

重回帰分析で競馬予想

江藤一樹

1 競馬の順位を予想するためには...

- ▶ 予想するレースに出る馬の一頭一頭のデータを集める。
- ▶ 重回帰分析を行って各馬のタイムを予想する。
- ▶ 予想したタイムから順位を決める。

2馬一頭のデータ

日付	間隔	開催	レース名	頭数	馬番	人気	着順	騎手	斤量	芝・ダート	距離	馬場状態	タイム
2017. 2.19		4 1東8	フェブラG1	16	5	9	15	和田竜二	57	1	1600	0	1366
2017. 1.22		7 1名4	東海SG2	16	1	2	8	和田竜二	57	1	1800	0	1536
2016.12. 4		4 4名2	チャンピG1	15	4	10	3	和田竜二	57	1	1800	0	1502
2016.11. 6		4 5京2	みやこSG3	16	11	2	14	和田竜二	58	1	1800	0	1515
2016.10.10		15 盛岡	マイルG1	14	8	3	4	太宰啓介	57	1	1600	1	1346
2016. 6.29		5 大井	帝王賞G1	12	1	3	6	太宰啓介	57	1	2000	3	2067
2016. 5.21		5 3京9	平安SG3	15	14	1	1	太宰啓介	58	1	1900	0	1562
2016. 4.16		8 2阪7	アンタレG3	16	16	3	2	太宰啓介	58	1	1800	0	1500
2016. 2.21		4 1東8	フェブラG1	16	4	7	3	太宰啓介	57	1	1600	2	1342
2016. 1.24		5 1名4	東海SG2	12	6	4	1	太宰啓介	56	1	1800	0	1519
2015.12.20		6 5阪6	ベテルギ	11	6	3	2	太宰啓介	56	1	2000	0	2040
2015.11. 8		24 5京2	みやこSG3	11	7	8	4	太宰啓介	56	1	1800	3	1483
2015. 5.23		5 3京9	平安SG3	16	4	9	15	太宰啓介	56	1	1900	0	1589
2015. 4.18		5 2阪7	アンタレG3	16	16	8	14	太宰啓介	56	1	1800	0	1523
2015. 3.11		3 船橋	ダイオG2	12	6	3	3	太宰啓介	55	1	2400	3	2342
2015. 2.14		14 2京5	アルデバH	15	1	4	1	太宰啓介	54	1	1900	0	1568
2014.11. 9		2 5京2	みやこSG3	16	16	8	10	太宰啓介	54	1	1800	0	1511
2014.10.25		11 4京6	観月橋S1600	16	4	4	1	太宰啓介	55	1	1800	0	1500
2014. 8.10		5 2新4	レパードG3	15	7	3	5	太宰啓介	56	1	1800	1	1513
2014. 7. 6		6 3名2	濃尾特別1000	15	10	2	1	太宰啓介	54	1	1800	1	1502
2014. 5.25		4 3京10	鳳雛S	11	3	4	2	太宰啓介	56	1	1800	0	1524
2014. 4.27		9 3京2	500万下*	13	3	2	1	太宰啓介	56	1	1800	0	1530
2014. 2.23		2 2京8	500万下*	11	1	1	4	和田竜二	56	1	1800	0	1548
2014. 2. 9		4 2京4	500万下*	11	4	1	2	和田竜二	56	1	1800	0	1524

2 馬一頭のデータ

- ▶ 各馬のデータは35行からできていて、すべての行に名前が付いている。
- ▶ 今まで出てきたレースの日付、頭数、馬場状態、騎手、タイム、体重...etc

3 過去のレースを予想

- ▶ 今回予想するレースは今年の2月に行われたG1のフェブラリーステークスというレース。
- ▶ G1とは競馬のレースで最もランクが高いレースのことを言います。
- ▶ 強い馬が集まるため過去のデータも集まりやすい

フェブラリーステークスとは...？

- ▶ 馬の数は最大16頭まで出れることができる。
- ▶ 距離1600メートル、ダートコース（土で出来ているコース）のレースとなっている。

4 説明変数と目的変数を決める

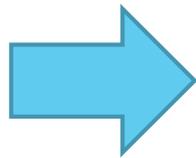
- ▶ 目的変数：タイム
- ▶ 説明変数：レース間隔、頭数、馬番、芝.ダート、距離、馬場状態、体重

