

第4章

データの再現性の評価

4.1 順位別の勝率分布

仮想データを実データと比較し、結果の妥当性を考察する。その判断基準として、今回は順位別に勝率の平均をとる。まず、実データにおける順位別の勝率分布を図 4.1、仮想データにおける順位別の勝率分布を図 4.2 に表した。

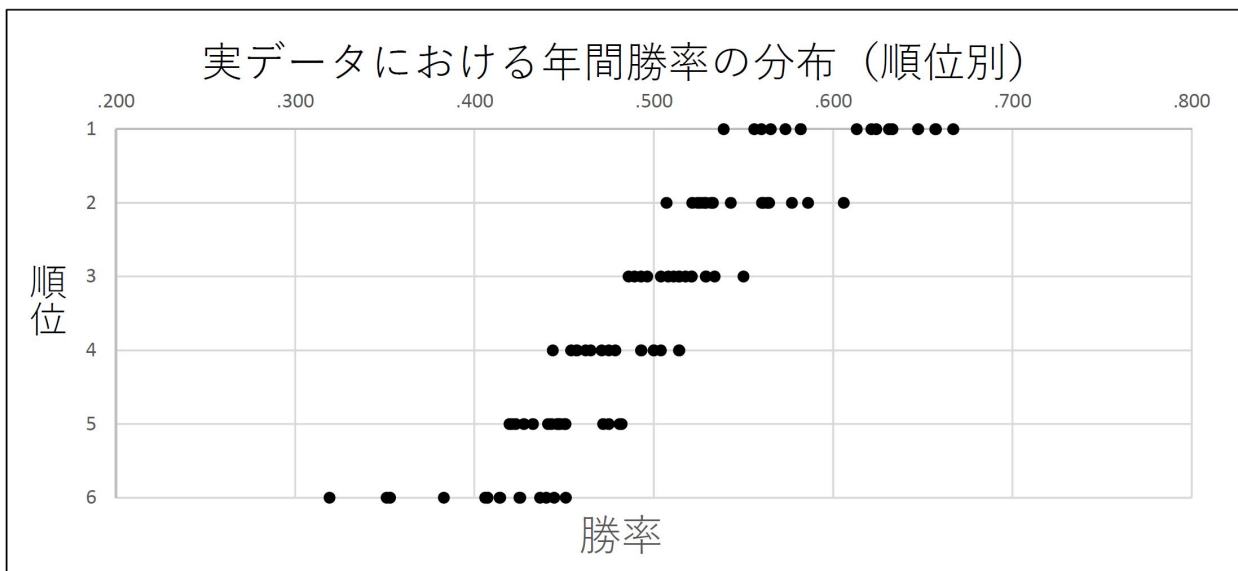


図 4.1 順位別勝率分布 (実データ)

