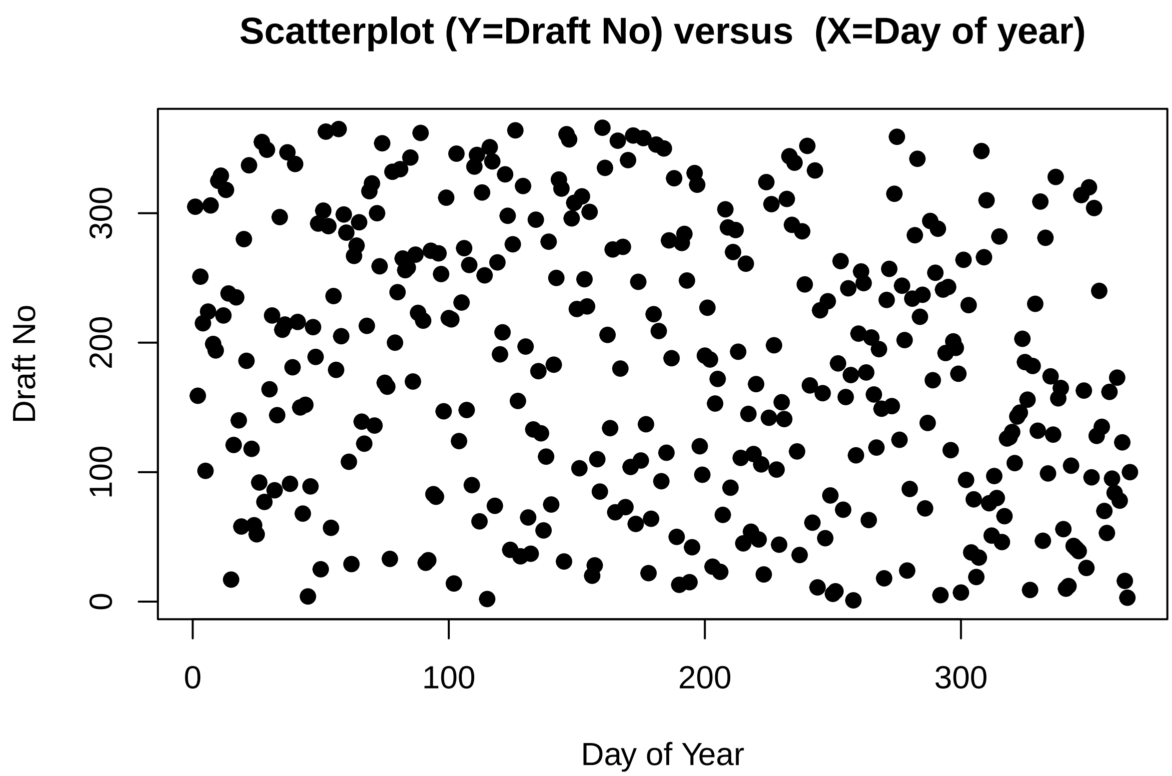
Lab 5

Ayesha Gamage-(ayega981)/Muditha Cherangani(mudch175)

