

統計解析の基礎

目的変数(賃料)と説明変数(駅距離、築年)および誤差項の関係式によりdataを発生させ、そのdataを解析することにより当初関係式を重回帰分析により求めるシミュレーションを行う。

関係式 賃料 = 1000 - 0.5 × (駅距離) - 10 × (築年) + 誤差項

駅距離 = 乱数発生(平均500, 標準偏差160の正規分布)

築年 = 乱数発生(平均15, 標準偏差5の正規分布)

誤差項 = 乱数発生(平均0, 標準偏差100の正規分布)

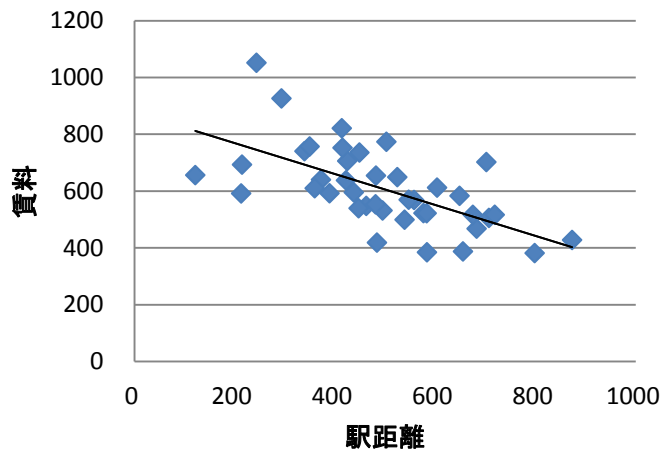
基本統計量

賃料		駅距離		築年		誤差項	
平均	608.5363	平均	494.4874	平均	14.76513	平均	3.431305
標準誤差	22.57561	標準誤差	26.03064	標準誤差	0.788691	標準誤差	16.59973
中央値 (X)	591.9302	中央値 (X)	482.6341	中央値 (X)	14.68663	中央値 (X)	7.799315
最頻値 (モ)	#N/A	最頻値 (モ)	#N/A	最頻値 (モ)	#N/A	最頻値 (モ)	#N/A
標準偏差	142.7807	標準偏差	164.6322	標準偏差	4.988122	標準偏差	104.9859
分散	20386.32	分散	27103.76	分散	24.88136	分散	11022.05
尖度	1.375549	尖度	-0.04777	尖度	0.914074	尖度	0.723025
歪度	0.840776	歪度	0.006872	歪度	0.597768	歪度	0.389053
範囲	670.1675	範囲	751.3881	範囲	24.70356	範囲	502.7923
最小	381.616	最小	121.3184	最小	4.990238	最小	-206.544
最大	1051.783	最大	872.7066	最大	29.6938	最大	296.2479
合計	24341.45	合計	19779.5	合計	590.6051	合計	137.2522
標本数	40	標本数	40	標本数	40	標本数	40

