Handball is a popular sport in many European countries such as Germany where it is said to have originated. The Bundesliga, for example, is a German men's professional handball league. Handball is typically played indoors on a rectangular court (20m x 40m). There are two goals (3m x 2m) on opposite sides of the court, the goal for each team is to score a goal

by getting the ball in the other team's goal. The challenge to this is that there is a semicircle with a radius of 6m surrounding the goal which players other than the goal keeper are not allowed in, making it challenging to score. Players run back and forth down the court passing the ball to each other and trying to score.

There are seven positions total in handball: the goalkeeper who defends the team's goal, left and rights backs are positioned on the left and right side of their half of the court to provide further defense, the center can move up and down the court and is usually the one trying to score, left and right wings can also move up and down the court, serving as offense when the team pushes for attack and defense when the opposing team tries to score, lastly, the pivot is considered strictly an offensive player as they are usually position in the opposing side of the court, they often work closely with the center. The Bundesliga regular season length is 34 games, with playoffs the maximum number of games a team could play is 41. Players do not play every game in the season and subbing is common,

generally speaking playing time goes to players with experience. This can create a bit of a disparity in which players with less playing time will not necessarily have statistics that accurately display their skills due to the smaller sample size.

Handball is considered a contact sport which means aggressive strategies are often used in games. Aggressiveness can be measured in the penalty statistics as players who tend to get more penalties are usually more considered more aggressive players overall. Their success can be measured with the handball performance index (HPI)(<https://www.liquimoly-hbl.de/en/s/handball-performance-index/2021-22/handball-performance-index--data-based--transparent--fair/>), a calculated statistic which essentially ranks how good a player is. The [handball\_bundesliga\_23.csv](handball_budnesliga_23.csv) data set could provide insight on if players or teams that are more aggressive as measured by penalties are more successful than those who are more passive.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| NAME | The name of the player. |
| CLUB | The club the player is on.  |
| POSITION | The position of the player.  |
| GP | The number of games the player was in that season. |
| total\_offense | The total offensive plays made by the player in the season. Calculated by adding the 6 offensive focused statistics from the original dataset. |
| total\_penalties | The total penalties the player had in the season. Calculated by adding the 5 penalty related statistics from the original dataset. |
| HPI | Handball performance index, complex formulaic calculation equivalent to how well the player performed in the season. Players with HPIs in the 70s are considered good, while players in the 60s are considered not as strong. (<https://www.liquimoly-hbl.de/en/s/handball-performance-index/2021-22/handball-performance-index--data-based--transparent--fair/>) |

1. Make a histogram of *total\_penalties* and describe the distribution.
	1. Fit the model: $\hat{HPI} = β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_offense+ε $
	2. Interpret $β\_{1}$ and $β\_{2}$ in the context of *HPI*.
	3. Create a scatterplot with a regression line, of *HPI* for against total\_penalties. Comment on the slope.
	4. Create a scatterplot a regression line of *HPI* against total\_offense. Comment on the slope.
2. Find and interpret a 98% confidence interval for the mean *HPI* of players with 30 *total\_penalties*.
3. Find and interpret a 98% prediction interval for the *HPI* of a player with 30 *total\_penalties*.
4. Using R, perform an ANOVA test to assess the overall fit of the model:

 $\hat{HPI} = β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_offense+ε $. Fill in the ANOVA table below and interpret the results.

**H0: Ha:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **d.f.** | **Sum of Squares** | **Mean Square** | **F** | **P-value** |
| **Model** |  |  |  |  |  |
| **Residual** |  |  |  |
| **Total** |  |  |

1. Total penalties vs. total offense
	1. Create a scatterplot of *total\_penalties* against *total\_offense* with a regression line.
	2. Based on the plot do you expect a strong correlation between *total\_penalties* and *total\_offense*, will it be positive or negative?
2. Find the correlation of *total\_penalties* and *total\_offense*.
3. Test the significance of the correlation between the *total\_offense* and the *total\_penalties* of a player. Provide an interpretation of the results.

**H0: Ha:**

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