

# UEFA EURO2024 優勝の方程式

2年4組35番 富岡佑斗

# UEFA EURO2024とは

- 欧州サッカー連盟 (UEFA) が主催する、ナショナルチームによるサッカーの大陸選手権大会のこと。EUROとも呼ばれる。
- 日本からするとAFCアジアカップ
- 欧州にはサッカーの強豪国が存在するのに加え、選手の知名度も高く、注目が集まる。
- 4年に1度開催され、それが今年の6/14~7/14にドイツで行われた。



TM

UEFA  
**EURO2024**  
GERMANY



ABEMA



UEFA  
**EURO2024**  
GERMANY

UEFA EURO 2024™

アベマで 日本史上初!

**全試合無料**

**生中継決定!**



- ベスト8が揃っ  
→自分が賭けた



なので今回はもしRを使ってデータ分析していたら、  
賭けに勝つことができたのかを分析してみる。

FIFA World Cup  
Qatar

優勝予想は難しい!!!

3

4



Costa

2022





# 分析方法

- 決定木を作成し、どの説明変数が一番優勝争いに影響しているのかを分析する。
- 基準変数は最終順位
- 決定木を作成したあと、予測された値と実際の値がどのくらい近いのかを分析

# 用いたデータ

- 前回大会EURO2020の24参加国のデータを使用
- 大会での成績に関係ありそうな項目

国名	FIFAランク	出場回数	過去の優勝回数	5大リーグ率	平均年齢	平均出場回数	予選勝率	予選得点	コンディション	予選失点	最終順位
イタリア	7	10	1	1	27.5	25.8	1	37	2	4	1
スイス	13	5	0	0.77	27.89	33.7	0.625	19	12	6	8
トルコ	29	5	0	0.54	25.19	21.2	0.7	18	24	3	24
ウェールズ	17	2	0	0.73	24.77	29.2	0.5	10	8	4	16
ベルギー	1	6	0	0.85	30.04	51.5	1	40	3	3	8
ロシア	38	12	0	0.12	27.04	21.4	0.8	33	18	8	24
デンマーク	10	9	1	0.81	24.19	30	0.5	23	6	6	4
フィンランド	54	1	0	0.31	26.12	32	0.6	16	17	10	24
ウクライナ	24	3	0	0.12	27.65	25.7	0.75	17	9	4	8
オランダ	16	10	1	0.46	25.46	24.3	0.75	24	5	7	16
オーストリア	23	3	0	0.85	24.42	29.6	0.6	19	11	9	16
北マケドニア	62	1	0	0.35	27.23	30.2	0.4	12	23	13	24
イングランド	4	10	0	1	27.04	20.9	0.875	37	1	6	2
クロアチア	14	6	0	0.54	27.54	35.1	0.625	17	13	7	16
チェコ	40	10	0	0.46	27.23	21	0.625	13	7	11	8
スコットランド	44	3	0	0.54	27.12	18.5	0.5	16	22	19	24
スペイン	6	11	3	1	24.42	27.73	0.8	31	4	5	4
ポーランド	20	4	0	0.69	25.96	28.2	0.8	18	21	5	24
スウェーデン	18	7	0	0.46	28.54	36.9	0.6	23	8	9	16
スロバキア	36	2	0	0.42	28.15	34.7	0.5	13	19	11	24
ドイツ	12	13	3	1	27.31	32.1	0.875	30	14	7	16
フランス	2	10	2	1	27.31	38.9	0.8	25	10	6	16
ポルトガル	5	8	1	0.77	27.46	40.2	0.625	22	15	6	16
ハンガリー	37	4	0	0.19	27.84	20.7	0.5	8	20	11	24



## 説明変数

- FIFAランク(2021年4月時点)
- 出場回数
- 過去の優勝回数
- 欧州5大リーグの率
- 平均年齢
- 平均出場回数(A代表の出場回数)
- 予選勝率・予選得点・予選失点
- 直近の試合のコンディション



## コード

```
1 library(rpart)
2 par(family= "HiraKakuProN-w3")
3 サッカーデータ <- read.csv("euro.csv", header = T, row.names = 1)
4 サッカー木 <- rpart(最終順位~., data=サッカーデータ, method="anova",minsplit = 5)
5 print(サッカー木)
6 plot(サッカー木, uniform=T, branch=0.6, margin=0.1)
7 text(サッカー木, all=T, use.n=T)
8 plotcp(サッカー木)
9 サッカー木2 <- prune(サッカー木,cp = 0.34)
10 print(サッカー木2)
11 plot(サッカー木2, uniform=T, branch=0.6, margin=0.1)
12 text(サッカー木2, all=T, use.n=T)
13 サッカー予測 <- predict(サッカー木,newdata=サッカーデータ,type= "vector")
14 cor(サッカーデータ$最終順位,サッカー予測)
15 plot(サッカーデータ$最終順位,サッカー予測)
16
```