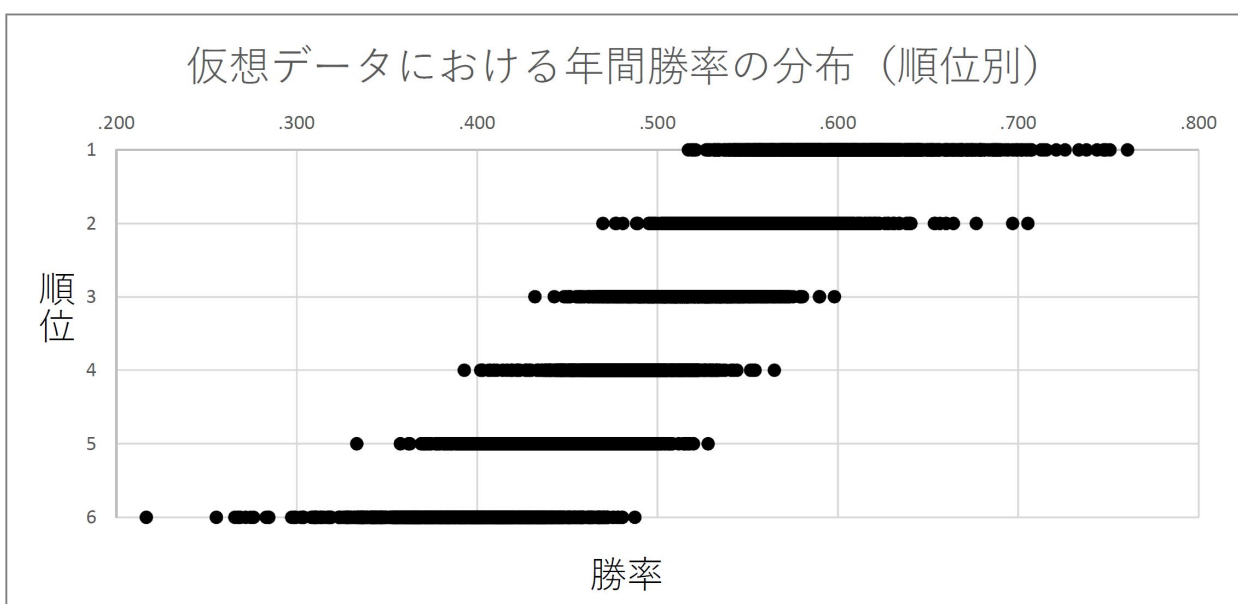


図 4.2 順位別勝率分布 (仮想データ)

これらの平均を実際に計算したところ、表 4.1 の様になった。

表 4.1 順位別平均勝率

順位	実データ	仮想データ	絶対誤差
1	.6068	.6059	.0009
2	.5476	.5539	.0062
3	.5136	.5149	.0012
4	.4801	.4830	.0029
5	.4474	.4472	.0002
6	.4053	.3967	.0086

4.2 考察

図 4.1 と図 4.2 を比較する限りでは、共通して順位別の勝率の最大値、最小値、平均値が順位通りになっており、分布の異常性は感じられない。一見データのばらつきが大きく異なる様に感じられるが、実データの勝率の標準偏差が 0.0716 であるのに対し仮想データの勝率の標準偏差は 0.0758 という近い数値*1であり、分散の違いは単にデータ数の違いによるものと考えられる。ここで、実データの勝率及び仮想データの勝率について、等分散性の検定を行う。

等分散性の検定

正規母集団 $N(\mu_x, \sigma_x^2)$ (母集団 1) から無作為に抽出された観測値を x_1, x_2, \dots, x_{n_x} 、正規母集団 $N(\mu_y, \sigma_y^2)$ (母集団 2) から無作為に抽出された観測値を y_1, y_2, \dots, y_{n_y} とする。このとき、検定統計量

$$F_0 = \frac{s_y^2}{s_x^2}$$

は、帰無仮説 $H_0: \sigma_y^2/\sigma_x^2 = 1$ のもとで、自由度 $(n_x - 1, n_y - 1)$ の F 分布 F_{n_x-1, n_y-1} に従う。ここで、 s_x^2, s_y^2 は、それぞれ x と y の不偏分散

$$s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$$

であり、 \bar{x}, \bar{y} は、それぞれ x と y の標本平均

$$\bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i, \quad \bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$$

である。

上記の検定方法について、 x を実データ、 y を仮想データとする。このとき $n_x = 96, n_y = 6000$ であり、検定統計量 F_0 を求めると $F_0 = 0.9002$ が得られる。 p 値は 0.5097 (> 0.0500) より、有意水準 5% で帰無仮説が受容される。よって、実データの勝率の分布と仮想データの勝率の分布は異なるとは言えない。

今後の課題

今回はモデリングとプログラムの実装に終始してしまった部分があるので、次回は本研究の結果を利用した分析や、異なる条件でシミュレーションを行った際の精度の比較などを行いたい。例として、今回はピタゴラス勝率をそのまま仮想データの勝率として扱ったが、実データにおいてはやはり実際の勝率とピタゴラス勝率の間に多少の誤差が生じる。この誤差を勝率の「補正值」として導入した場合に仮想データの再現精度が上昇するか否かを検証したい。

*1 1 位から 6 位まで全体の勝率の標準偏差である。

補足資料

実データ

以下の項目について、次頁以降の表 4.3~4.5 にプロ野球データ（実データ）を掲載した。ただし紙面の都合上、リーグ名の記載を省略している。（各年度について上段がパシフィック・リーグ、下段がセントラル・リーグである。）

表 4.2 データ集に掲載した項目（表記を本文から変更していないものは割愛）

記号	意味
BA	打率 (Batting Average)
OBP	出塁率
SLG	長打率
RS/G	1 試合平均得点
RS	年間総得点 (Runs Scored)
RA	年間総失点 (Runs Allowed)
Py	ピタゴラス勝率 (指数部は 2)
W%	年間勝率