誤差項σ=100

$$y = -0.5445x + 877.79$$

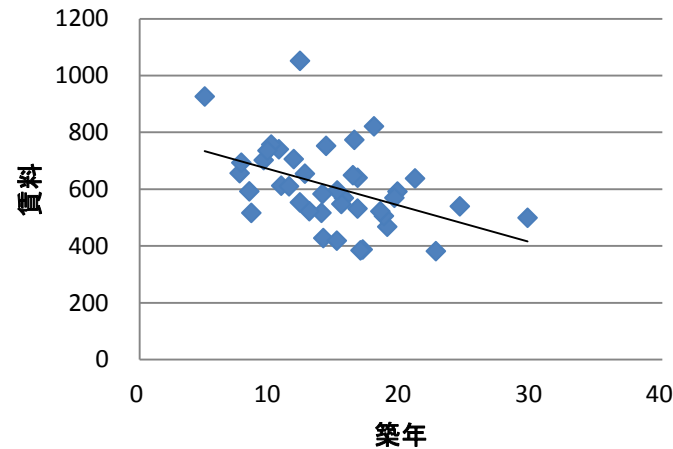
$$R^2 = 0.3942$$



σ=100

$$y = -12.871x + 798.58$$

$$R^2 = 0.2022$$



賃料	駅距離	築年	誤差項
532	495	17	-54
468	682	19	-2
382	798	23	7
516	675	9	-61
638	422	21	59
592	213	20	-105
612	603	11	22
640	372	17	-7
505	707	19	46
654	482	13	22
569	558	16	4
517	719	14	15
656	121	8	-207
592	389	8	-129
522	576	13	-59
706	424	12	36
702	702	10	148
610	359	11	-96
752	415	14	103
821	413	18	207
428	873	14	5
548	462	15	-67
419	484	15	-189
553	481	12	-84
595	437	15	-35
649	524	16	74
740	339	11	17
540	447	25	8
926	293	5	123
757	349	10	32
499	539	30	65
570	547	20	38
774	502	16	189
1052	243	12	296
693	215	8	-122
387	655	17	-114
583	648	14	47
736	449	10	58
522	583	18	-2
384	583	17	-155

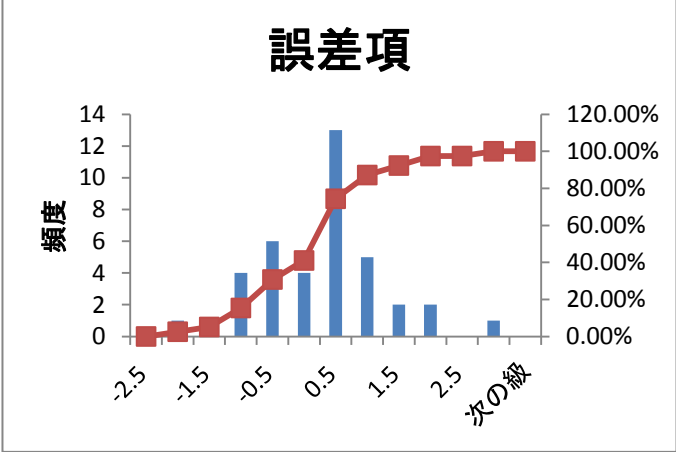
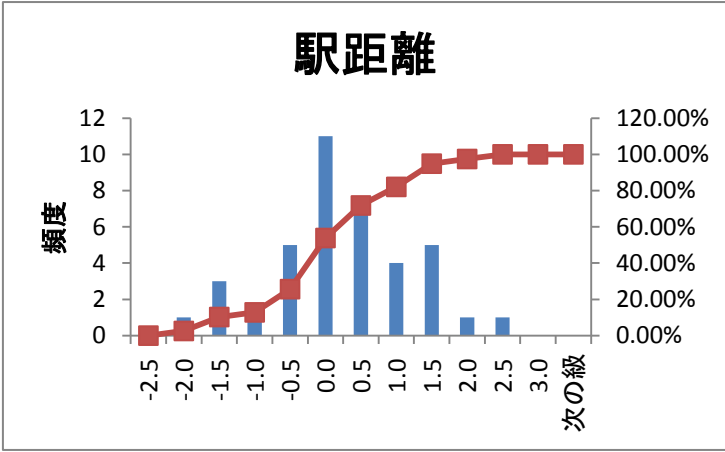
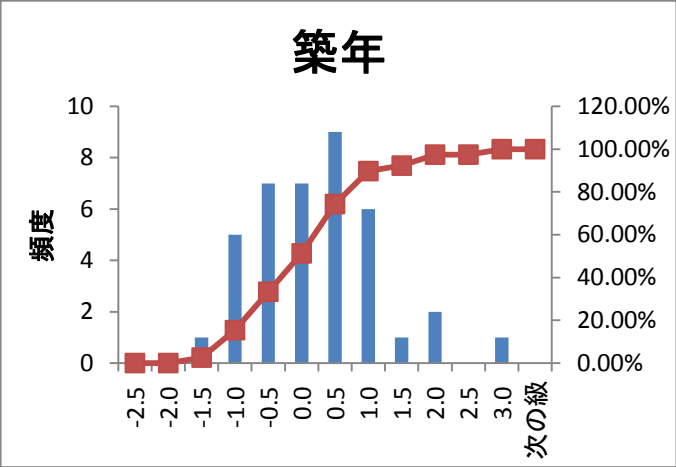
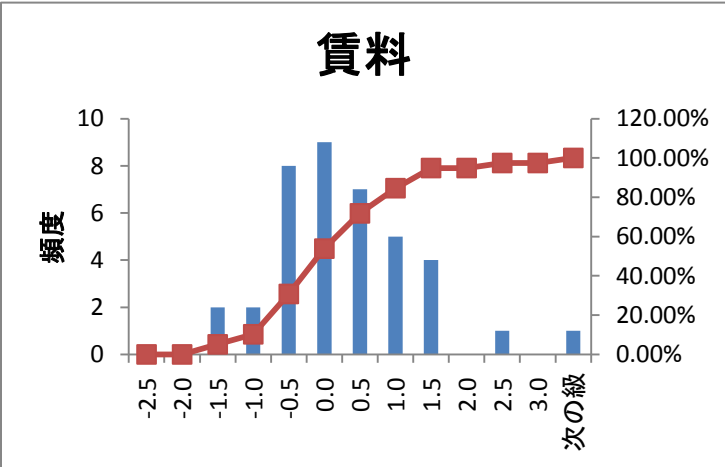
標準化することにより、偏回帰係数の大きさを目的変数の寄与度を判定することができる。

