

Tarea

Regresión Lineal

Juan Sebastian Velasquez Acevedo - 1744936

Programa: 3743

Introducción	1
Regresión Lineal simple 1 - Posesión - Oportunidades de gol	1
Regresión Lineal simple 2 - Cantidad de pases - Posesión	3
Regresión Lineal Múltiple - Tiros Libres - Posesión - Oportunidades de gol	5
Repositorio de Github con código R	6

Introducción

Para esta tarea se decidió tomar un dataset de Kaggle que contiene datos públicos de la copa del mundo FIFA 2018.

Link: <https://www.kaggle.com/mathan/fifa-2018-match-statistics>

Regresión Lineal simple 1 - Posesión - Oportunidades de gol

La cuestión que en este caso queremos analizar es si entre más **posesión de balón** de un equipo, se realizaron **más oportunidades de gol**.

Call:

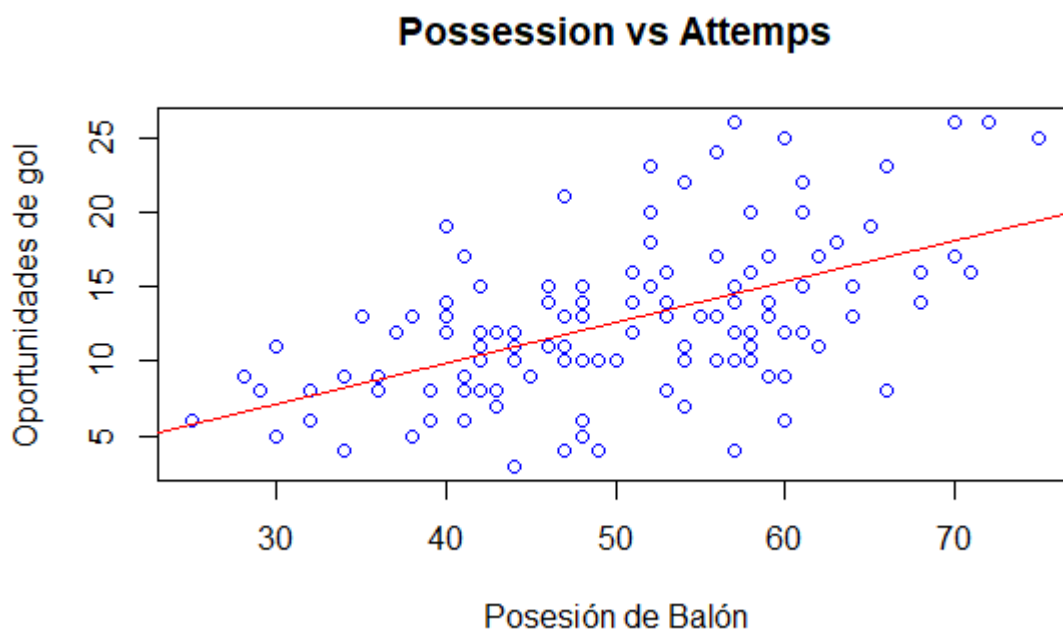
```
lm(formula = data$Attempts ~ data$Ball.Possession..)
```

Coefficients:

```
(Intercept) data$Ball.Possession..  
-0.9954      0.2718
```

Al hacer el modelo lineal, nos dieron los datos anteriores: Lo que indica un comportamiento lineal $Y=0.2718X - 0.9954$ teniendo el valor 0.2718 como pendiente de la línea de regresión β_1 y el valor -0.9954 que es el intercepto de la línea con el eje Y el cual me representa β_0 . Para así poder cumplir con la fórmula: $Y = \beta_0 + \beta_1X + \epsilon$. El valor positivo de β_1 me indica que la posesión del balón contribuye a la creación de oportunidades de gol.

El comportamiento lo podemos ver en el siguiente gráfico:



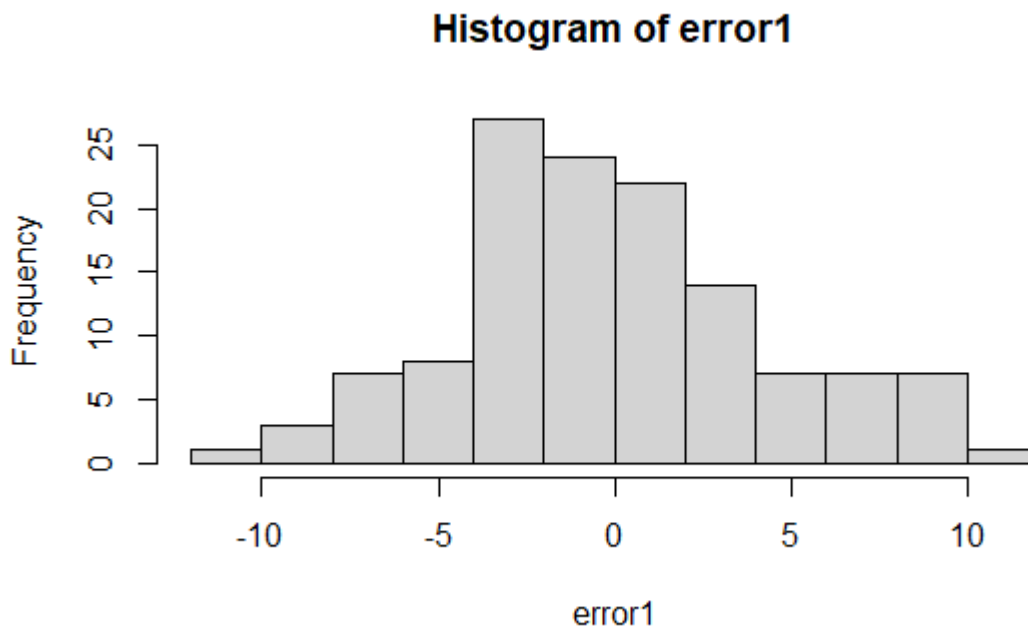
A partir del gráfico podemos concluir que a mayor posesión de balón, más oportunidades de gol son creadas por el equipo. Estos datos pueden ser muy relevantes para los analistas del fútbol, entrenadores y cuerpo técnico para así establecer estrategias de posesión de balón en sus partidos.

También podemos analizar el coeficiente de determinación r^2 que como sabemos, cuanto más cerca esté r^2 de 1 mejor será el ajuste del modelo lineal y entre más cerca esté r^2 de 0 peor será el ajuste del modelo lineal. Para esto se usa la función `summary` en R que me arroja lo siguiente:

```
r^2 = summary(modLinealSimple)$r.square = 0.2928813
```

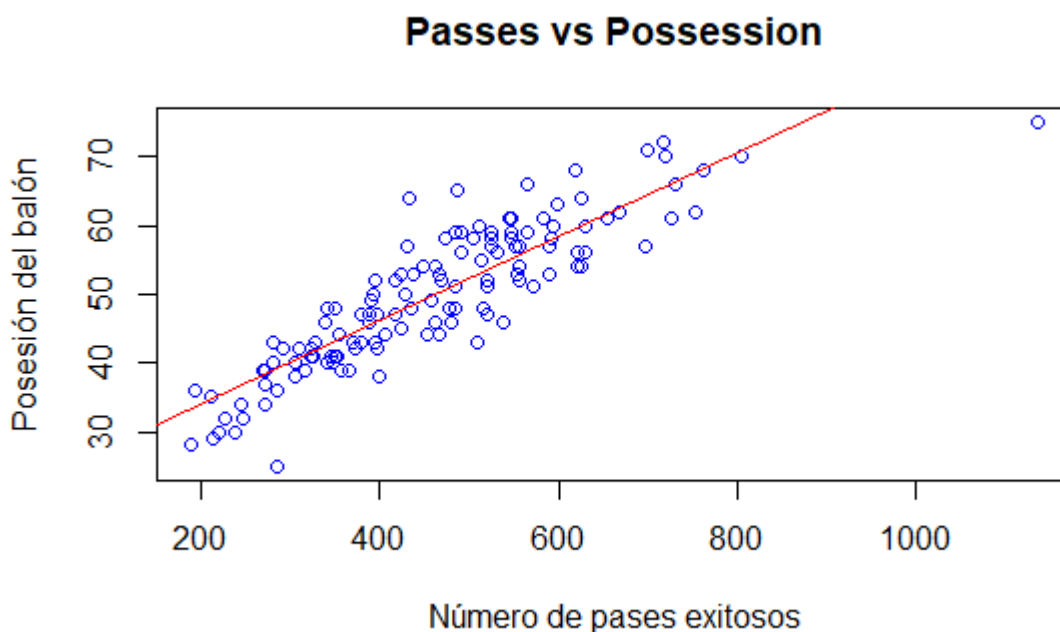
En este caso no tenemos el mejor ajuste porque los datos no me permiten verificar una relación completamente lineal como para decir que el modelo lineal simple funciona al 100%.

