1. Using R find and interpret a 98% confidence interval for the mean HPI of a player with 30 total\_penalties.

```{r}

mod <- lm(HPI ~ total\_penalties, data=handball\_df)

newx <- data.frame(total\_penalties=30)

predict(mod, newx,interval="confidence", level = 0.98)

```

**(69.11, 69.97)**

We are 98% confident that the mean HPI for all players with 30 total penalties for the season is between 69.11 and 69.97.

1. Using R find and interpret a 98% prediction interval for the HPI of a player with 30 total\_penalties.

```{r}

mod <- lm(HPI ~ total\_penalties, data=handball\_df)

newx <- data.frame(total\_penalties=30)

predict(mod, newx,interval="prediction", level = 0.98)

```

**(62.3, 76.78)**

We are 98% confident that a player with 30 total penalties for the season will have an HPI between 62.3 and 76.78.

* 1. Use R to create a scatterplot of the mean HPI for clubs against the mean club penalties. Add a regression line to the plot.

A graph with a line and a blue line

Description automatically generated

```{r}

mod\_plot <- handball\_clean |> group\_by(CLUB) |> summarise(club\_offense = mean(total\_offense), club\_pen = mean(total\_penalties), club\_hpi = mean(HPI))

ggplot(data = mod\_plot, aes(x = club\_pen, y = club\_hpi)) +

geom\_point() +

geom\_smooth(method = "lm") + theme\_minimal() + labs(title = "Scatterplot of club\_hpi against club\_offense with regression line")

```

* 1. Use R to create a scatterplot of the mean HPI for clubs against the mean club offensive plays. Add a regression line to the plot.

A graph with a line and dots

Description automatically generated```{r}

mod\_plot <- handball\_clean |> group\_by(CLUB) |> summarise(club\_offense = mean(total\_offense), club\_pen = mean(total\_penalties), club\_hpi = mean(HPI))

ggplot(data = mod\_plot, aes(x = club\_offense, y = club\_hpi)) +

geom\_point() +

geom\_smooth(method = "lm") + theme\_minimal() + labs(title = "Scatterplot of club\_hpi against club\_offense with regression line")

```

* 1. In comparing the two plots, what do you expect from the model:

?

Given that the regression line in the club\_hpi against club\_penalties have a negative slope, I expect the coefficient for club\_penalties to be negative. Since the regression line in the club\_hpi against club\_offense plot has a positive slope, I expect club\_offense to have a positive coefficient.

* 1. Use R to fit the model:

A close-up of numbers

Description automatically generated

```{r}

hpi\_mod <- lm(HPI~total\_offense+total\_penalties, data=handball\_clean)

summary(hpi\_mod)

```

* 1. Interpret and in the context of HPI.

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 R, perform an ANOVA test to assess the overall fit of . Fill in the ANOVA table below and interpret the results.

**H0:**  **Ha:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **d.f.** | **Sum of Squares** | **Mean Square** | **F** | **P-value** |
| **total\_offense** | **1** | **358.92** | **358.92** | **48.367** | **2.146e-11** |
| **total\_penalties** | **1** | **309.59** | **309.59** | **41.720** | **4.136e-10** |
| **Residual** | **306** | **2270.72** | **7.42** |
| **Total** | ***n-1 =* 309-1 = 308** | ***SSModel + SSError =* 668.51 + 2270.72 = 2939.23** |

**Conclusion:** Reject H0

p-value < 0.05, p-value < 0.05

We have significant evidence that both total\_offense and total\_penalties are effective predictors of HPI in handball.

```{r}

anovamod <- lm(HPI~total\_offense+total\_penalties, data=handball\_df)

summary(anovamod)

anova(anovamod)

```

6.

1. Use R to create a scatterplot of total\_penalties against total\_offense with a regression line. Color the points by POSITION.

A graph with a line and a line

Description automatically generated

```{r}

ggplot(data = handball\_clean, aes(x = total\_offense, y = total\_penalties)) + geom\_point(alpha = 0.5, aes(colour = POSITION)) + theme\_minimal() + geom\_smooth(method = "lm") + labs(title = "Scatterplot of total\_penalties against total\_offense with a regression line")

```

1. Based on the plot what do you expect the correlation between total\_penalties and total\_offense to be?

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

1. Using R, find the correlation of total\_penalties and total\_offense.

**r = 0.7341583**

```{r}

cor(handball\_df$total\_penalties, handball\_df$total\_offense)

```

1. Using R, 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:**

P-value= **< 2.2e-16**

**Conclusion:**

r = 0.7341583p-value < 0.05

Reject H0

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

```{r}

cor.test(handball\_clean$total\_offense, handball\_clean$total\_penalties)

```

1. Could it be concluded that having more penalties increases the skill and 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.