

⑫～⑮統計量比較図

⑪に出力された、分析結果のうち、標準偏回帰係数、P 値、および重相関係数に関わるものをグラフ化して提示します。⑮では、グラフの表示域などの設定を行えます。いずれのグラフも、コピーしたりグラフ編集で加工して利用できます。

**例題 13：** 集団健診で得られた 572 例の個人特性と臨床検査値（12 項目）は下表のとおりである。これから、HDL-C と年齢、性別、喫煙度、飲酒度、BMI との関連の強さを分析せよ。また、同じ条件で全 12 検査について同時分析し、標準偏回帰係数の比較表を作成せよ。



このデータは [SampleData] フォルダの [DB 型 健診 12 検査結果.SFD6] で、読み込むとデータシートは次のようになります。なお、12 検査のうち、中性脂肪、AST、ALT、GGT は、対数変換 (べき乗値 p=0.0) の指定がされています。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
	年齢	性別	飲酒度	喫煙度	BMI	TP	ALB	尿酸	血糖	HbA1c	中性脂肪	総コレステロール	HDL-C	AST	ALT	LDH	GGT	
行1	39	1	0	1		22.4	6.8	4.3	3.0	94	4.9	37	143	62	15	11	335	9
行2	36	0	0	1		23.1	7.1	4.5	6.0	101	5.3	58	187	47	21	19	293	32
行3	66	0	0	1		19.4	7.5	4.4	3.8	106	5.8	47	252	110	31	17	353	28
行4	68	0	0	2		26.6	7.1	4.3	7.7	112	4.9	51	169	38	17	14	318	17
行5	45	0	0	2		22.5	6.9	4.4	6.7	89	5.3	149	208	45	20	26	273	35
行6	71	0	0	2		24.0	6.9	3.8	4.9	87	5.6	41	217	100	28	20	489	28
行7	62	0	0	2		18.4	7.0	4.2	4.5	105	5.4	56	208	61	21	20	290	16
行8	71	0	0	3		25.3	6.7	3.9	7.4	95	5.1	161	155	42	35	45	339	142
行9	70	0	0	3		25.1	7.3	4.5	4.5	104	4.8	90	179	37	24	18	408	21
行10	59	0	0	3		23.5	6.8	4.1	6.5	93	5.3	63	239	52	22	17	450	13
行11	62	0	0	4		22.2	7.0	4.6	4.4	88	5.4	77	261	62	25	22	508	19
行12	55	0	0	4		22.1	7.6	4.8	5.3	87	5.5	77	226	51	20	15	389	17
行13	49	0	0	4		23.5	7.5	4.6	7.4	84	5.2	125	250	34	17	18	312	18
行14	63	0	0	5		22.7	6.9	4.1	5.5	111	6.1	86	177	41	18	15	318	13
行15	74	0	0	6		26.8	6.7	4.2	7.8	100	5.5	85	200	52	22	15	410	21



計算出力例 1-1

目的変数に HDL-C を、説明変数に年齢、性別、飲酒度、喫煙度、BMI を選んで実行します。重回帰分析表の  $std\beta$  は標準偏回帰係数と呼ばれ、目的変数と各説明変数との純粋な (他の説明変数の影響を除外した) 相関係数 (偏相関係数) を表します。多重共線性や交互作用項をの設定がなければ、その値は通常  $-1.0 \leq std\beta \leq 1.0$  の範囲に入ります。

目的変数: HDL-C 有効データ数=572 [第1頁: 群1]

次数	変数名	$\beta$	SE( $\beta$ )	$std\beta$	t-val	df	P-val
0		84.43	5.786				
1	年齢	0.1099	0.05645	0.0762	1.9471	566	0.05202
2	性別	4.826	1.535	0.1649	3.1442	566	0.00175
3	飲酒度	1.613	0.5225	0.1415	3.0861	566	0.00213
4	喫煙度	-1.277	0.3970	-0.1609	-3.2161	566	0.00137
5	BMI	-1.487	0.2020	-0.2865	-7.3606	566	0.00000

< 回帰の適合度指標 >

重相関係数R=0.3975 決定係数R<sup>2</sup>=0.1580

決定係数調整後adjR<sup>2</sup>=0.1506



計算出力例 1-2 (詳細出力) (上と同じデータ: 詳細出力で実行)

stdβ の信頼区間、多重共線性の判定指標となる分散拡大係数 (VIF) とトレランス (Tol) が追加表示され、また多重共線性の度合いを表す条件番号 (CN) や回帰の適合度のより詳細なデータが出力されます。VIF は 10 以上、Tol はその逆数で 0.1 以下、CN は 30 以上が問題とされますが、詳細は、数理解説の項を参照してください。

目的変数: HDL-C 有効データ数=572 [第1頁: 群1]