$\sigma = 100$

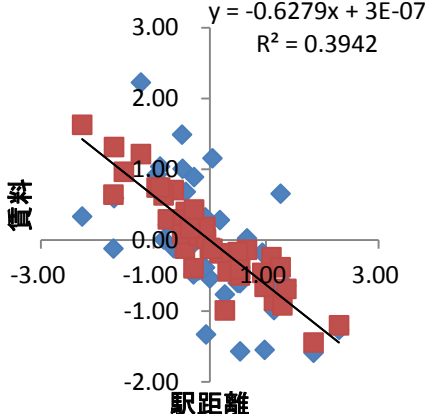
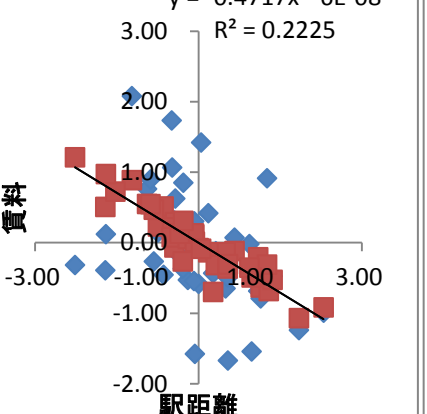
賃料	駅距離	築年	誤差項
-0.54	0.00	0.38	-0.55
-0.99	1.14	0.84	-0.05
-1.59	1.84	1.59	0.04
-0.65	1.10	-1.25	-0.61
0.20	-0.44	1.27	0.53
-0.12	-1.71	1.00	-1.03
0.03	0.66	-0.79	0.18
0.22	-0.74	0.39	-0.10
-0.72	1.29	0.79	0.40
0.32	-0.08	-0.42	0.18
-0.28	0.38	0.17	0.00
-0.64	1.36	-0.17	0.11
0.33	-2.27	-1.42	-2.00
-0.11	-0.64	-1.28	-1.26
-0.60	0.50	-0.36	-0.60
0.68	-0.43	-0.59	0.31
0.66	1.26	-1.06	1.38
0.01	-0.82	-0.67	-0.95
1.01	-0.48	-0.10	0.95
1.49	-0.49	0.64	1.94
-1.27	2.30	-0.14	0.01
-0.43	-0.20	0.14	-0.67
-1.33	-0.07	0.07	-1.83
-0.39	-0.08	-0.50	-0.83
-0.09	-0.35	0.07	-0.36
0.28	0.18	0.31	0.68
0.92	-0.94	-0.82	0.13
-0.48	-0.29	1.95	0.05
2.23	-1.22	-1.96	1.14
1.04	-0.88	-0.94	0.28
-0.77	0.27	2.99	0.59
-0.27	0.32	0.95	0.33
1.16	0.05	0.34	1.77
3.10	-1.53	-0.50	2.79
0.59	-1.70	-1.40	-1.19
-1.55	0.98	0.46	-1.12
-0.18	0.93	-0.15	0.42
0.89	-0.28	-1.00	0.52
-0.61	0.54	0.73	-0.06
-1.57	0.54	0.43	-1.51