# 結果

node), split, n, deviance, yval  
\* denotes terminal node

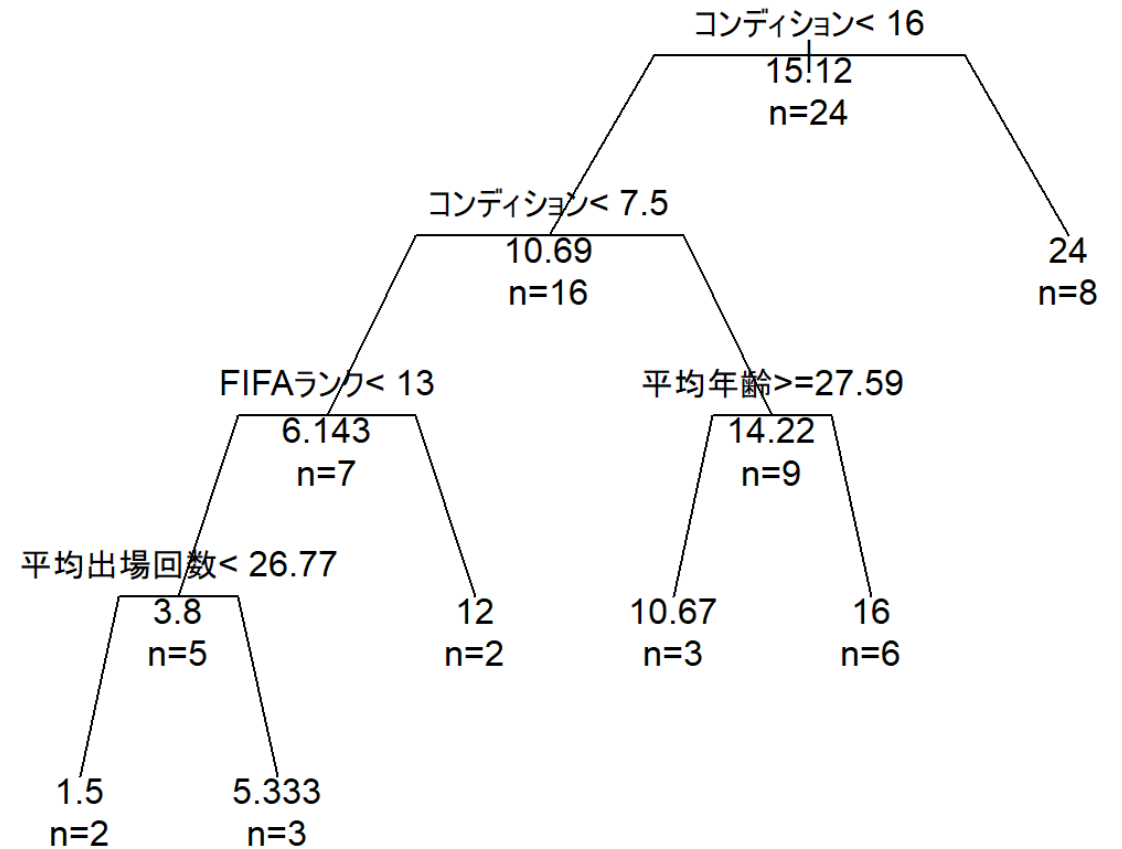
```
1) root 24 1458.62500 15.125000
2) コンディション< 16 16 513.43750 10.687500
4) コンディション< 7.5 7 156.85710 6.142857
8) FIFAランク< 13 5 28.80000 3.800000
16) 平均出場回数< 26.765 2 0.50000 1.500000 *
17) 平均出場回数>=26.765 3 10.66667 5.333333 *
9) FIFAランク>=13 2 32.00000 12.000000 *
5) コンディション>=7.5 9 99.55556 14.222220
10) 平均年齢>=27.595 3 42.66667 10.666670 *
11) 平均年齢< 27.595 6 0.00000 16.000000 *
3) コンディション>=16 8 0.00000 24.000000 *
```