次数	変数名	β	SE(β)	stdβ	stdβ 95%CI	t-val	df	P-val	VIF	Tol
0		84.43	5.786							
1	年齢	0.1099	0.05645	0.0762	-0.001~0.153	1.9471	566	0.05202	1.029	0.972
2	性別	4.826	1.535	0.1649	0.062~0.268	3.1442	566	0.00175	1.848	0.541
3	飲酒度	1.613	0.5225	0.1415	0.051~0.232	3.0861	566	0.00213	1.414	0.707
4	喫煙度	-1.277	0.3970	-0.1609	-0.259~-0.063	-3.2161	566	0.00137	1.682	0.595
5	BMI	-1.487	0.2020	-0.2865	-0.363~-0.210	-7.3606	566	0.00000	1.019	0.982

VIF=分散拡大係数 Tol=トレランス

< 回帰の適合度指標 >

重相関係数R=0.3975 決定係数R<sup>2</sup>=0.1580 [F=21.2437 (df1=5, df2=566) P=2.220E-016]

決定係数調整後adjR<sup>2</sup>=0.1506

赤池の情報量規準AIC=4602.5051

多重共線性判定指数(条件番号CN)=2.3732

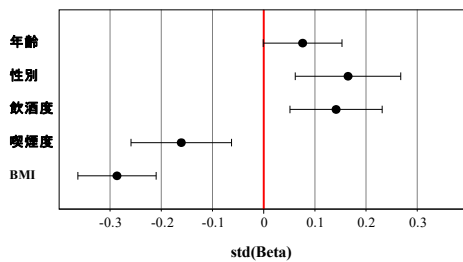
残差変動SE=102040.3866 (総変動ST=121189.8112 回帰変動SR =19149.4246)

回帰の標準誤差 Sy=13.4270



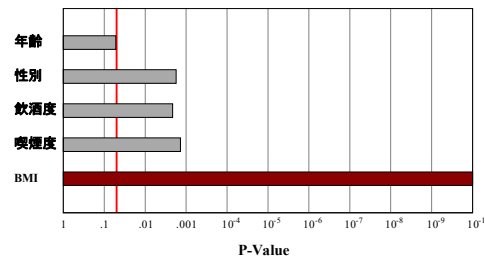
統計量の比較分析

● stdβ 比較図



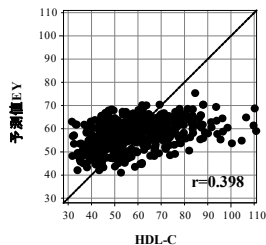
stdβ の表示域は操作パネル右上の「設定」タブ (下記) で調整できます。

● P 値比較図



P 値の表示域は操作パネル右上の「設定」タブ (下記) で変更できます。

● 予測値の適合度



観察値 Y と予測値 Y の相関図で、図中の相関係数 r が重相関係数 R に相当します。

● 統計量作図設定

The settings panel for statistical charts. It includes checkboxes for 'stdβ (rp) 表示域自動設定' and 'P値 表示域自動設定'. Below these are input fields for 'rp 下限値' (-1.00), 'rp 上限値' (1.00), 'P 下限値' (0.0001), and 'P 上限値' (1). At the bottom, there is a setting for 'stdβ 信頼区間確率: 0~100' set to 95.0.



## 計算出力例 2

次に、対数変換指定済み(べき乗=0.0)の中性脂肪を目的変数に実行すると、結果は次のようになります。

目的変数: 中性脂肪 有効データ数=572 【第1頁: 群1】

次数	変数名	$\beta$	SE( $\beta$ )	std $\beta$	std $\beta$ 95%CI	t-val	df	P-val	VIF	Tol
0		3.155	0.1868							
1	年齢	-8.368E-5	1.822E-3	-0.0018	-0.078~0.075	-0.0459	566	0.96339	1.029	0.972
2	性別	0.03088	0.04955	0.0326	-0.070~0.135	0.6233	566	0.53334	1.848	0.541
3	飲酒度	-6.257E-3	0.01687	-0.0170	-0.107~0.073	-0.3710	566	0.71079	1.414	0.707
4	喫煙度	0.07052	0.01282	0.2743	0.176~0.372	5.5022	566	0.00000	1.682	0.595
5	BMI	0.05395	6.519E-3	0.3210	0.245~0.397	8.2749	566	0.00000	1.019	0.982

VIF=分散拡大係数 Tol=トレランス

## &lt; 回帰の適合度指標 &gt;

重相関係数 $R=0.4045$  決定係数 $R^2=0.1636$  [F=22.1404 (df1=5, df2=566) P=1.000E-024]

決定係数調整後adj $R^2=0.1562$

赤池の情報量規準AIC=674.7665

多重共線性判定指数(条件番号CN)=2.3732

残差変動SE=106.3194 (総変動ST=127.1140 回帰変動SR =20.7946)

回帰の標準誤差 Sy=0.4334



## 説明変数固定実行の出力例

初期設定の目的変数は HDL-C ですが、残る 11 の変数も含めて、全 12 変数を順に目的変数として自動実行するには、操作パネル右下の⑧のタブ「説明変数固定実行」を利用します。その出力は、std $\beta$  と P 値の 2 通りから選びます。両方を選んで実行すると結果は次のようになります。

