# [NBA選手のピークはいつか？ PER評価指標を用いた分析](http://kokiando.hatenablog.com/entry/%E2%96%A0)

[NBA](http://kokiando.hatenablog.com/archive/category/NBA) [Data Manipulation](http://kokiando.hatenablog.com/archive/category/Data%20Manipulation)

この記事は[スポーツアナリティクス Advent Calendar 2019](https://adventar.org/calendars/3930)の14日目の記事になります。

* [PERとは](http://kokiando.hatenablog.com/preview#PER%E3%81%A8%E3%81%AF)
* [データ](http://kokiando.hatenablog.com/preview#%E3%83%87%E3%83%BC%E3%82%BF)
* [分析](http://kokiando.hatenablog.com/preview#%E5%88%86%E6%9E%90)
  + [ポジションごとの年齢分布](http://kokiando.hatenablog.com/preview#%E3%83%9D%E3%82%B8%E3%82%B7%E3%83%A7%E3%83%B3%E3%81%94%E3%81%A8%E3%81%AE%E5%B9%B4%E9%BD%A2%E5%88%86%E5%B8%83)
  + [2変量分析：年齢とPER](http://kokiando.hatenablog.com/preview#2%E5%A4%89%E9%87%8F%E5%88%86%E6%9E%90%E5%B9%B4%E9%BD%A2%E3%81%A8PER)
  + [考察](http://kokiando.hatenablog.com/preview#%E8%80%83%E5%AF%9F)
* [まとめ](http://kokiando.hatenablog.com/preview#%E3%81%BE%E3%81%A8%E3%82%81)
* [問題点](http://kokiando.hatenablog.com/preview#%E5%95%8F%E9%A1%8C%E7%82%B9)
* [バスケットのデータ分析に興味のある人へ](http://kokiando.hatenablog.com/preview#%E3%83%90%E3%82%B9%E3%82%B1%E3%83%83%E3%83%88%E3%81%AE%E3%83%87%E3%83%BC%E3%82%BF%E5%88%86%E6%9E%90%E3%81%AB%E8%88%88%E5%91%B3%E3%81%AE%E3%81%82%E3%82%8B%E4%BA%BA%E3%81%B8)

先日[Journal of Sports Analytics](https://www.journalofsportsanalytics.com/)を徘徊していたら、以下のようなペーパーを発見しました。

[content.iospress.com](https://content.iospress.com/articles/journal-of-sports-analytics/jsa0021)

**サッカー選手のパフォーマンスピークが何歳の時に訪れるか**、といったペーパーになります。選手の評価指標として [WhoScore.com](http://whoscore.com/)というサイトが公開している10段階評価のrating指標を用いて分析した結果、全体としては**25~27才**の時にピークを迎える。しかし、ポジションごとに違いが見られ、[フォワ](http://d.hatena.ne.jp/keyword/%A5%D5%A5%A9%A5%EF)ードは25才あたりで[ディフェンダー](http://d.hatena.ne.jp/keyword/%A5%C7%A5%A3%A5%D5%A5%A7%A5%F3%A5%C0%A1%BC)は27才あたりにパフォーマンスがピークになる、といった興味深い結果が提示されています。

自分が興味のあるバスケット、特に[NBA](http://d.hatena.ne.jp/keyword/NBA)ではPERという指標が多く用いられ選手が評価されているので、自分なりにやってみようと思ったため、筆を取りました。

# PERとは

大まかな説明になりますが…

PERとは、[ESPN.com](https://www.espn.com/nba/)のコラムニストによって開発された出場時間毎の**エフィシエンシー(Player Efficiency Rating)を示す指標**になります。他のスタッツと異なる点は、プラスに働いたスタッツ(ポイントとかシュート成功率など…)から、マイナスに働いたスタッツ(シュートミスやファールなど…)を差し引いたもので、選手を評価・比較する上で重宝されています。

(原文)The Player Efficiency Rating (PER) is a per-minute rating developed by [ESPN](http://d.hatena.ne.jp/keyword/ESPN).com columnist John Hollinger. In John's words, "The PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance." It appears from his books that John's database only goes back to the 1988-89 season. I decided to expand on John's work and calculate PER for all players since minutes played were first recorded (1951-52).