表 4.3 公式戦データ ('18~'16)

年	順位	球団	BA	OBP	SLG	OPS	RS/G	RS	RA	Py	W%
'18	1	L	.273	.352	.454	.806	5.50	792	653	.595	.624
	2	H	.266	.326	.451	.777	4.76	685	579	.583	.577
	3	F	.251	.329	.393	.722	4.09	589	586	.503	.529
	4	B	.244	.308	.365	.673	3.74	538	565	.476	.471
	5	M	.247	.324	.355	.679	3.71	534	628	.420	.421
	6	E	.241	.307	.368	.675	3.61	520	583	.443	.414
	1	C	.262	.349	.431	.780	5.01	721	651	.551	.582
	2	S	.266	.347	.402	.749	4.57	658	665	.495	.532
	3	G	.257	.325	.403	.728	4.34	625	575	.542	.486
	4	DB	.250	.307	.415	.722	3.97	572	642	.443	.475
	5	D	.265	.325	.380	.705	4.15	598	654	.455	.447
	6	T	.253	.330	.361	.691	4.01	577	628	.458	.440
'17	1	H	.259	.331	.421	.752	4.43	638	483	.636	.657
	2	L	.264	.332	.420	.752	4.79	690	560	.603	.564
	3	E	.254	.324	.390	.714	4.06	585	528	.551	.550
	4	B	.251	.316	.380	.696	3.74	539	598	.448	.444
	5	F	.242	.313	.357	.670	3.53	509	596	.422	.420
	6	M	.233	.297	.351	.648	3.33	479	647	.354	.383
	1	C	.273	.345	.424	.769	5.11	736	540	.650	.633
	2	T	.249	.327	.371	.698	4.09	589	528	.554	.561
	3	DB	.252	.311	.391	.702	4.15	597	598	.499	.529
	4	G	.249	.318	.373	.691	3.72	536	504	.531	.514
	5	D	.247	.300	.365	.665	3.38	487	623	.379	.428
	6	S	.234	.306	.338	.644	3.28	473	653	.344	.319
'16	1	F	.266	.340	.385	.725	4.30	619	467	.637	.621
	2	H	.261	.341	.386	.727	4.42	637	479	.639	.606
	3	M	.256	.326	.363	.689	4.05	583	582	.501	.514
	4	L	.264	.335	.395	.730	4.30	619	618	.501	.457
	5	E	.257	.324	.368	.692	3.78	544	654	.409	.443
	6	B	.253	.317	.355	.672	3.47	499	635	.382	.407
	1	C	.272	.343	.421	.764	4.75	684	497	.654	.631
	2	G	.251	.310	.384	.694	3.60	519	543	.477	.507
	3	DB	.249	.309	.385	.694	3.97	572	588	.486	.493
	4	T	.245	.312	.351	.663	3.51	506	546	.462	.457
	5	S	.256	.331	.378	.709	4.13	594	694	.423	.451
	6	D	.245	.309	.353	.662	3.47	500	573	.432	.414

表 4.4 公式戦データ ('15~'13)

