

The goal of a basketball team is similar to that of a baseball team, making the playoffs.

So how many games does a team need to win in order to make the playoffs?

Recall that in the lecture we found this number by looking at a graph.

Here in R, let's use the table command to figure this out for the NBA.

So that's just `table(NBA$W, NBA$Playoffs)`.

So our table pops up.

Let's scroll to the top so we see what's going on.

OK so as we typed in, we've got the number of wins here as the rows.

And 0 if a team didn't make the playoffs, 1 if a team did make the playoffs in the columns.

So for all of our data, for example, consider all the times that a team won 17 games.

So this happened 11 times in total.

And all 11 times the teams didn't make it to the playoffs when they won 17 games.

Let's scroll down and look at a much higher number for contrast.

For example, 61 wins.

If a team 61 games then 10 of those times they made it to the playoffs, and 0 times they didn't.

So it seems like if you win 61 games you are definitely going to make it to the playoffs.

But I'm sure we can find a much better threshold.

Let's take a look at the table, say around the middle section.

OK, so here we can see that a team who wins say about 35 games or fewer almost never makes it to the playoffs.

We see a lot of 0s and 1s in this column up until 35.

After 35 we start seeing some numbers over here.

So teams are starting to make it to the playoffs.

And if we scroll down, we see that after about 45 wins, teams almost always make it to the playoffs.

We see very few 1s and 0s in the category of not making it.

So it seems like a good goal would be to try to win about 42 games.

If a team can win about 42 games then they have a very good chance of making it to the playoffs.

So in basketball, games are won by scoring more points than the other team.

Can we use the difference between points scored and points allowed throughout the regular season in order to predict the number of games that a team will win?

Let's give it a try.

First we add a variable that is the difference between points scored and points allowed.

Let's call this `NBA$PTSdiff`.

And that's just the difference between points scored, which is points, and points allowed, which is opponent's points.

All right, so we've created a variable.

Let's first make a scatter plot to see if it looks like there's a linear relationship between the number of wins that a team wins and the point difference.

So this is easy to do just with the `Plot` command.

`NBA$PTSdiff` and `NBA$W`.

So our graph pops up and it looks like there's an incredibly strong linear relationship between these two variables.

So it seems like linear regression is going to be a good way to predict how many wins a team will have given the point difference.

Let's try to verify this.

So we're going to have points diff as our independent variable in our regression, and `W` for wins as the dependent variable.

So let's call this WinsReg.

And we just use the lm command as before progressing w on the points diff and using the NBA data.

All right, so we've created our regression.

Let's take a look at the summary.

OK, so the first thing that we notice is that we've got very significant variables over here.

And an R squared of 0.9423, which is very high.

And this is verifying the scatter plot we saw before that there's a very strong linear relationship between the wins and the points difference.

So let's write down the regression equation that we found.

We see that the number of wins,  $W$ , is equal to 41.

That's coming from the coefficient estimate for the intercept.

Plus  $0.0326 \cdot \text{PTSdiff}$ .

And that 0.0326 is coming from the coefficient estimate for points difference.

So we saw earlier with the table that a team would want to win about at least 42 games in order to have a good chance of making it to the playoffs.

So what does this mean in terms of their points difference?

Well, we can calculate it.

If we want this to be greater than or equal to 42, that means that the points difference would need to be greater than or equal to 42 minus 41 divided by 0.0326.

So if we actually do that calculation, we see that this is equal to 30.67.

So we need to score at least 31 more points than we allow in order to win at least 42 games.