この記事では、PER指標がどのように算出されているかについて深堀りは致しませんので、詳細は[こちら](https://www.basketball-reference.com/about/per.html)の公式ページをご覧ください。

参考までに今シーズンの上位10選手を提示しておきます。

今シーズンの上位10選手 (2019/12/10現在)

[Hollinger's NBA Player Stats - ESPN Insider - National Basketball Association - ESPN](http://insider.espn.com/nba/hollinger/statistics)

名だたる選手が散見されますが、興味深いのが上位二名の選手たち。どちらも非常に若く、2位の選手は今季2年目かつ現在20才の[スロベニア](http://d.hatena.ne.jp/keyword/%A5%B9%A5%ED%A5%D9%A5%CB%A5%A2)人ルカ・ドンチッチ選手になります。

つい先日まで[ティー](http://d.hatena.ne.jp/keyword/%A5%C6%A5%A3%A1%BC)ンエイジャーだった選手が現在の[NBA](http://d.hatena.ne.jp/keyword/NBA)候補です

# データ

今回分析対象にしたデータは[こちら](https://www.kaggle.com/drgilermo/nba-players-stats#Seasons_Stats.csv)より拝借し、2000~01から2017~18シーズンの間の20~40才の選手で、そのシーズン20試合以上プレーした選手に絞りました。

また選手を比較し分類するために今回は**バックコートとフロントコートポジションの二種類**にカテゴライズしました。

**バックコート**：PG([ポイントガード](http://d.hatena.ne.jp/keyword/%A5%DD%A5%A4%A5%F3%A5%C8%A5%AC%A1%BC%A5%C9))とSG([シューティングガード](http://d.hatena.ne.jp/keyword/%A5%B7%A5%E5%A1%BC%A5%C6%A5%A3%A5%F3%A5%B0%A5%AC%A1%BC%A5%C9))。 [NBA](http://d.hatena.ne.jp/keyword/NBA)選手でいうと、カイリー・[アービング](http://d.hatena.ne.jp/keyword/%A5%A2%A1%BC%A5%D3%A5%F3%A5%B0)やステフォン・カリー。[スラムダンク](http://d.hatena.ne.jp/keyword/%A5%B9%A5%E9%A5%E0%A5%C0%A5%F3%A5%AF)でいうところの宮城や三井です。どちらかというとゲムメイクやアウトサイドでプレーするポジションです。

**フロントコート**：SF([スモールフォワード](http://d.hatena.ne.jp/keyword/%A5%B9%A5%E2%A1%BC%A5%EB%A5%D5%A5%A9%A5%EF%A1%BC%A5%C9))、PF([パワーフォワード](http://d.hatena.ne.jp/keyword/%A5%D1%A5%EF%A1%BC%A5%D5%A5%A9%A5%EF%A1%BC%A5%C9))、C(センター)。 [NBA](http://d.hatena.ne.jp/keyword/NBA)選手いうところの[シャキール・オニール](http://d.hatena.ne.jp/keyword/%A5%B7%A5%E3%A5%AD%A1%BC%A5%EB%A1%A6%A5%AA%A5%CB%A1%BC%A5%EB)や[レブロン](http://d.hatena.ne.jp/keyword/%A5%EC%A5%D6%A5%ED%A5%F3)・ジェームス。[スラムダンク](http://d.hatena.ne.jp/keyword/%A5%B9%A5%E9%A5%E0%A5%C0%A5%F3%A5%AF)でいうところの桜木、流川、ゴリです。比較的ゴール近辺でプレーし[インサイド](http://d.hatena.ne.jp/keyword/%A5%A4%A5%F3%A5%B5%A5%A4%A5%C9)で得点を稼ぐ選手です。