年	順位	球団	BA	OBP	SLG	OPS	RS/G	RS	RA	Py	W%
'15	1	H	.267	.340	.408	.748	4.52	651	491	.637	.647
	2	F	.258	.330	.378	.708	4.27	615	581	.528	.560
	3	M	.257	.320	.368	.688	3.90	561	563	.498	.514
	4	L	.263	.335	.406	.741	4.38	631	573	.548	.500
	5	B	.249	.321	.354	.675	3.60	519	548	.473	.433
	6	E	.241	.311	.338	.649	3.22	463	612	.364	.407
	1	S	.257	.322	.377	.699	3.99	574	518	.551	.539
	2	G	.243	.313	.354	.667	3.40	489	443	.549	.528
	3	T	.247	.317	.343	.660	3.23	465	550	.417	.496
	4	C	.246	.312	.368	.680	3.51	506	474	.533	.493
	5	D	.253	.313	.344	.657	3.28	473	504	.468	.446
	6	DB	.249	.306	.373	.679	3.53	508	598	.419	.437
'14	1	H	.280	.344	.396	.740	4.22	607	522	.575	.565
	2	B	.258	.334	.382	.716	4.06	584	468	.609	.563
	3	F	.251	.321	.379	.700	4.12	593	569	.521	.518
	4	M	.251	.314	.378	.692	3.86	556	642	.429	.465
	5	L	.248	.329	.384	.713	3.99	574	600	.478	.450
	6	E	.255	.327	.356	.683	3.81	549	604	.452	.444
	1	G	.257	.321	.391	.712	4.14	596	552	.538	.573
	2	T	.264	.335	.376	.711	4.16	599	614	.488	.524
	3	C	.272	.337	.420	.757	4.51	649	610	.531	.521
	4	D	.258	.325	.364	.689	3.96	570	590	.483	.479
	5	DB	.253	.317	.383	.700	3.94	568	624	.453	.472
	6	S	.279	.339	.412	.751	4.63	667	717	.464	.426
'13	1	E	.267	.338	.379	.717	4.36	628	537	.578	.582
	2	L	.257	.331	.363	.694	3.96	570	562	.507	.529
	3	M	.262	.331	.374	.705	3.97	572	584	.490	.521
	4	H	.274	.339	.409	.748	4.58	660	562	.580	.514
	5	B	.256	.323	.366	.689	3.56	513	529	.485	.475
	6	F	.256	.326	.368	.694	3.71	534	604	.439	.451
	1	G	.262	.326	.400	.726	4.15	597	508	.580	.613
	2	T	.255	.326	.358	.684	3.69	531	488	.542	.521
	3	C	.248	.319	.367	.686	3.87	557	554	.503	.489
	4	D	.245	.315	.359	.674	3.65	526	599	.435	.454
	5	DB	.262	.325	.390	.715	4.38	630	686	.458	.448
	6	S	.253	.327	.377	.704	4.01	577	682	.417	.407

表 4.5 公式戦データ ('12~'11)

年	順位	球団	BA	OBP	SLG	OPS	RS/G	RS	RA	Py	W%
'12	1	F	.256	.315	.363	.678	3.54	510	450	.562	.556
	2	L	.251	.316	.353	.669	3.58	516	518	.498	.533
	3	H	.252	.304	.349	.653	3.14	452	429	.526	.508
	4	E	.252	.307	.332	.639	3.41	491	467	.525	.500
	5	M	.257	.320	.350	.670	3.47	499	502	.497	.481
	6	B	.241	.301	.337	.638	3.08	443	525	.416	.425
	1	G	.256	.326	.367	.693	3.71	534	354	.695	.667
	2	D	.245	.311	.334	.645	2.94	423	405	.522	.586
	3	S	.260	.325	.361	.686	3.47	499	514	.485	.511
	4	C	.233	.297	.330	.627	2.97	427	454	.469	.462
	5	T	.236	.302	.316	.618	2.85	411	438	.468	.423
	6	DB	.233	.298	.320	.618	2.93	422	571	.353	.351
'11	1	H	.267	.323	.384	.707	3.82	550	351	.711	.657
	2	F	.251	.304	.356	.660	3.35	482	418	.571	.526
	3	L	.253	.318	.366	.684	3.97	571	522	.545	.504
	4	B	.248	.307	.342	.649	3.32	478	518	.460	.504
	5	E	.245	.298	.323	.621	3.00	432	464	.464	.482
	6	M	.241	.301	.316	.617	3.00	432	533	.396	.406
	1	D	.228	.298	.330	.628	2.91	419	410	.511	.560
	2	S	.244	.312	.343	.655	3.36	484	504	.480	.543
	3	G	.243	.298	.354	.652	3.27	471	417	.561	.534
	4	T	.255	.307	.354	.661	3.35	482	443	.542	.493
	5	C	.245	.305	.324	.629	3.05	439	496	.439	.441
	6	DB	.239	.296	.333	.629	2.94	423	587	.342	.353

仮想データ

紙面の都合上, 3年度分のみを掲載する。(ピタゴラス勝率 P_y の指数部は 1.69)

表 4.6 仮想データ (年度 1~年度 3)

