1. Make a histogram of *total\_penalties* and describe the distribution.

```{r}

hist(handball\_clean$total\_penalties)

```



The distribution is heavily right skewed meaning its more common to have a lower amount of total\_penalties.

* 1. Fit the model: $\hat{HPI} = β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_offense+ε $

```{r}

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

summary(hpi\_mod)

```



* 1. Interpret $β\_{1}$ and $β\_{2}$ 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. 1. Create a scatterplot with a regression line, of *HPI* for against total\_penalties. Comment on the slope.

```{r}

plot(handball\_clean$HPI ~ handball\_clean$total\_penalties)

abline(hpi\_mod)

```

The plot has a slightly positive slope. Note that the slope is positive even though the coefficient in the fitted model is negative, this is likely because the two variables are both heavily dependent on playing time

* 1.  Create a scatterplot a regression line of *HPI* against total\_offense. Comment on the slope.

```{r}

plot(handball\_clean$HPI ~ handball\_clean$total\_offense)

abline(hpi\_mod)

```

The plot has a steep positive slope.

1. Find and interpret a 98% confidence interval for the mean *HPI* of players 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. 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. Using R, perform an ANOVA test to assess the overall fit of the model:

 $\hat{HPI} = β\_{0}+β\_{1}total\\_offense+β\_{2} total\\_offense+ε.$

**H0:**  $β\_{1}=β\_{2}=0 $**Ha:** $β\_{1 } OR β\_{2}\ne 0$

```{r}

summary(hpi\_mod)

```



**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.

1. Total penalties vs. total offense
	1. Create a scatterplot of *total\_penalties* against *total\_offense* with a regression line.

```{r}

mod <- lm(total\_penalties ~ total\_offense, data = handball\_clean)

plot(handball\_clean$total\_penalties ~ handball\_clean$total\_offense)

abline(mod)

```

* 1. Based on the 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. Find the correlation of *total\_penalties* and *total\_offense*.

```{r}

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

```

**r = 0.7341583**

1. Test the significance of the correlation between the *total\_offense* and the *total\_penalties* of a player. Provide an interpretation of the results.

**H0:** $ρ=0$ **Ha:** $ρ \ne 0$

```{r}

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

```

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.

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.