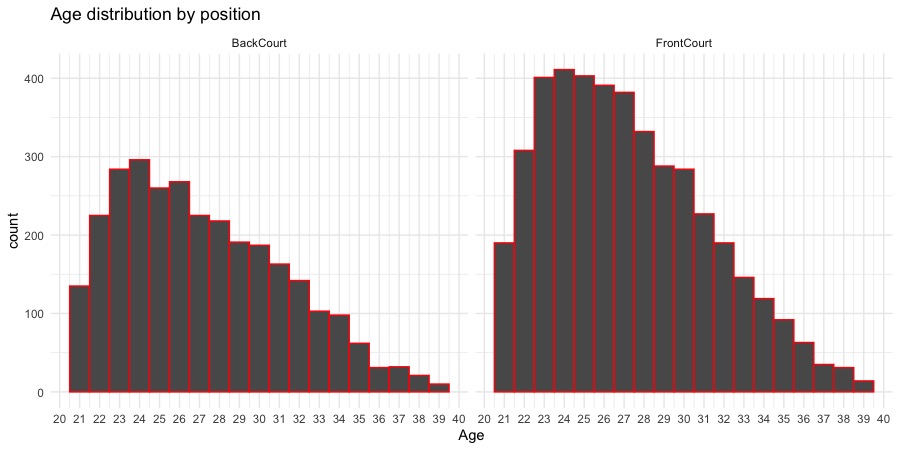
木暮くんはどっちかというとフロントコートポジションのはず。

(画像は[*こちら*](https://nekonekonoheya.com/archives/7218)からお借りしています。)

# 分析

## ポジションごとの年齢分布

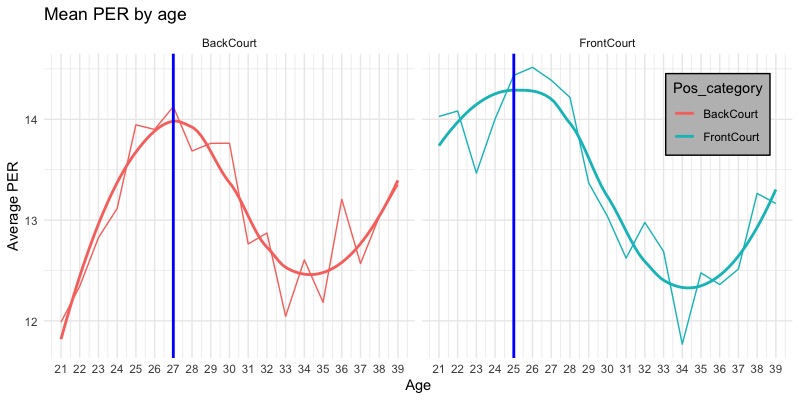
最初に確認するのはポジションごとの年齢の分布です。前述した通り、分析対象とした選手は20~40才の選手のみに絞られています。

ポジションごとの年齢分布

両者共に24才がもっとも観測された階級となります。また30才以後の選手の数が減少していることから多くの選手のキャリアが30代前半で終わっていることがわかります。

## 2変量分析：年齢とPER

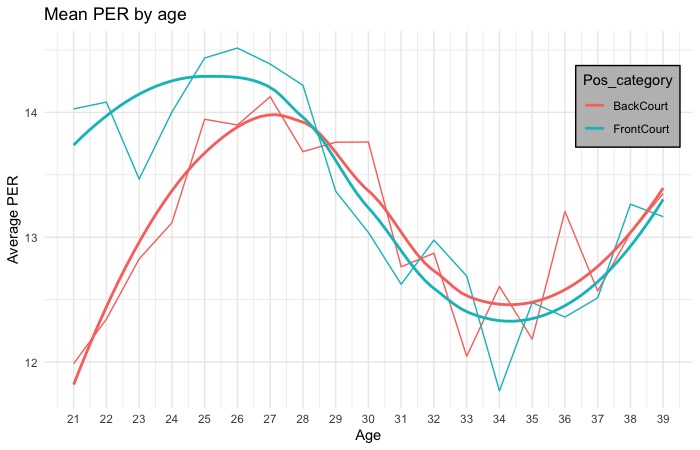
年齢とPERを[多項式](http://d.hatena.ne.jp/keyword/%C2%BF%B9%E0%BC%B0)回帰を用いたフィッティングを行いトレンドを確認してみます。