サンプル No.	年度	リーグ	順位	OBP	SLG	OPS	得点数	失点数	P_y
2	1	1	1	.345	.423	.768	695	488	.645
5			2	.324	.411	.736	634	520	.583
1			3	.327	.373	.700	567	474	.575
6			4	.302	.368	.670	509	569	.453
4			5	.332	.360	.692	551	684	.410
3			6	.289	.350	.639	452	630	.363
12		2	1	.361	.479	.840	831	643	.607
8			2	.325	.408	.733	629	539	.565
9			3	.369	.452	.821	795	805	.495
10			4	.358	.411	.768	696	764	.461
7			5	.314	.364	.678	524	578	.459
11			6	.302	.364	.665	501	690	.368
13	2	1	1	.330	.372	.702	570	463	.587
18			2	.340	.426	.767	693	567	.584
16			3	.317	.366	.683	535	541	.495
14			4	.319	.351	.670	509	516	.494
15			5	.330	.408	.738	638	675	.476
17			6	.318	.342	.661	492	603	.415
20		2	1	.339	.386	.725	615	448	.631
19			2	.339	.398	.737	637	479	.618
24			3	.342	.384	.726	616	574	.530
21			4	.329	.374	.703	573	672	.433
23			5	.330	.404	.734	631	818	.392
22			6	.304	.336	.639	452	605	.379
26	3	1	1	.334	.412	.746	654	574	.555
27			2	.330	.370	.700	567	512	.543
28			3	.316	.371	.687	542	499	.535
29			4	.324	.346	.670	509	485	.520
25			5	.327	.390	.717	599	613	.490
30			6	.309	.367	.677	523	641	.415
36		2	1	.340	.427	.767	693	427	.694
31			2	.332	.369	.701	568	532	.528
32			3	.304	.362	.666	503	499	.503
34			4	.309	.346	.655	481	522	.465
33			5	.305	.356	.662	495	564	.445
35			6	.280	.280	.560	301	567	.255

プログラムのソースコード (MATLAB)

```
clc
clear

size=6000; % サンプル数 ( 12 の倍数にする事 )
mu = [0.320 0.371]; % OBP と SLG の平均
sigma = [0.000185 0.000291; 0.000291 0.000822]; % OBP と SLG の分散共分散行列
S=zeros(size/6,3); % 12 球団の合計得点数、合計失点数を格納するための行列
U=zeros(size/12,2);
R1=zeros(size,3);
win=zeros(6,3);
WINerror=zeros(6,2);

win(:,1)=[0.606833887 0.547634759 0.513617858 0.480084463 0.447420546 0.405277376];

for n=1:size
    R1(n,1)=n;
    R1(n,2)=ceil(n/12); % 年度の区別
    if rem(ceil(n/6),2)==0 % リーグの区別
        R1(n,3)=2; % リーグ2
    else
        R1(n,3)=1; % リーグ1
    end
end

rng default % For reproducibility
R2 = mvnrnd(mu,sigma,size); % 2 次元正規乱数による OBP と SLG の再現
figure
plot(R2(:,1),R2(:,2),'.') % 散布図

C=horzcat(R1,R2);

year=randi(6,size/12,1); % 各年度・リーグに対して 1 ~ 6 のランダムな整数を割り当てる
nu=12; % 自由度
U(:,1)=trnd(nu,size/12,1); % t 分布による乱数
U(:,2)=U(:,1)*39.5089+49.5962; % 交流戦におけるパ・リーグの得失点差

C(:,6)=C(:,4)+C(:,5); % OPS
C(:,7)=round(C(:,6)*1890.9-757,0); % 単回帰式から予想得点を求め、整数値に丸める
C(:,8)=normrnd(550.99,76.07,[size,1]); % 仮失点数の生成

for i=1:size/6
    for j=1:6
        S(i,1)=S(i,1)+C((i-1)*6+j,7);
        % C7 (得点) を 6 球団分足して年度・リーグごとの合計を記録
        S(i,2)=S(i,2)+C((i-1)*6+j,8);
        % C8 (失点) を 6 球団分足して年度・リーグごとの合計を記録
    end
    for j=1:6
        if C((i-1)*6+j,3)==1
            % リーグ 1 の処理
            C((i-1)*6+j,9)=round(C((i-1)*6+j,8)*(S(i,1)-U(ceil(i/2),2))/S(i,2));
            % ( 各球団の失点数 ) * ( ( 合計得点 ) - ( 交流戦のリーグ 1 得失点差 ) ) / ( 合計仮失点 )
            % ( ( 合計得点 ) - ( 交流戦のリーグ 1 得失点差 ) ) を調整後のリーグ合計失点とする
        else
            % リーグ 2 の処理
            C((i-1)*6+j,9)=round(C((i-1)*6+j,8)*(S(i,1)+U(ceil(i/2),2))/S(i,2));
            % ( 各球団の失点数 ) * ( ( 合計得点 ) + ( 交流戦のリーグ 1 得失点差 ) ) / ( 合計仮失点 )
            % ( ( 合計得点 ) + ( 交流戦のリーグ 1 得失点差 ) ) を調整後のリーグ合計失点とする
        end
        S(i,3)=S(i,3)+C((i-1)*6+j,9);
        % 調整後の合計失点を記録 ( 丸め誤差のため、両リーグの得失点差の和が 0 になるとは限らない )
    end
    if rem(i,2)==0
```



```

        C((i-1)*6+year(i/2),9)=C((i-1)*6+year(i/2),9)+S(i-1,1)+S(i,1)-S(i-1,3)-S(i,3);
        % 調整しても両リーグの得失点差の和が 0 にならない場合、
        % リーグ 2 のランダムな 1 球団の失点数に差を足して無理矢理合わせる
    end
end

C(:,10)=C(:,7).^2./(C(:,7).^2+C(:,9).^2);
C(:,11)=C(:,7).^1.69./(C(:,7).^1.69+C(:,9).^1.69);
% ピタゴラス勝率により勝率を推定する

Sort1=sortrows(C,[2 3 10],{'ascend' 'ascend' 'descend'});

% 年度・リーグごとに勝率の降順に並べ替える
for n=1:size
    Sort1(n,12)=rem(n,6); % 順位の決定
    if rem(n,6)==0
        Sort1(n,12)=Sort1(n,12)+6;
    end
end

for k=1:6
    for l=1:size/6
        win(k,2)=win(k,2)+Sort1(k+(l-1)*6,10);
        win(k,3)=win(k,3)+Sort1(k+(l-1)*6,11);
    end
end

win=win/(size/6);
win(:,1)=win(:,1)*size/6;

for m=1:2
    for k=1:6
        WINerror(k,m)=win(k,m+1)-win(k,1);
    end
end
end