4 説明変数と目的変数を決める

- ▶ 芝の時は0、ダートの場合は1を芝・ダートの行に記入する。
- ▶ 馬場状態（走るコースの状態のこと）
良、稍重、重、不良の4つの状態がある。
良 = 0、稍重 = 1、重 = 2、不良 = 3とする。

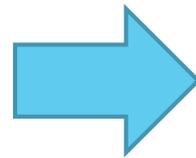
4 説明変数と目的変数を決める

芝・ダート
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ
ダ



芝・ダート
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

馬場状態
良
良
重
稍
良
稍
重
不
稍
良
良
良
良
重
良
稍
稍
良
良
重
良
良
重



馬場状態
0
0
0
0
1
3
0
0
2
0
0
3
0
0
3
0
0
0
1
1
0
0
0
2

5 回帰式を求めて予想

▶ **タイム** = (間隔) * x 1 + (頭数) * x 2 +
(馬番) * x 3 + (芝、ダート) * x 4 +
(距離) * x 5 + (体重) * x 6 +
(馬場状態) * x 7

5 回帰式を求めて予想

	予想タイム	実際のタイム	誤差
アスカノロマン	1399.664	1366	33.664
インカンステーション	1379.807	1365	14.807
エイシンバッケン	1358.5757	1354	4.5757
カフジテイク	1372.799	1352	20.799
キングズガード	1360.7866	1360	0.7866
ケイティブレイブ	1311.648	1356	-44.352
ゴールドドリーム	1571.733	1351	220.733
コパノリッキー	1361.572	1365	-3.428
サウンドトゥルー	1376.213	1357	19.213
デニムアンドルピー	1442.624	1367	75.624
ニシケンモノノフ	1375.693	1356	19.693
ノンコノユメ	1253.064	1357	-103.936
ブライトライト	1366.046	1360	6.046
ベストウォーリア	1501.351	1351	150.351
ホワイトフーガ	1398.869	1359	39.869
モーニン	1366.043	1361	5.043

実際の順位	予想した順位
ゴールドドリーム	ノンコノユメ
ベストウォーリア	ケイティブレイブ
カフジテイク	エイシンバッケン
エイシンバッケン	キングズガード
ケイティブレイブ	コパノリッキー
ニシケンモノノフ	モーニン
サウンドトゥルー	プライトライト
ノンコノユメ	カフジテイク
ホワイトフーガ	ニシケンモノノフ
キングズガード	サウンドトゥルー
プライトライト	インカンステーション
モーニン	ホワイトフーガ
インカンステーション	アスカノロマン
コパノリッキー	デニムアンドルピー
アスカノロマン	ベストウォーリア
デニムアンドルピー	ゴールドドリーム

5 回帰式を求めて予想

	出たレースの数	タイム誤差
ゴールドドリーム	8	220.733
モーニン	11	5.043
ノンコノユメ	16	-103.936
ケイティブレイブ	17	-44.352
ホワイトフーガ	17	39.869
エイシンバッケン	19	4.5757
デニムアンドルピー	19	75.624
カフジテイク	22	20.799
キングズガード	23	0.7866
インカンステーション	25	14.807
アスカノロマン	27	33.664
コパノリッキー	27	-3.428
ベストウォーリア	27	150.351
ニシケンモノノフ	32	19.693
プライトライト	36	6.046
サウンドトゥルー	40	19.213

6 結果

▶ 予想した順位と実際の順位が一致した馬の数は...

0 頭

大失敗...(泣)

まとめ

- ▶ 予想タイムと実際のタイムに差がある馬の数が多すぎる。
- ▶ 出たレースの数がタイムの誤差に出ているわけではない。
- ▶ タイムに強く影響する説明変数が距離以外にないため細かいタイムまで予想するのが難しい。
- ▶ タイムに影響する要因が今回挙げた7つでは少なすぎたため良い結果に結びつかなかったと思われる。