年齢別のPER平均

こちらのプロットでは[ヒストグラム](http://d.hatena.ne.jp/keyword/%A5%D2%A5%B9%A5%C8%A5%B0%A5%E9%A5%E0)で確認したような結果と逆の結果で見れました。 バックコートのピークが27才、フロントコートのピークが25才と読み取れます。またピークのレンジとしては、順に26~28才、24~27才あたりであると読み取れます。

また、どちらのポジションでも34~35才あたりまで減少が続き、そこから跳ね返るように増加していることもグラフから読み取れます。 自分の考察を述べると…

この底をつくあたりで平均的な選手は引退をを迎える一方で、All-Star級の選手は現役を続行し、(全盛期ほどではないものの)年齢の割りに良いパフォーマンスを継続するため、このような現象が見受けられると考えました。

ポジションごとのトレンド比較

2つ目のプロットではトレンドの比較を行いました。興味深い点は、PERがピークになる**若いうちはバックコートの方が良いパフォーマンスを発揮する**ものの、**キャリアの後半になれば、フロントコートの方がより良いパフォーマンスを発揮する傾向がある**ことです。

一概に述べることはできませんが、先にも記述した通り、バックコートポジションの選手にはゲームをコン[トロール](http://d.hatena.ne.jp/keyword/%A5%C8%A5%ED%A1%BC%A5%EB)するスキルが求められます。そのためベテラン選手でも十分にコートでの存在感を発揮することが可能です。対照的に、フロントコートポジションの選手にはよりフィジカル性の高い環境である[インサイド](http://d.hatena.ne.jp/keyword/%A5%A4%A5%F3%A5%B5%A5%A4%A5%C9)でプレーすることが求められます。そのためある程度の「若さ」という要因がパフォーマンス指標に現れ、このような歳をとるにつれて逆転してしまうという現象が確認できると考えました。

## 考察

非常に興味深い結果や示唆を得られたと思います。

発見としては、

* 選手全体的にパフォーマンスのピークは**25~27才の間**に起きる
* **バックコートの選手は27才**あたりにピークを迎える傾向がある
* **フロントコートの選手は25才**あたりにピークを迎える傾向がある
* パフォーマンスはどちらのポジションも**30才を目前に大きく減少**する傾向がある
* またPERの**減少度合い**はバックコートに比べるとフロントコートの選手がとても強い
* **若いうちはフロントコートの選手の方がPERが高いが、キャリアの後半になるにつれその現象は逆転する傾向にある**

図らずとも参照したサッカーデータの分析と近い値が得られました。

# まとめ

他の記事でも提示されているパフォーマンスがピークになる年齢を提示した上で、今回の記事では新しく、PERを用いた検証の可能性を示唆できたのではないかと思います。

# 問題点

* 選手の分類でフロントコートとバックコートの区分が少し大きすぎる
* 推定の行い方がナイーブすぎる

この二つが現状の問題点だと思います。他の論文でもしばしば見受けられる指摘としては、セレクションバイアスがかかっているという点です。キャリアを通じてパフォーマンスが良い選手はキャリアの長さが長く、そのためサンプルデータの最初と最後のあたりにバイアスがのしかかっているということです。

この部分を解消できる手段を探して実装することで次につなげていきたいと思います。

# バスケットのデータ分析に興味のある人へ

これまで自分が参考にしてきたサイトやソースなどを提示して記事を締め括りたいと思います。

* [Basketball Reference](https://www.basketball-reference.com/): バスケのデータが豊富にあり、まず初めにチェックすべきサイトです。
* [Kaggle Notebooks](https://www.kaggle.com/notebooks): バスケットに関する分析レポートが多く公開されています。おすすめはスペイン在住の[Xavier氏](https://www.kaggle.com/xvivancos/kernels)の[カーネル](http://d.hatena.ne.jp/keyword/%A5%AB%A1%BC%A5%CD%A5%EB)。
* [バスケのデータ分析](https://twitter.com/b__s__k__t): 日本人の方が運営されているバスケデータ分析に関するアカウントです。
* [NBAloveR](https://github.com/koki25ando/NBAloveR): 恐縮ですが自分が開発したRのバッケージです。現在利用は可能ですが、メンテナンスは一時停止中です。今後アップデートしていきたいです。