全ての要素の
平均 $\div 0$
標準偏差 $\div 1$
となった。

賃料		駅距離		築年		誤差項	
平均	7.65E-08	平均	2.92E-07	平均	-6.1E-07	平均	-3.2E-09
標準誤差	0.158114	標準誤差	0.158114	標準誤差	0.158114	標準誤差	0.158114
中央値 (メ)	-0.1163	中央値 (メ)	-0.072	中央値 (メ)	-0.01574	中央値 (メ)	0.041606
最頻値 (モ)	#N/A	最頻値 (モ)	#N/A	最頻値 (モ)	#N/A	最頻値 (モ)	#N/A
標準偏差	1	標準偏差	1	標準偏差	1	標準偏差	1
分散	1	分散	1	分散	1	分散	1.000001
尖度	1.375549	尖度	-0.04777	尖度	0.914074	尖度	0.723025
歪度	0.840776	歪度	0.006872	歪度	0.597768	歪度	0.389053
範囲	4.693684	範囲	4.564041	範囲	4.952477	範囲	4.789141
最小	-1.58929	最小	-2.26668	最小	-1.95963	最小	-2.00004
最大	3.104391	最大	2.297358	最大	2.992843	最大	2.789104
合計	3.06E-06	合計	1.17E-05	合計	-2.4E-05	合計	-1.3E-07
標本数	40	標本数	40	標本数	40	標本数	40



	誤差項 $\sigma = 0$						誤差項 $\sigma = 50$					
相関	賃料 駅距離 築年						賃料 駅距離 築年					
	賃料	1					賃料	1				
	駅距離	-0.89957	1				駅距離	-0.81117	1			
	築年	-0.69315	0.3087	1			築年	-0.60394	0.3087	1		
重相関	回帰統計						回帰統計					
	重相関 R	1					重相関 R	0.892272				
	重決定 R2	1					重決定 R2	0.79615				
	補正 R2	1					補正 R2	0.785131				
	標準誤差	1.84E-16					標準誤差	0.46354				
	観測数	40					観測数	40				
	分散分析表						分散分析表					
	自由度	変動	分散	割られた分散	有意 F		自由度	変動	分散	割られた分散	有意 F	
	回帰	2	39.00003	19.50002	5.74E+32	0	回帰	2	31.04982	15.52491	72.25286	1.67E-13
	残差	37	1.26E-30	3.4E-32			残差	37	7.950159	0.214869		
	合計	39	39.00003				合計	39	38.99998			
	係数	標準誤差	t	P-値	下限 95%	上限 95%	係数	標準誤差	t	P-値	下限 95%	上限 95%
	切片	1.56E-17	2.91E-17	0.53506	0.59581	-4.3E-17 7.47E-17	切片	-3.9E-07	0.073292	-5.4E-06	0.999996	-0.1485 0.148503
	駅距離	-0.75781	3.1E-17	-2.4E+16	0	-0.75781 -0.75781	駅距離	-0.69055	0.078037	-8.84893	1.15E-10	-0.84866 -0.53243
	築年	-0.45921	3.1E-17	-1.5E+16	0	-0.45921 -0.45921	築年	-0.39076	0.078037	-5.00742	1.38E-05	-0.54888 -0.23265
赤: 予測値 青: 観測値	誤差項 $\sigma = 0$						誤差項 $\sigma = 50$					
	$y = -0.8996x + 3E-07$ $R^2 = 0.8092$						$y = -0.8112x - 1E-07$ $R^2 = 0.658$					
	$\sigma = 0$ $y = -0.6931x - 4E-07$ $R^2 = 0.4805$						$\sigma = 50$ $y = -0.6039x - 7E-07$ $R^2 = 0.3647$					

	誤差項 σ = 100						誤差項 σ = 150					
相関	賃料 駅距離 築年						賃料 駅距離 築年					
	賃料	1					賃料	1				
	駅距離	-0.62786	1				駅距離	-0.47166	1			
	築年	-0.44965	0.3087	1			築年	-0.32315	0.3087	1		
重相関	回帰統計						回帰統計					
	重相関 R	0.683041					重相関 R	0.507252				
	重決定 R2	0.466545					重決定 R2	0.257305				
	補正 R2	0.43771					補正 R2	0.217159				
	標準誤差	0.74986					標準誤差	0.884783				
	観測数	40					観測数	40				
	分散分析表						分散分析表					
	自由度	変動	分散	割られた分散	有意 F		自由度	変動	分散	割られた分散	有意 F	
	回帰	2	18.19524	9.097622	16.17958	8.94E-06	回帰	2	10.03489	5.017445	6.409275	0.004074
	残差	37	20.80474	0.56229			残差	37	28.96513	0.782841		
	合計	39	38.99998				合計	39	39.00002			
	係数	標準誤差	t	P-値	下限 95%	上限 95%	係数	標準誤差	t	P-値	下限 95%	上限 95%
	切片	6.26E-08	0.118563	5.28E-07	1	-0.24023 0.240232	切片	-2E-07	0.139897	-1.4E-06	0.999999	-0.28346 0.283457
	駅距離	-0.54056	0.126239	-4.28204	0.000126	-0.79635 -0.28478	駅距離	-0.41107	0.148954	-2.75974	0.008943	-0.71288 -0.10926
	築年	-0.28278	0.126239	-2.24001	0.031183	-0.53856 -0.02699	築年	-0.19625	0.148954	-1.31754	0.195757	-0.49806 0.105556
赤: 予測値 青: 観測値	誤差項σ=100						誤差項σ=150					
	y = -0.6279x + 3E-07 R ² = 0.3942						y = -0.4717x - 6E-08 R ² = 0.2225					
	σ=100 y = -0.4496x - 2E-07 R ² = 0.2022						σ=150 y = -0.3232x - 4E-07 R ² = 0.1044					
												