2023-12-04

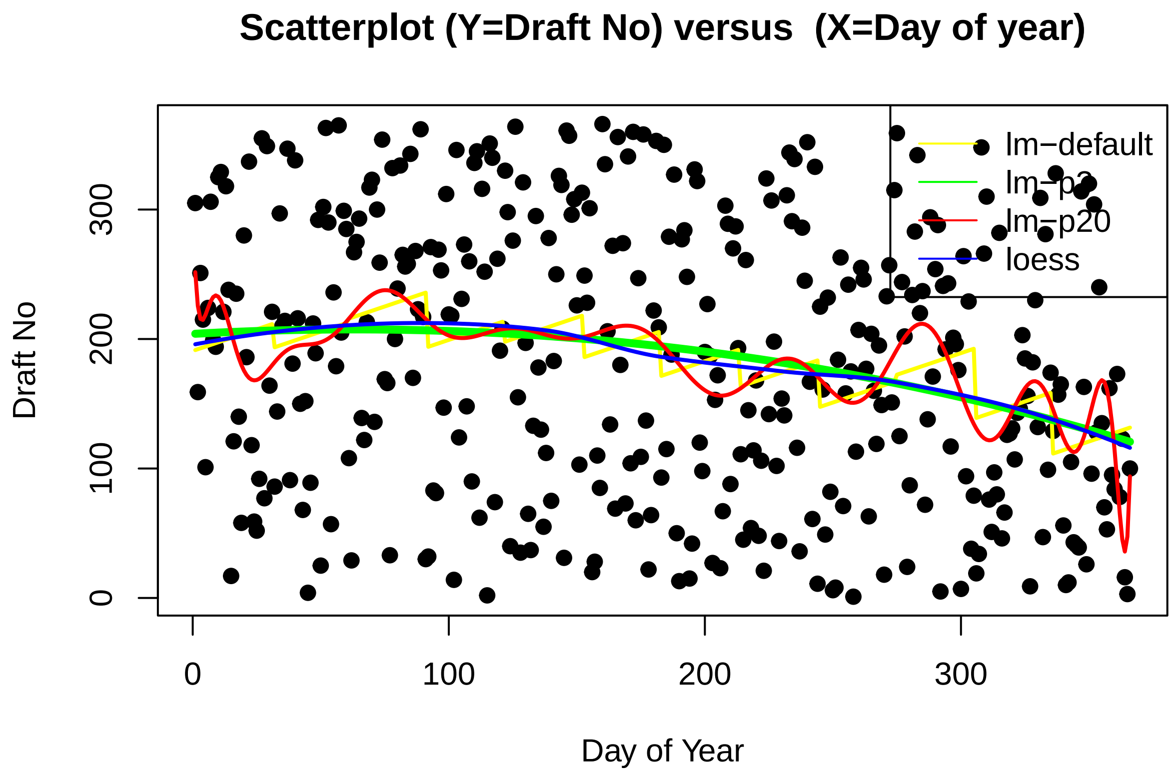
Question 1: Hypothesis testing

1.Scatterplot of Y versus X



The data points are spread out in this graph. This means there is no data trend; thus, there is no pattern visible in the scatterplot.

2. Fit a curve to the data



The yellow, red, and green line represents the prediction of the linear regression model.The blue line represents the prediction of the loess model. From the point of view of the lines, there is a clear relationship between parameters.There are correlation between X and Y. However, these curves are not enough to identify which parameters are significantly responsible for non–randomness. When consider the order ,order is significant for the curve because the highest order given the approximately equel curve to loess model's curve.

