**STAT2300 Group Project – Example Proposal**

**Subject area**: In the sport of basketball, two teams of five players each compete to score the most points. During the game, a personal foul may be called by a referee if illegal contact occurs. Personal fouls are generally to be avoided, as there are a range of penalties that can happen, including free throws by the opposing team or disqualification of players who commit too many personal fouls during a game. Consequently, teams generally wish to avoid committing personal fouls. However, there are certain situations in the game where personal fouls can be strategically employed by teams for a variety of reasons. One such example is when a team is trailing by a small margin of points near the end of the game. Committing a personal foul may result in the opposing team scoring points via free throws, but it increases the number of chances the team has to make up the deficit.

In this project, we wish to explore the distribution of fouls throughout the early part of the game for randomly sampled data from the NBA, and compare this distribution to the league-wide foul rate for an entire season over the entire course of a game. As situations for strategic fouling are rare early in a game, we expect the rate of fouls to be less frequent early in the game than the average over the course of the entire game.

**Data set**: The website basketball-reference.com contains historical game data for the NBA, including play-by-play data. We will take a random sample of 50 regular season games from the 2023-24 NBA season and record the time between the start of the game and the first personal foul, the time between the first personal foul and the second personal foul, and the time between the second personal foul and the third personal foul. We will also obtain the league-wide foul rate (a single value) for the 2023-24 NBA season from basketball-reference.com for comparison.

We will share the random sample of data to ensure that the project is entirely reproducible.

**Topics covered**: We have learned that the waiting time until the third event in a Poisson process may be modelled with a gamma distribution with shape parameter a = 3 and scale parameter b. With this in mind, we will apply the following methods from STAT2300 to the data set:

1. We will use **maximum likelihood estimation** to estimate the scale parameter b, which represents the average time between fouls within the first 3 personal fouls of an NBA game during the 2023-24 season.
2. Next, we will produce a **QQ plot** to visually compare the time until the third foul with a gamma distribution, assuming the shape parameter is a = 3 and the scale parameter b is equal to the MLE derived in part 1. Then, we will employ a **Kolmogorov-Smirnov test** to formally test whether it is reasonable to assume the waiting time until the third personal foul follows a gamma distribution with these parameters.
3. Then, we will produce a **95% confidence interval** for the scale parameter b based on a likelihood interval obtained using the Newton-Raphson method. This will give a range of plausible values for the scale parameter b.
4. Finally, we will use a **Wald test** compare the scale parameter b to the known constant based on the entire season. In this way, we will be able to determine whether the rate of personal fouls varied throughout the game during the 2023-24 NBA season.