まとめ

I. 当初設定した関係式 賃料＝1000 - 0.50×(駅距離) - 10.0×(築年) + 誤差項 に対して各 σ に応じて下記の回帰式を得た。

回帰式($\sigma = 0$) 賃料＝1000 - 0.50×(駅距離) - 10.0×(築年)

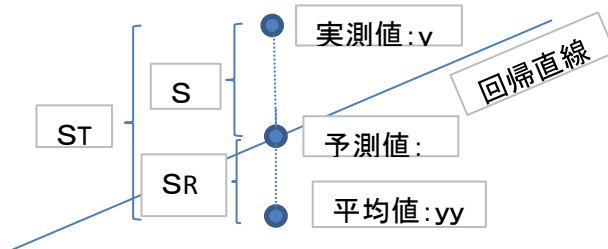
回帰式($\sigma = 50$) 賃料＝ 980 - 0.48×(駅距離) - 9.0×(築年)

回帰式($\sigma = 100$) 賃料＝ 960 - 0.47×(駅距離) - 8.1×(築年)

回帰式($\sigma = 150$) 賃料＝ 940 - 0.45×(駅距離) - 7.1×(築年)

	Y:賃料		駅距離			築年			$a\sigma / \sigma_x$	$b\sigma / \sigma_z$	切片
	μ :平均値	σ :標準偏	a:係数	μ_x :平均値	σ_x :標準	b:係数	μ_z :平均	σ_z :標準			
$\sigma = 0$	605	109	-0.75781	494	165	-0.45921	15	5	-0.50	-10.0	1000
$\sigma = 50$	607	115	-0.69055	494	165	-0.39076	15	5	-0.48	-9.0	980
$\sigma = 100$	609	143	-0.54056	494	165	-0.28278	15	5	-0.47	-8.1	960
$\sigma = 150$	610	182	-0.41107	494	165	-0.19625	15	5	-0.45	-7.1	940

II.



全変動 $ST = \sum (y_i - \bar{y})^2$ 全分散 $VT = 1$

回帰変動 $SR = \sum (\hat{y}_i - \bar{y})^2$ 回帰分散 $VR = SR/2$

残差変動 $SE = \sum (y_i - \hat{y}_i)^2$ 残差分散 $VE = SE/37$ 標準誤差 $\sigma_E = \sqrt{VE}$

重決定 $R^2 = SR/ST$ 、 $ST = SR + SE$

	全変動 ST	回帰変動 ST	残差変動 SE	重決定 R^2 SR/ST	回帰分散 VR	残差分散 VE	分散比 VR/VE	有意F	係数のP-値	
									駅距離	築年
$\sigma = 0$	39.0000	39.0000	0	1	19.5000	3.40E-32	5.74E+32	0	0	0
$\sigma = 50$	39.0000	31.0498	7.9502	0.7962	15.5249	0.21487	7.23E+01	1.67E-13	1.15E-10	1.38E-05
$\sigma = 100$	39.0000	18.1952	20.8047	0.4377	9.0976	0.56229	16.17958	8.94E-06	0.000126	0.031183
$\sigma = 150$	39.0000	10.0349	28.9651	0.2573	5.0174	0.78284	6.409275	0.004074	0.008943	0.195757

以上のシミュレーションにより、どの程度実態に近似した回帰式を得ることができるのかを、重決定 R^2 のレベルに応じて把握することができた。