1. The scatterplots below show the HPI of each player against clubs’ mean penalties and offense. In comparing the two plots, what do you expect the signs of $β\_{1}$ and $β\_{2}$ to be in the model: $\hat{HPI} = β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_offense+ε ?$



Given that the regression line in the HPI against penalties has an even slope, I expect the coefficient for total\_penalties to be zero. Since the regression line in the HPI against total\_offense plot has a positive slope, I expect total\_offense to have a positive coefficient.

1. Using the values provided in this summary table, interpret $β\_{1}$ and $β\_{2}$ in the context of *HPI* for the model: $\hat{HPI}=β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_penalties+ε.$



For every additional offensive play, a player’s HPI will increase by 0.017539, provided their total penalties stay constant.

For every additional penalty, a player’s HPI will decrease by 0.071746, provided their total offensive plays remain constant.

1. Using calculations, fill in the below ANOVA table and perform an ANOVA test to assess the overall fit of the model: $\hat{HPI}=β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_penalties+ε$. The dataset has 309 players.

**H0:**  $β\_{1}=β\_{2}=0$

**Ha:** $β\_{1 } OR β\_{2}\ne 0$

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **d.f.** | **Sum of Squares** | **Mean Square** | **F** | **P-value** |
| **Model** | *k =* **2** | **668.51** | *SSModel/k=*668.51/2=**334.26** | *MSModel/MSE*=334.26/7.42=**45.05** | **0.00** |
| **Residual** | *n-k-1=*309-2-1= **306** | **2270.72** | *SSE/(n-k-1)=*2270.72/306=**7.42** |
| **Total** | *n-1 =* 309-1 **= 308** | *SSModel + SSError =* 668.51 + 2270.72 = **2939.23** |

**Conclusion:**

p-value < 0.05

Reject H0

We have significant evidence that at least one of total\_offense or total\_penalties is effective predictors of HPI in handball.

1. The scatterplot below shows *total\_penalties* against *total\_offense* with a regression line. Based on this plot do you expect a strong correlation between *total\_penalties* and *total\_offense*, will it be positive or negative?



Given that the regression line shows total\_penalties increasing with total\_offense, I expect them to have a strong positive correlation

1. Using a correlation of 0.734, test the significance of the correlation between the *total\_offense* and the *total\_penalties* of a player. Provide an interpretation of the results. The dataset has 309 players.

**H0:** $ρ=0$ **Ha:** $ρ \ne 0$ $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^{2}}} =\frac{0.73\sqrt{309-2}}{\sqrt{1-0.73^{2}}}=18.71$

Compare on t distribution 307 df.

P-value= **0**

**Conclusion:**p-value < 0.05

Reject H0

We have significant evidence of a strong positive correlation between total\_penalties and total\_offense, meaning they increase together.

1. Could it be concluded that having more penalties impacts the success of a player in the form of HPI?

Having more penalties in some ways decreases a players success as they can detract from the playing time of a player. However, it seems that being a more aggressive player or a player with more penalties, tends to leads towards players being more offensively aggressive as well which does improve their success. Additionally if players play in more games they are likely to have more penalties and an overall higher HPI which could also be a reason for this trend.