## 重回帰分析

std $\beta$ 一覧

目的変数	n	R	年齢	性別	飲酒度	喫煙度	BMI
TP	572	0.2844	0.0222	0.0617	-0.0505	-0.1615	0.1633
ALB	572	0.2579	-0.2284	-0.0748	-0.0069	-0.0418	0.0896
尿酸	572	0.5605	-0.0859	-0.5128	-0.0059	-0.0020	0.1609
血糖	572	0.2143	0.1128	-0.0825	0.0688	-0.0226	0.1410
HbA1c	572	0.2795	0.2243	0.0955	-0.0715	0.1192	0.0765
中性脂肪	572	0.4045	-0.0018	0.0326	-0.0170	0.2743	0.3210
総コレステロール	572	0.2883	0.1589	0.1562	-0.0546	0.0277	0.1564
HDL-C	572	0.3975	0.0762	0.1649	0.1415	-0.1609	-0.2865
AST	572	0.3389	0.2333	-0.1572	0.0708	-0.0196	0.1669
ALT	572	0.4897	-0.0488	-0.2993	-0.0298	0.0251	0.3539
LDH	572	0.4884	0.3709	0.2084	0.0116	-0.1100	0.1120
GGT	572	0.6335	-0.0597	-0.2995	0.1962	0.1919	0.2296

## P値一覧

目的変数	n	R	年齢	性別	飲酒度	喫煙度	BMI
TP	572	0.2844	0.58671	0.26059	0.29196	0.00210	0.00007
ALB	572	0.2579	0.00000	0.17584	0.88666	0.42826	0.02919
尿酸	572	0.5605	0.01535	0.00000	0.88630	0.96415	0.00001
血糖	572	0.2143	0.00694	0.13973	0.15942	0.67160	0.00072
HbA1c	572	0.2795	0.00000	0.08228	0.13691	0.02311	0.06072
中性脂肪	572	0.4045	0.96339	0.53334	0.71079	0.00000	0.00000
総コレステロール	572	0.2883	0.00011	0.00446	0.25474	0.59623	0.00013
HDL-C	572	0.3975	0.05202	0.00175	0.00213	0.00137	0.00000
AST	572	0.3389	0.00000	0.00359	0.13250	0.70234	0.00003
ALT	572	0.4897	0.19011	0.00000	0.49418	0.59807	0.00000
LDH	572	0.4884	0.00000	0.00003	0.79058	0.02108	0.00260
GGT	572	0.6335	0.07113	0.00000	0.00000	0.00001	0.00000

例題のデータを、HDL-C を目的変数にして同じ条件で R を使って実行すると次の結果が得られます。



### R による分析結果 (1)

```
> dat = read.csv("DB型 健診12検査結果.csv", header=T)
> res = lm(data=dat, HDL.C~年齢+性別+飲酒度+喫煙度+BMI)
> summary(res)

Call:
lm(formula = HDL.C ~ 年齢 + 性別 + 飲酒度 + 喫煙度 + BMI, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-30.625  -9.222  -1.960   7.853  51.599

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  84.42537    5.78583   14.592 < 2e-16 ***
年齢           0.10992    0.05645    1.947  0.05202 .
性別           4.82621    1.53495    3.144  0.00175 **
飲酒度        1.61257    0.52253    3.086  0.00213 **
喫煙度       -1.27696    0.39705   -3.216  0.00137 **
BMI          -1.48659    0.20196   -7.361 6.48e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.43 on 566 degrees of freedom
Multiple R-squared:  0.158,    Adjusted R-squared:  0.1506
F-statistic: 21.24 on 5 and 566 DF,  p-value: < 2.2e-16
```

また、対数変換指定のある中性脂肪を目的変数にし、同じ条件で R を使って実行すると次の結果が得られます。



### R による分析結果 (2)

```
> dat = read.csv("DB型 健診12検査結果.csv", header=T)
> res = lm(data=dat, log(中性脂肪)~年齢+性別+飲酒度+喫煙度+BMI)
> summary(res)

Call:
lm(formula = log(中性脂肪) ~ 年齢 + 性別 + 飲酒度 + 喫煙度 +
    BMI, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-1.2588 -0.2883 -0.0265  0.3093  1.2528

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.155e+00  1.868e-01  16.896 < 2e-16 ***
年齢       -8.368e-05  1.822e-03  -0.046  0.963
性別        3.088e-02  4.955e-02  0.623  0.533
飲酒度     -6.257e-03  1.687e-02  -0.371  0.711
喫煙度      7.052e-02  1.282e-02  5.502 5.69e-08 ***
BMI         5.395e-02  6.519e-03  8.275 9.26e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4334 on 566 degrees of freedom
Multiple R-squared:  0.1636,    Adjusted R-squared:  0.1562
F-statistic: 22.14 on 5 and 566 DF,  p-value: < 2.2e-16
```