Si analizamos los errores residuales, vemos que se comporta como una distribución normal:



Regresión Lineal simple 2 - Cantidad de pases - Posesión

Entonces, siguiendo con lo que desearía un entrenador ¿Cómo se puede lograr mayor posesión? Pues trabajando en la parte técnica del pase, porque aunque parece obvio, podemos concluir también que entre más **cantidad de pases exitosos por partido** se dan, más **posesión del balón** se tiene. Como lo observamos en el siguiente gráfico:



La ecuación de la línea recta es:

Call:

```
lm(formula = data$Ball.Possession.. ~ data$Passes)
```

Coefficients:

```
(Intercept) data$Passes
```

```
21.84775    0.06083
```

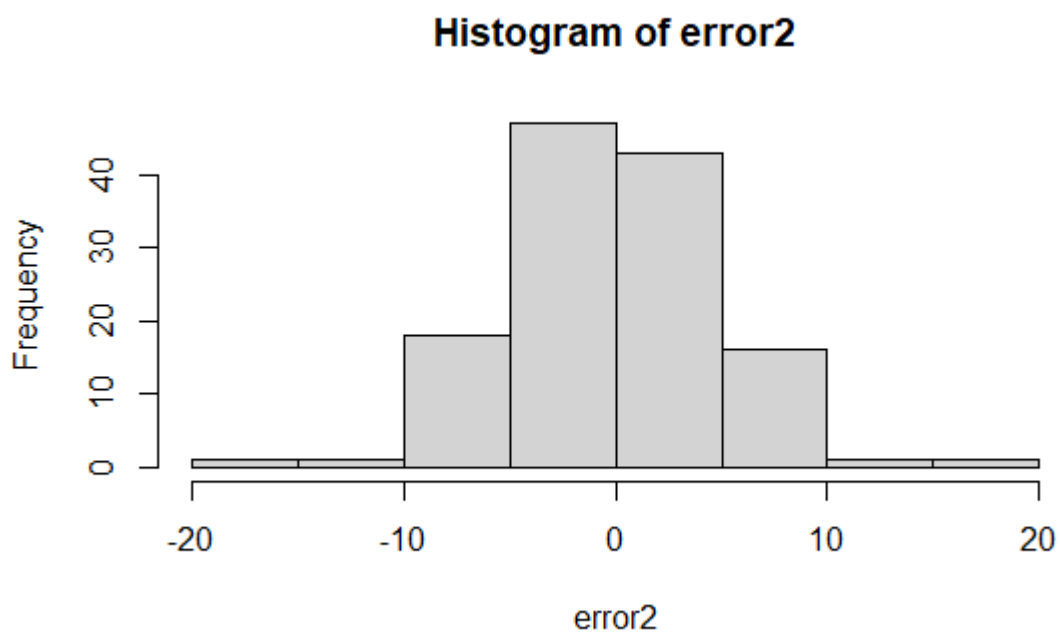
$Y = 0.06083X + 21.84775$

Para el anterior modelo también se decide analizar el coeficiente de determinación, que en este caso es:

```
r^2 = summary(modLinealSimple2)$r.square = 0.7754755
```

En este caso tenemos un mejor ya que r^2 está más cerca de 1, entonces la calidad del modelo para replicar los resultados es buena.

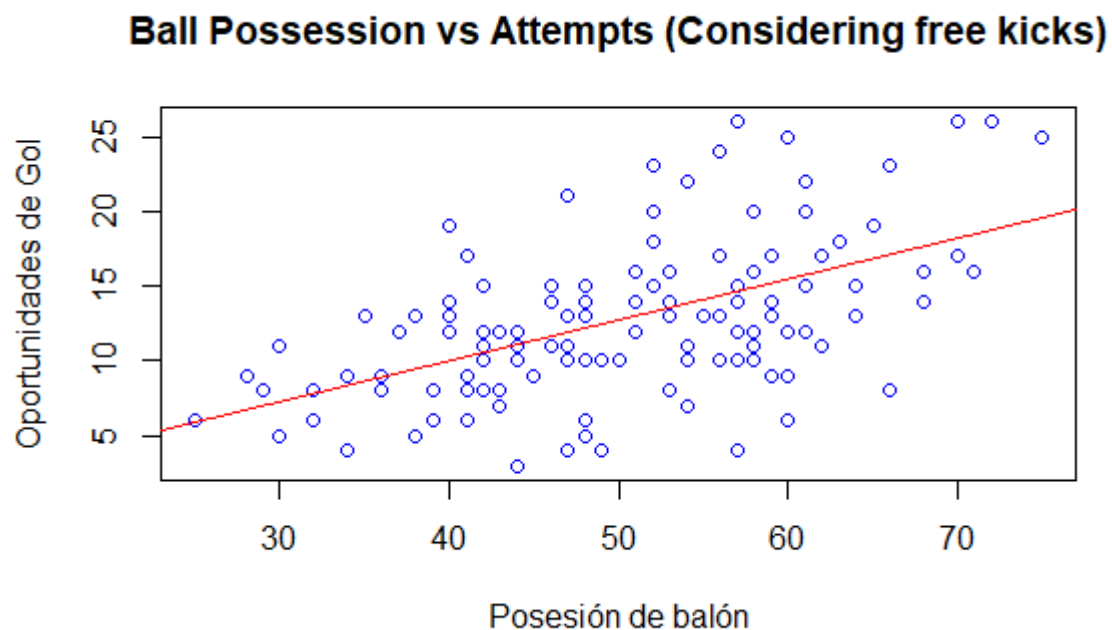
Si analizamos los errores residuales, vemos que se comporta como una distribución normal:



