Applying Concepts From Performance Management: Using R to Work With Performance Data II

Patrick J. Ferguson

University of Melbourne

April 2021

Data Task: The Setting

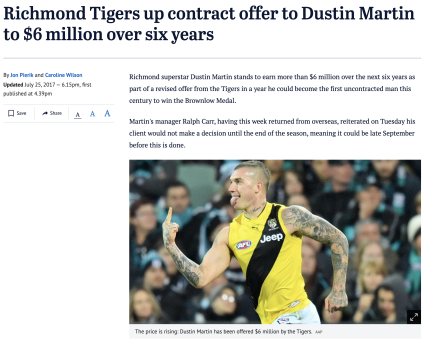
The Australian Football League is the national Australian-rules football competition. In 2019, total club earnings exceeded $1 billion.

- On the off chance you are unfamiliar with Australian-rules football, here is a quick primer on the sport: http://www.youtube.com/watch?v=XMZYZcoAcU0

Data Task: Performance Evaluation in the AFL

In the AFL, coaches/managers evaluate players after each game. These evaluations are used to determine which players get selected for upcoming games, and which players have their contracts renewed at the end of the season.

- In conducting these reviews, coaches consult a large number of measures ('stats') that capture each players' performance during a game.



Richmond Tigers up contract offer to Dustin Martin to $6 million over six years

By Jon Pierik and Caroline Wilson
Updated July 25, 2017 — 6.15pm, first published at 4.39pm

Save Share A A A

Richmond superstar Dustin Martin stands to earn more than $6 million over the next six years as part of a revised offer from the Tigers in a year he could become the first uncontracted man this century to win the Brownlow Medal.

Martin's manager Ralph Carr, having this week returned from overseas, reiterated on Tuesday his client would not make a decision until the end of the season, meaning it could be late September before this is done.

The price is rising: Dustin Martin has been offered $6 million by the Tigers. AAP

Data Task: The Data Set

You will be working with a simulation of the actual data that AFL coaches use to evaluate players.

- This data set is at the player-game level.
- 1,474 observations for nine different performance measures. 60+ unique players in games across 3 AFL seasons.

Game_ID	Player_ID	Position	GameTotalMins	GameTotalDistance_km	Disposals	Disposal_efficiency	Goals
1	2	DEFENCE	108.5	13.2552	17	58.8	0
1	5	DEFENCE	115.4	14.2282	21	66.7	0
1	6	MIDFIELD	90.1	12.2865	12	66.7	0
1	7	DEFENCE	109.8	13.1969	19	78.9	1
1	8	FORWARD	109.3	15.2498	14	78.6	3
1	13	DEFENCE	116.5	13.2598	15	86.7	0
1	14	MIDFIELD	107.8	14.2644	17	76.5	0
1	16	MIDFIELD	95.8	11.8750	21	71.4	3
1	21	DEFENCE	108.2	13.7973	11	81.8	0
1	22	MIDFIELD	91.8	12.0372	14	71.4	1

Data Task: Objective

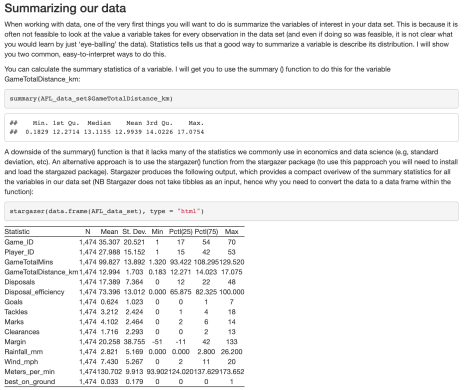
I will get you to explore and analyze this data set as a way to apply the concepts on performance management that you have covered during this course.

- How do we describe the properties of performance measures? What makes a 'good' performance measure? How can we determine if a measure is 'sensitive'? If it is 'noisy'?
- How do we measure performance in team settings? How do we filter out 'common shocks'?

Data Task: Preparation

I have broken this data task into two components: preparation material and a workshop.

- The preparation material is a self-directed exercise where I give you instructions, code, output, etc to work through. You will learn how to install R and how to run basic commands on our data set.
- I will also provide you with the data set itself and a data dictionary (a document that defines the variables in the data set).



Summarizing our data

When working with data, one of the very first things you will want to do is summarize the variables of interest in your data set. This is because it is often not feasible to look at the value a variable takes for every observation in the data set (and even if doing so was feasible, it is not clear what you would learn by just 'eye-balling' the data). Statistics tells us that a good way to summarize a variable is describe its distribution. I will show you two common, easy-to-interpret ways to do this.

You can calculate the summary statistics of a variable. I will get you to use the summary () function to do this for the variable GameTotalDistance_km:

```
summary(AFL_data_set$GameTotalDistance_km)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1829 12.2714 13.1155 12.9939 14.0226 17.0754
```

A downside of the summary() function is that it lacks many of the statistics we commonly use in economics and data science (e.g, standard deviation, etc). An alternative approach is to use the stargazer() function from the stargazer package (to use this papproach you will need to install and load the stargazed package). Stargazer produces the following output, which provides a compact overivew of the summary statistics for all the variables in our data set (NB Stargazer does not take tibbles as an input, hence why you need to convert the data to a data frame within the function):

```
stargazer(data.frame(AFL_data_set), type = "html")
```

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Game_ID	1,474	35.307	20.521	1	17	54	70
Player_ID	1,474	27.988	15.152	1	15	42	53
GameTotalMins	1,474	99.827	13.892	1.320	93.422	108.295	129.520
GameTotalDistance_km	1,474	12.994	1.703	0.183	12.271	14.023	17.075
Disposals	1,474	17.389	7.364	0	12	22	48
Disposal_efficiency	1,474	73.396	13.012	0.000	65.675	82.325	100.000
Goals	1,474	0.624	1.023	0	0	1	7
Tackles	1,474	3.212	2.424	0	1	4	18
Marks	1,474	4.102	2.464	0	2	6	14
Clearances	1,474	1.716	2.293	0	0	2	13
Margin	1,474	20.258	38.755	-51	-11	42	133
Rainfall_mm	1,474	2.821	5.169	0.000	0.000	2.800	26.200
Wind_mph	1,474	7.430	5.267	0	2	11	20
Meters_per_min	1,474	130.702	9.913	93.902	124.020	137.629	173.652
best_on_ground	1,474	0.033	0.179	0	0	0	1

Data Task: Workshop

I have broken this data task into two components: preparation material and a workshop.

- The workshop is a live coding session in RStudio where I will lead you through a set of questions on performance management in the context of our data set/institutional setting.
- I will share my screen with you over Zoom. I will write and run code that you will need to replicate on your own machine in RStudio. Prior to class, I will also circulate a handout that contains a set of questions that will guide the session.