```

参考文献

- [1] <http://npb.jp/> : NPB. jp 日本野球機構, 最終アクセス日 : 2019.5.17
- [2] <http://baseballldata.jp/> : データで楽しむプロ野球, 最終アクセス日 : 2018.10.15
- [3] <https://baseball-data.com/> : プロ野球データ Freak, 最終アクセス日 : 2018.9.13
- [4] <https://ja.wikipedia.org/wiki/> : Wikipedia—野球の各種記録, 最終アクセス日 : 2018.5.15
- [5] <https://bellcurve.jp/statistics/glossary/660.html> : 統計 WEB—統計用語集—因子負荷量, 最終アクセス日 : 2018.5.19
- [6] 小西貞則, 多変量解析入門—線形から非線形へ, 岩波書店, 2010
- [7] 森田浩, 図解入門ビジネス 多変量解析の基本と実践がよ〜くわかる本, 秀和システム, 2014
- [8] 室淳子, 石村貞夫, Excel でやさしく学ぶ多変量解析 [第 2 版], 東京図書株式会社, 2007
- [9] 下川敏雄, 実践のための基礎統計学, 講談社, 2016
- [10] 廣瀬英雄, 実例で学ぶ確率・統計, 日本評論社, 2014
- [11] データスタジアム株式会社, 野球 × 統計は最強のバッテリーである セイバーメトリクスとトラッキングの世界, 中央公論新社, 2015
- [12] 有馬哲, 石村貞夫, 多変量解析のはなし, 東京図書株式会社, 2002
- [13] 石村貞夫, 石村光資郎, 入門はじめての多変量解析, 東京図書株式会社, 2007
- [14] <https://bellcurve.jp/statistics/course/12931.html> : 統計 WEB—統計学の時間—正規性の確認, 最終アクセス日 : 2019.4.19
- [15] Amy N. Langville, Carl D. Meyer (訳) 岩野和生, 中村英史, 清水咲里, レイティング・ランキングの数理, 共立出版, 2015
- [16] <https://jp.mathworks.com/help/stats/kstest.html> : MATLAB—ドキュメンテーション—1 標本コルモゴロフ・スミルノフ検定, 最終アクセス日 : 2019.4.19
- [17] <https://bellcurve.jp/statistics/glossary/1360.html> : 統計 WEB—統計用語集—コルモゴロフ=スミルノフ検定, 最終アクセス日 : 2019.4.19
- [18] <https://bellcurve.jp/statistics/glossary/2071.html> : 統計 WEB—統計用語集—正規 Q-Q プロット, 最終アクセス日 : 2019.4.19