MLR Lab Report I
Xinyi (Anny) Cui

Dear Sir,

By collating and analyzing physical data of 13 players from the study "The relationship between selected physical performance variables and football punting ability" by the Department of Health, Physical Education and Recreation at the Virginia Polytechnic Institute and State University, 1983. I build a model that estimates the length of a punter's punting distance by looking at how his ball stagnated in the air and his overall leg strength.

Below is specific information about this multiple linear regression model:

```
Call:
lm(formula = Distance ~ Hang + O_Strength, data = footballData)

Residuals:
    Min      1Q  Median      3Q     Max
-25.268  -8.090   2.622   8.091  19.898

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.3004    35.8019  -0.036   0.9717
Hang         26.8985    12.9852   2.071   0.0651 .
O_Strength    0.2246     0.1328   1.692   0.1216
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.4 on 10 degrees of freedom
Multiple R-squared:  0.7438,    Adjusted R-squared:  0.6926
F-statistic: 14.52 on 2 and 10 DF,  p-value: 0.001103
```

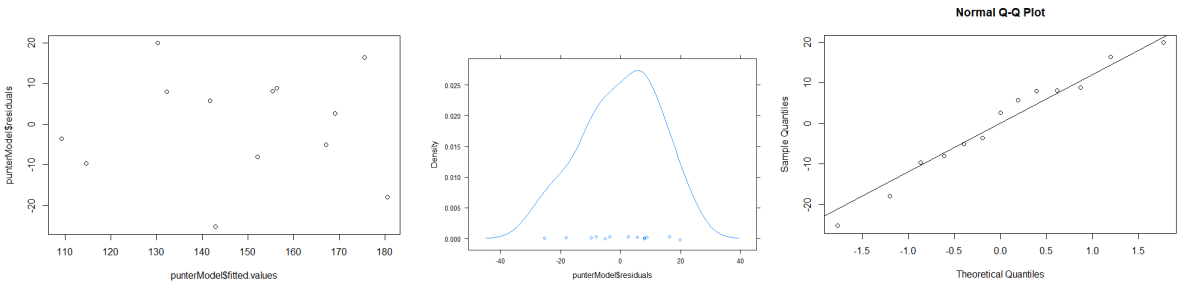
From the observation, the formula is

$$Distance = -1.3 + 26.9(Hang) + 0.2(O\_Strength)$$

This model means that Hang time will be a very important explanatory variable, and the distance of the ball may increase by about 27 feet for each additional second the ball stays in the air. And the overall leg strength of the punter is also important, with every extra pound of leg strength increasing the likelihood of 0.2 feet in ball distance.

The test of the model can prove the rationality of the model. The value of p-value is only 0.001103, is less than 0.05. This proves that there is an obvious linear relationship between the response and explanatory variables. At the same time, the value of F-statistic of 14.52 also indicates that the model is better than the constant one. When looking at the value of R^2, the coefficient of determination, 0.7438 means that 74.38% of the variation in distance can be explained by the variables Hang time and Overall leg strength. This is a very good percentage.

To further confirm the linearity of the model, I construct some graphs.



The graph of the fitted values v.s. residuals seems displays a zero mean value scatter plot. The density plot of the residuals shows a good normal distribution, and the quantile graph gives a good linear scatter plot. Those graphs verify the linearity of the multiple linear regression model.

When watching high school students, I focus on their punting time and overall leg strength. I also predict the punting distance of an average punter and the pitching distance of a star player I'm looking for. When watching high school students, I focus on their punting time and overall leg strength. I also predict the punting distance of an average punter and the pitching distance of a star player I'm looking for. When watching high school students, I focus on their punting time and overall leg strength. I also predict the punting distance of an average punter and the pitching distance of a star player I'm looking for.

```
> averageKicker
       fit      lwr      upr
1 148.2331 114.9397 181.5265
> star
       fit      lwr      upr
1 200.5827 160.8698 240.2956
```

The average punter has a 3.9 seconds of hang time and his overall leg strength is 196 pounds, which predicting his final punting distance is around 148 ponds with no exceeding 182 feet. However, I'm looking for a star punter whose hang time can be 5 seconds and the overall leg strength is 300 pounds, he can punt about 200 feet and the highest possible distance he can do is 240 feet, which is wonderful!

Best,
Anny