Division III Women’s Soccer 2022 Data: Correlations KEY

1. Before calculating the correlation value between home\_score and away\_score, do you think that the value will be positive or negative? Explain.

A correlation value that is positive in this case means that a home team scores a lot of goals, and the away team scores a lot of goals as well.

A correlation value that is negative means that either the home team or the away team scores a lot of goals, while the other team does not.

In soccer, we typically do not see high scoring games from both the home team and the away team. This means that we expect a negative correlation value between home\_score and away\_score.

1. Calculate the correlation value for home\_score vs. away\_team. Interpret the correlation value in context.

cor(d3\_wsoc2022$away\_score, d3\_wsoc2022$home\_score) = -0.3108331

The correlation between away\_score and home\_score is -0.311. This means that one team either the home team or the away team, scores a lot of goals while the other team does not. This correlation value makes sense in the context of soccer.

1. Perform a test of association to see if the correlation value that you calculated above is statistically significant. Make sure to state the hypotheses clearly and follow the appropriate steps.

Ho: 𝜌=0

Ha: 𝜌≠0

r = -0.311

t = = = -20.029

In R:

cor.test(d3\_wsoc2022$away\_score, d3\_wsoc2022$home\_score)

Pearson's product-moment correlation

data: d3\_wsoc2022$away\_score and d3\_wsoc2022$home\_score

t = -20.029, df = 3751, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.3394524 -0.2816389

sample estimates:

cor

-0.3108331

p-value = ~0

Conclusion: We reject Ho. This means that we have convincing evidence that there is an association between home\_score and away\_score.

1. Plot home\_score vs. away\_score as a scatterplot. Does there appear to be a relationship between the two variables? If so, describe it. If not, explain the problem.

plot(d3\_wsoc2022$away\_score, d3\_wsoc2022$home\_score)

A graph of a number

Description automatically generated

This scatterplot shows that there seems to be a concern about normality. In a Pearson correlation, both variables are assumed to be normally distributed.