12.22	
10	84.92	-26.92	
11	72.08	-2.08	
12	41.39	-3.39	
13	45.67	0.33	
14	84.92	9.08	
15	117.02	4.98	
16	34.97	-0.97	
17	39.25	-1.25	
18	62.78	-0.78	
19	47.81	6.19	
20	75.62	14.38	

AIC: 赤池情報量基準(Akaike's Information Criterion)

- モデルの当てはまり度を表す統計量
- 小さい数値ほどモデルに合っている

$$AIC = n \left(\ln \left(2\pi \frac{S_e}{n} \right) + 1 \right) + 2(p + 2)$$

n : サンプルサイズ
 S_e : 残渣の平方和
 π : 円周率
 p : 説明変数の数