Regresión Lineal Múltiple - Tiros Libres - Posesión - Oportunidades de gol

Queremos ver a través de esta regresión múltiple si la posesión y los tiros libres, contribuyen a que haya más oportunidades de gol.

Vemos que el gráfico no cambia mucho si lo comparamos con la [regresión 1](#):



Si analizamos el modelo lineal para este caso tenemos que:

Call:

```
lm(formula = data$Attempts ~ data$Ball.Possession.. + data$Free.Kicks)
```

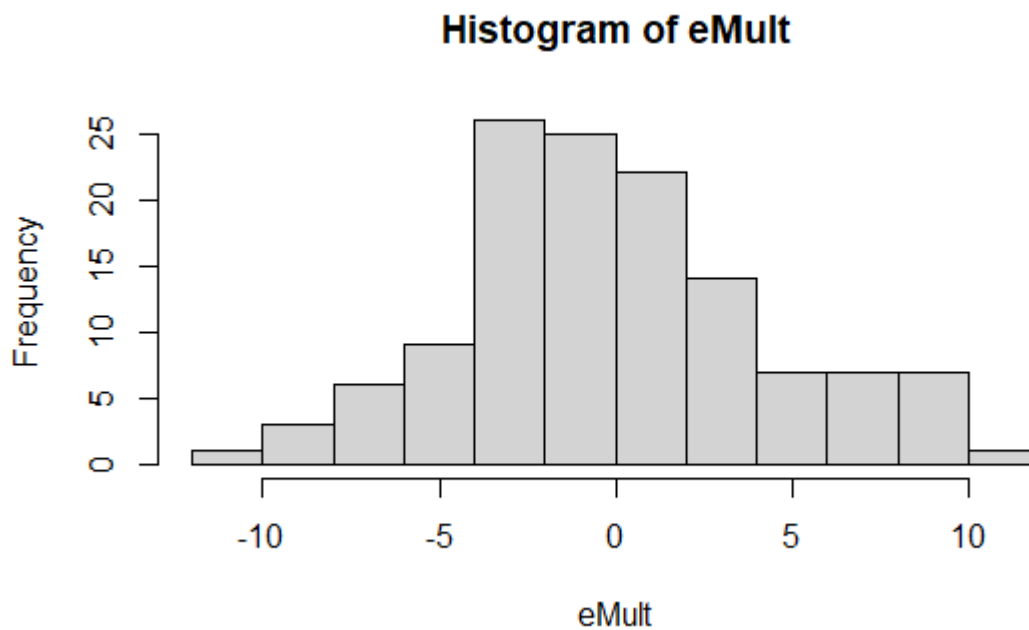
Coefficients:

(Intercept)	data\$Ball.Possession..	data\$Free.Kicks
-0.918723	0.272917	-0.008814

Esto indica que la ecuación de la recta en este caso es: $Y = 0.272917 X_1 - 0.008814 X_2 - 0.918723$. Lo que me indica que la posesión ayuda a crear más oportunidades de gol, y los tiros libres no siempre ayudan a crear más oportunidades de gol, ya

que vemos que este valor es negativo, pero al ser tan pequeño es casi nulo el aporte.

Si analizamos los errores residuales, vemos que se comporta como una distribución normal:



Repositorio de Github con código R

https://github.com/Odzen/Probability_Statistics/tree/main/RegresionLineal/Tarea