## Statistical Data Analysis Report

In order to help answer whether certain features are more responsible in determining which side wins, statistical data analysis is performed on selected features. The selected features are as follows:

1. Gamelength
2. Time of first tower taken
3. Number of each type of dragon
    a. Earth
    b. Ocean
    c. Air
    d. Infernal
    e. Elder
4. Mean gold of winning team by role
    a. Top
    b. Jungle
    c. Mid
    d. ADC
    e. Support

These features were selected because they are the most suitable features to perform analysis on. Many of the other features are simply booleans created from dummy encodings.

In order to analyze these features, bootstrap analysis was performed on each of these features to calculate the p-value. Using this p-value, it will be possible to determine if there is a significant difference between the winning red team and the winning blue team values.

The function to calculate the p-value is shown below. In order to properly analyze the values, some preprocessing steps were used to get the features in the proper format. This included formatting the values of the winning teams into a single column and calculating summary statistics such as the mean.

```
def calculate_p(blue, red, feature, N=10000):
    bs_mean_diff = np.empty(N)

    mean_total   = np.mean(feature)
    blue_shifted = blue - np.mean(blue) + mean_total
    red_shifted  = red - np.mean(red) + mean_total

    for i in range(N):
        bs_blue = np.random.choice(blue_shifted, size=len(league))
        bs_red = np.random.choice(red_shifted, size=len(league))
        bs_mean_diff[i] = np.mean(bs_blue) - np.mean(bs_red)

    emperical_mean_diff = np.mean(blue) - np.mean(red)
    p = np.sum(bs_mean_diff >= emperical_mean_diff) / len(gamelength_mean_diff)
    return p
```

The results are summarized in the table below.

| Feature | p-value |
|---|---|
| Gamelength | 1.0 |
| Time of first tower taken | 0.4908 |
| Earth Dragon | 0.999 |
| Ocean Dragon | 0.999 |
| Air Dragon | 1.0 |
| Infernal Dragon | 0 |
| Elder Dragon | 1.0 |
| Top Gold | 1.0 |
| Jungle Gold | 1.0 |
| Mid Gold | 1.0 |
| ADC Gold | 1.0 |
| Supp Gold | 1.0 |

The alpha value was chosen to be 0.05 to represent the 95% confidence interval. All of the calculated p-values were greater than the alpha value except for the infernal dragon, indicating the null hypothesis should be accepted for those features. Regarding the results of the bootstrap test for the infernal dragon, the results state that there is a significant difference between the number of infernal dragons taken by the blue team when they win versus the red team when they win. This may be explained by the champions that are more likely to be chosen by the blue side versus the red side, or even the geography of the game map. Nonetheless, the number of infernal dragons retrieved is clearly a significant indicator of the result of the game within this dataset.