3.Check if the lottery is random

The statistic consider,

$$S = \sum_{i=1}^{n} |\hat{Y}_i - \overline{Y}|$$

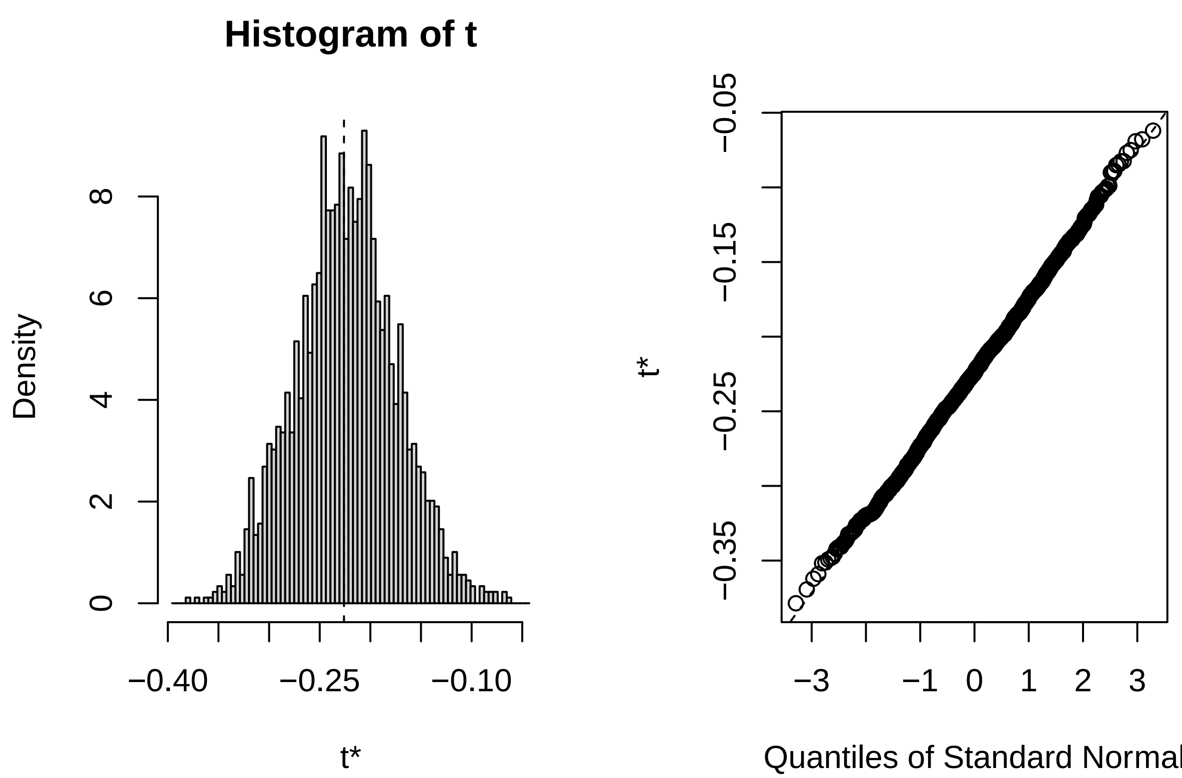
```
## S =  8238.649
```

Estimate S's distribution through a non–parametric bootstrap, B = 2000

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = res)
##
```

```
## Intervals :
## Level      Normal              Basic
## 95%   (-0.3251, -0.1318 )   (-0.3254, -0.1328 )
##
## Level     Percentile            BCa
## 95%   (-0.3193, -0.1267 )   (-0.3212, -0.1314 )
## Calculations and Intervals on Original Scale
```

```
## P-values : 0.4106409
```



The p–value of the observed value of S,

```
## p-value: 0.497
```

According to BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS, confidence intervals levels 95%. Computed p value is \$0.4955 > 0.05 \$, reject H_0 ,then lottery wasn't random.

4.Based on the test statistic S, implemented a function that tests the hypothesis,

$$H_0 : Lottery\ is\ random versus H_1 : Lottery\ is\ non\breve{}random.$$

The function return the value of S and its p–value, based on 2000 bootstrap samples.

```r
  set.seed(1234)
# Function to test the hypothesis
test_hypothesis <- function(data, loess_model, B = 2000) {
  # Using boot function
  res <- boot(data, statistic_S, R = B, loess_model = loess_model)

  # Calculate the observed S statistic
  observed_S <- statistic_S(data, 1:nrow(data), loess_model)

  # Calculate the p-value
  p_value <- mean(res$t >= observed_S)

  # Return results
  result <- list(observed_S = observed_S, p_value = p_value)
  return(result)
}
  # Example call
  result <- test_hypothesis(data, loess_model)

  # Display results
  cat("Observed S:", result$observed_S, "\n")
```

```
## Observed S: 8238.649
```

```r
  cat("p-value:", result$p_value, "\n")
```

```
## p-value: 0.497
```

5.Create a dataset of the same dimensions as the original data.Here I used consecutive dates and blocks of consecutive dates for create data.

```
## numeric(0)

## Results for Consecutive Dataset:

## Observed S: 14638.71

## p-value: 0.4935

## Accept H0: FALSE

## Results for Block Dataset:

## Observed S: 3881.383

## p-value: 0.496

## Accept H0: FALSE
```

```
## Results for Consecutive Dataset:
## Observed S: 2878.794
## p-value: 0.492
## Reject H0: FALSE
##
## Results for Block Dataset:
## Observed S: 2744.261
## p-value: 0.4805
## Reject H0: FALSE
## Results for Consecutive Dataset:
## Observed S: 4054.109
## p-value: 0.5
## Reject H0: FALSE
##
## Results for Block Dataset:
## Observed S: 4563.602
## p-value: 0.494
## Reject H0: FALSE
```

In this case, computed p-values are approximately equal, and all p-values are greater than 0.05. Then H_0 rejected.