split = 分岐条件

n = ノードに含まれる個体数

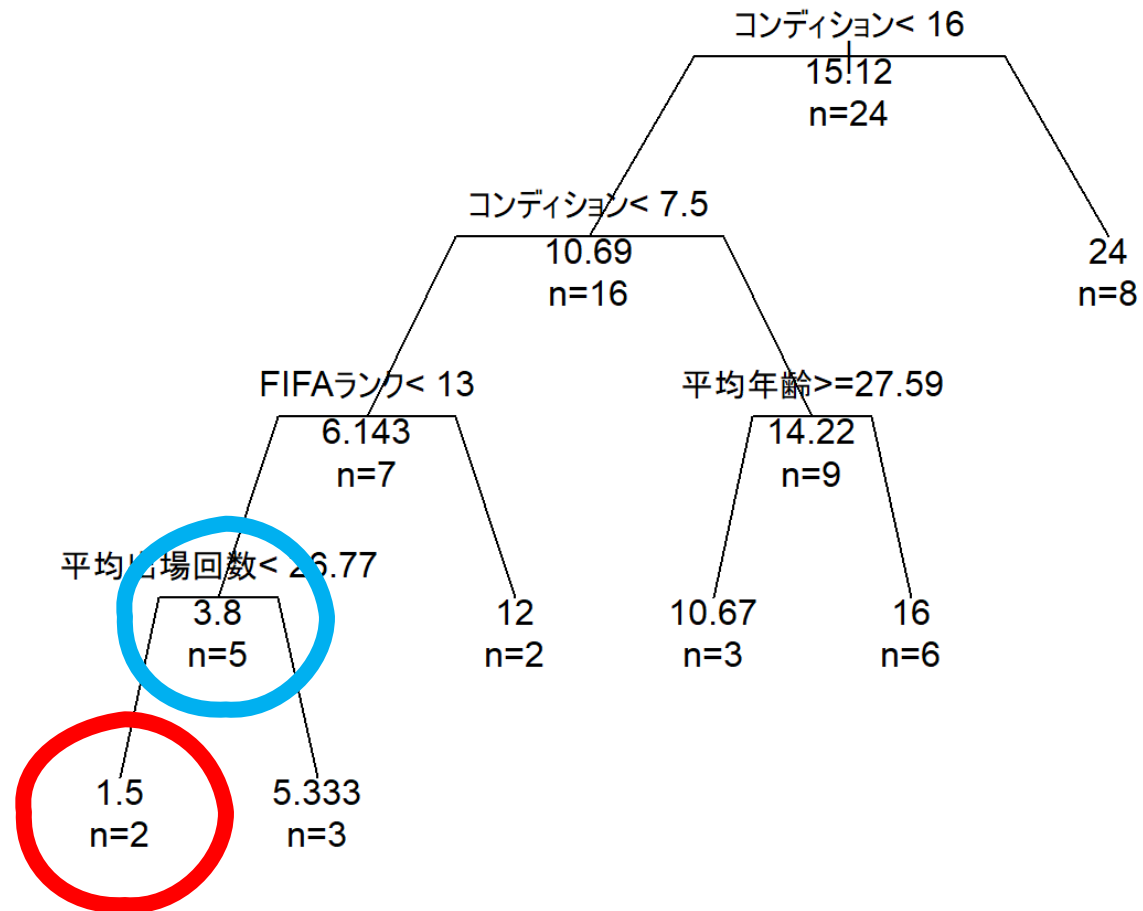
deviance = ノード内の偏差平方和

yval = 基準変数の平均





# 結果



条件1:

コンディション7位以上かつFIFAランク12以上

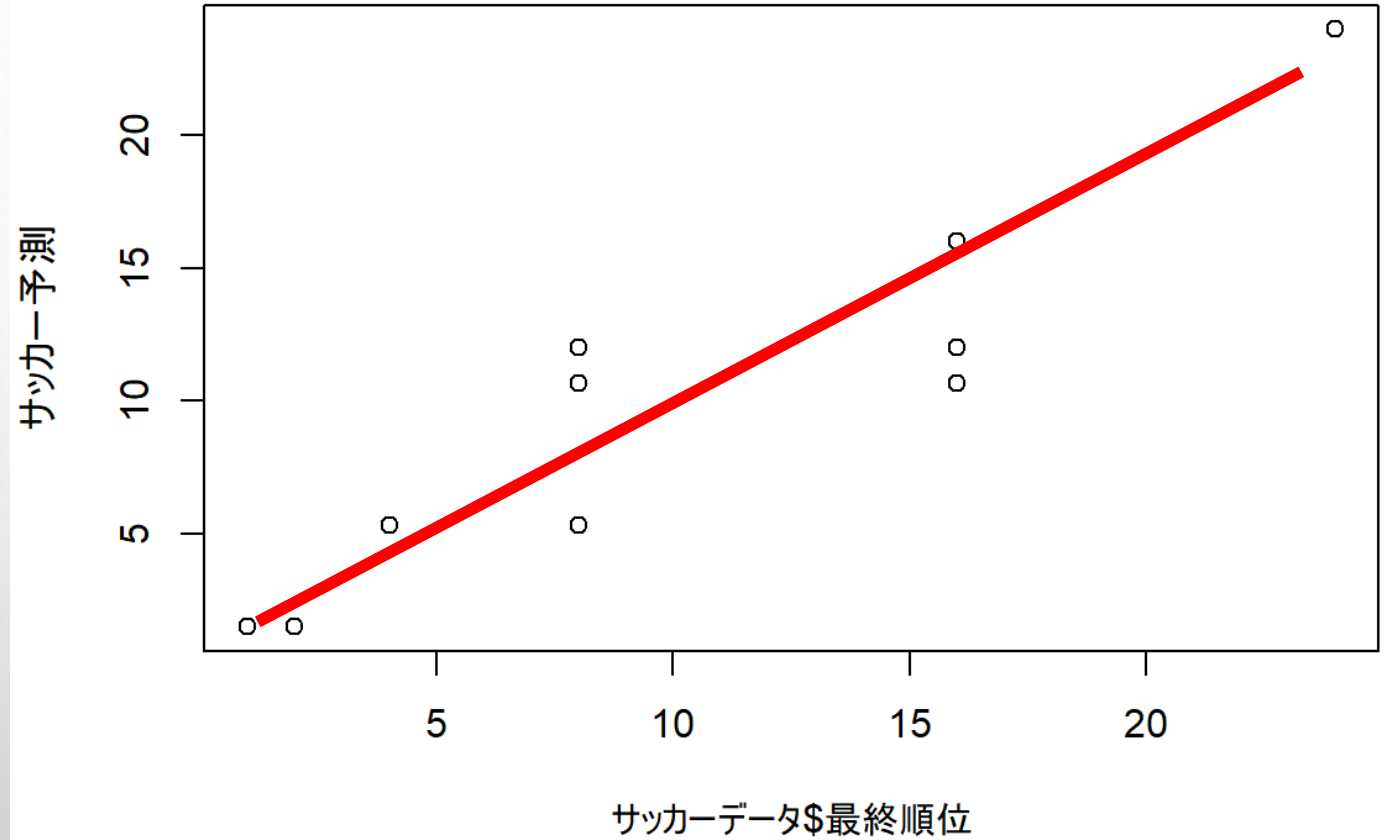
この時点でベスト4

条件2:

条件1を満たし平均出場回数26回以下だと50%の確率で優勝(トップ2)

# 結果

予測値と実測値の  
相関係数は0.97



# EURO2024の結果と照らし合わせてみる

優勝 : スペイン(CD 1位・FIFAランク8位)

準優勝 : イングランド(CD 4位・FIFAランク5位)

ベスト4 : オランダ(CD 2位・FIFAランク7位)

フランス(CD 7位・FIFAランク2位)

ベスト8 : ドイツ(CD 3位・FIFAランク16位)

ポルトガル(CD 8位・FIFAランク6位)

トルコ(CD 5位・FIFAランク42位)

スイス(CD 6位・FIFAランク19位)



# 条件1を満たすのは

優勝 : スペイン(CD 1位・FIFAランク8位)

準優勝 : イングランド(CD 4位・FIFAランク5位)

ベスト4 : オランダ(CD 2位・FIFAランク7位)

フランス(CD 7位・FIFAランク2位)

ベスト8 : ドイツ(CD 3位・FIFAランク16位)

ポルトガル(CD 8位・FIFAランク6位)

トルコ(CD 5位・FIFAランク42位)

スイス(CD 6位・FIFAランク19位)

## 4カ国の条件2について

### 4カ国の平均出場回数

- スペイン: 21.23
- イングランド: 25.08
- オランダ: 26.77
- フランス: 32.85

## 条件2を満たすのは

### 4カ国の平均出場回数

- スペイン: 21.23
- イングランド: 25.08
- オランダ: 26.77
- フランス: 32.85

2024年の結果と全く一緒に！！！！！！！！！！



# 分析の結果

- 予測値と実測値の間には強い相関関係があった。
- 優勝への方程式は直近の試合のコンディションを上げ、FIFAランクもあげること。
- ベスト8が決まった段階でベスト4・トップ2を予想してみると、2024の結果と全く一緒になった。
- ちなみにトップ2はスペイン・イングランドと予想できたが、この2チームなら確実にスペインを選んでいた。

## まとめと感想

- めっちゃうまくいった。
- 分析してから勝負してたら、奢る羽目になってなかった。むしろ奢ってもらえていた。
- 出場回数が少ないほうが上位になりやすいのは意外だった。
- データを自分で0から集めるのは大変。
- 1回分のデータだけでやってこのレベルなら、何回分かのデータを集めてみたらとんでもないのでは??