

- ① 導入
- ② 準備
- ③ プログラムの実装
- ④ 結果と考察
- ⑤ 終わりに

- 「OBP, SLG」「失点数」に対し正規乱数⁵を, 「交流戦におけるリーグ間の得失点差」に対し t 分布乱数を使用して仮想データを生成したい
- 上記の実データについて, 正規性 (正規分布に従うこと) を仮定する必要がある
- Q-Q プロットが 1 直線上に並んでいることを目視で判断し, さらにコルモゴロフ・スミルノフ検定を行う⁶

⁵正規分布に従う擬似乱数

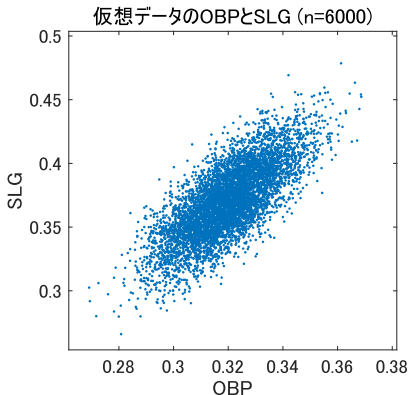
⁶検定結果は資料を参照されたい.

OBP, SLG の出力結果

- 以下の平均 μ , 分散共分散行列 Σ による 2 次元正規分布に従う乱数

$$\mu = \begin{pmatrix} 0.320 \\ 0.371 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.000185 & 0.000291 \\ 0.000291 & 0.000822 \end{pmatrix}$$

- 以下の散布図は出力された 6000 組のデータのプロット



① 得点数

- OPS = OBP + SLG という式で OPS を計算する
- OPS は得点との連動性が高いことを利用し、以下の単回帰式で得点数を確定させる

$$(\text{得点数}) = -757.003 + 1890.903 \times (\text{OPS})$$

② 失点数

- 平均 $\mu = 550.99$, 標準偏差 $\sigma = 76.07$ の正規乱数⁷を用いる。ただし、ここで得た数値は仮の失点数（仮失点）として扱い、微調整をかけた上で失点数を決定する。
- 調整する要因：(12 球団の得点数の合計) = (12 球団の失点数の合計) という関係を満たさなければならないため。（詳細は次のスライド）

⁷OPS と SLG を出力した正規乱数とは独立したもの。

合計得点と合計失点の関係

- (パ (セ) 6 球団の得点数の合計) = (パ (セ) 6 球団の失点数の合計) という関係性は、交流戦が始まると崩れる。
- しかしペナントレースで NPB 以外の球団と対戦することは無いので、先述の (12 球団の得点数の合計) = (12 球団の失点数の合計) は常に成り立つ。

Table : '19 得失点 (5/18 終了時点)

パ・リーグ				セ・リーグ			
順位	球団	得点	失点	順位	球団	得点	失点
1	H	180	154	1	G	208	154
2	F	167	153	2	C	165	164
2	M	171	161	3	S	213	210
4	E	201	198	3	T	170	188
5	L	207	222	5	D	144	158
6	B	134	172	6	DB	158	184
合計		1060	1060	合計		1058	1058

交流戦の得失点差

- 仮想データでの勝率の並び方がいびつ⁸にならない様に、交流戦のリーグ間得失点差を「違和感の無いレベル」に抑えたい
- 正規乱数を用いるには実データのサンプル数が少なすぎるため、自由度 $n - 2$ の t 分布乱数 ($\mu = 49.60, \sigma = 39.51$) を使用する

Table : 交流戦におけるリーグ間の得失点差 ('06~'18)

年度	2018	2017	2016	2015	2014	2013	2012
試合数	18	18	18	18	24	24	24
パ得点 - セ得点	64	28	66	72	58	99	20
18 試合換算	64	28	66	72	43.5	74.25	15
年度	2011	2010	2009	2008	2007	2006	
試合数	24	24	24	24	24	36	
パ得点 - セ得点	139	155	25	62	37	-63	
18 試合換算	104.25	116.25	18.75	46.5	27.75	-31.5	

⁸極端な例としては、「片方のリーグは勝率 6 割以上が 3 球団あり、もう片方のリーグは勝率 5 割未満の球団が首位」といった異常な状況にもなり得る。

失点数の調整

$$(\text{リーグ 1 各球団の失点数}) = (\text{仮失点数}) \times \frac{3408 - 43}{3355.14}$$

$$(\text{リーグ 2 各球団の失点数}) = (\text{仮失点数}) \times \frac{3976 + 43}{3196.64}$$

Table : 年度 1 の仮失点と失点（交流戦のリーグ間得失点差：43）

リーグ 1				リーグ 2			
No.	得点	仮失点	失点	No.	得点	仮失点	失点
1	567	472.61	474	7	524	459.88	578
2	695	486.24	488	8	629	428.48	539
3	452	628.02	630	9	795	640.32	805
4	551	682.23	684	10	696	607.98	764
5	634	518.96	520	11	501	548.36	690
6	509	567.09	569	12	831	511.62	643
合計	3408	3355.14	3365	合計	3976	3196.64	4019

- プロ野球のペナントレースにおいて、順位の設定要素は勝率である。

$$(\text{勝率}) = \frac{(\text{勝利数})}{(\text{勝利数}) + (\text{敗戦数})}$$

- 本来であれば年間勝率は公式戦を 143 試合戦ってようやく確定するものだが、本研究では最適化したピタゴラス勝率の式

$$\frac{1}{1 + \rho^{1.69}} \quad \left(\rho = \frac{\text{失点数}}{\text{得点数}} \right)$$

に得点数と失点数を代入して求める。つまり、純粋に得点が増えるほど勝率は上がり、失点が増えるほど勝率は下がる

- 仮想データの順位はピタゴラス勝率の序列で機械的に決定する。