



# The Statistical Analysis of the Birth Dates of Thoroughbred Racehorses and the Subsequent Impact on their Racing Career

By Eva Murphy and Lynn Sweetnam

## Introduction

The racing industry in Ireland annually contributes more than 1 billion euro to the economy and so when we discovered the rule which declares all thoroughbred racehorses to share the universal birthday of January 1st, we were compelled to investigate whether this rule allowed for a bias towards horses born closest to the beginning of the year.

## Method

- We began our project by collecting our sample selection, the names and the corresponding birth dates of over 500 thoroughbred horses.
- In order to analyse our data we converted the birth dates of the horses into days from January 1st.
- We then proceeded to analyse this data using the following analytic tests:
  - Frequency Distribution
  - Scatterplot Analysis
  - Descriptive Statistics
  - Hypothesis test

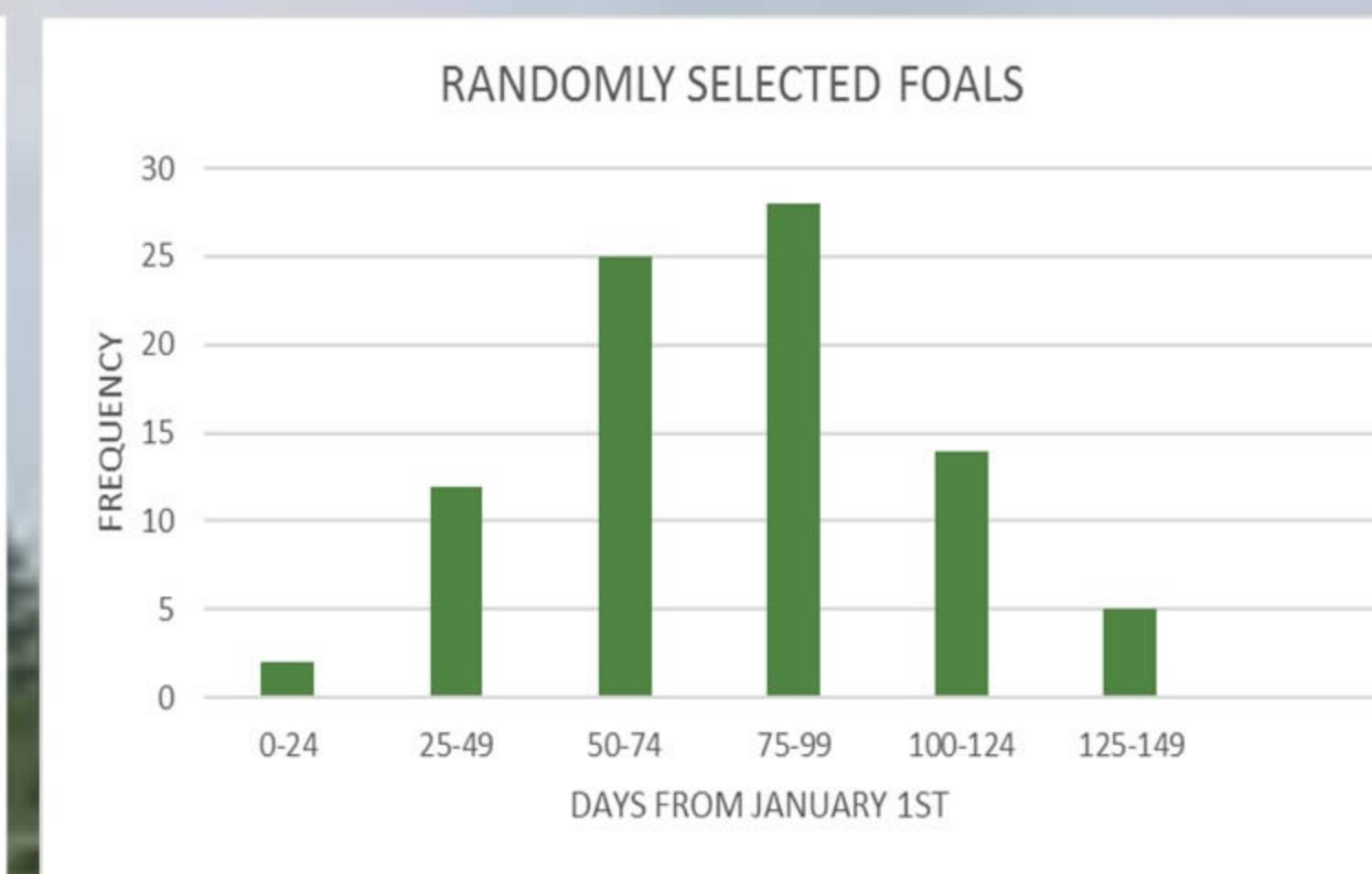
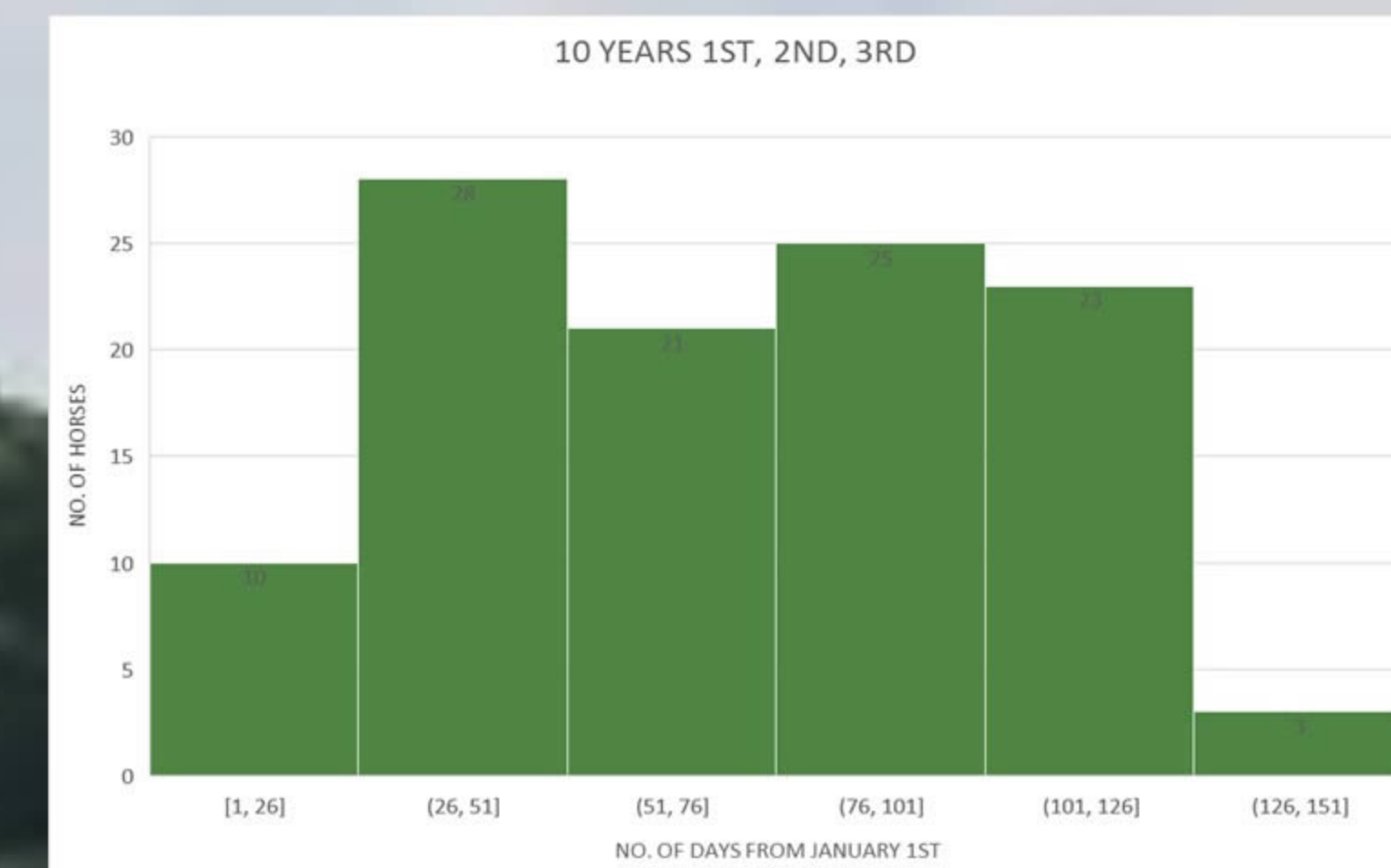
## Sample selection

Our sample selection was imperative to the success of our project as it was vital we obtained an accurate and unbiased range of data.

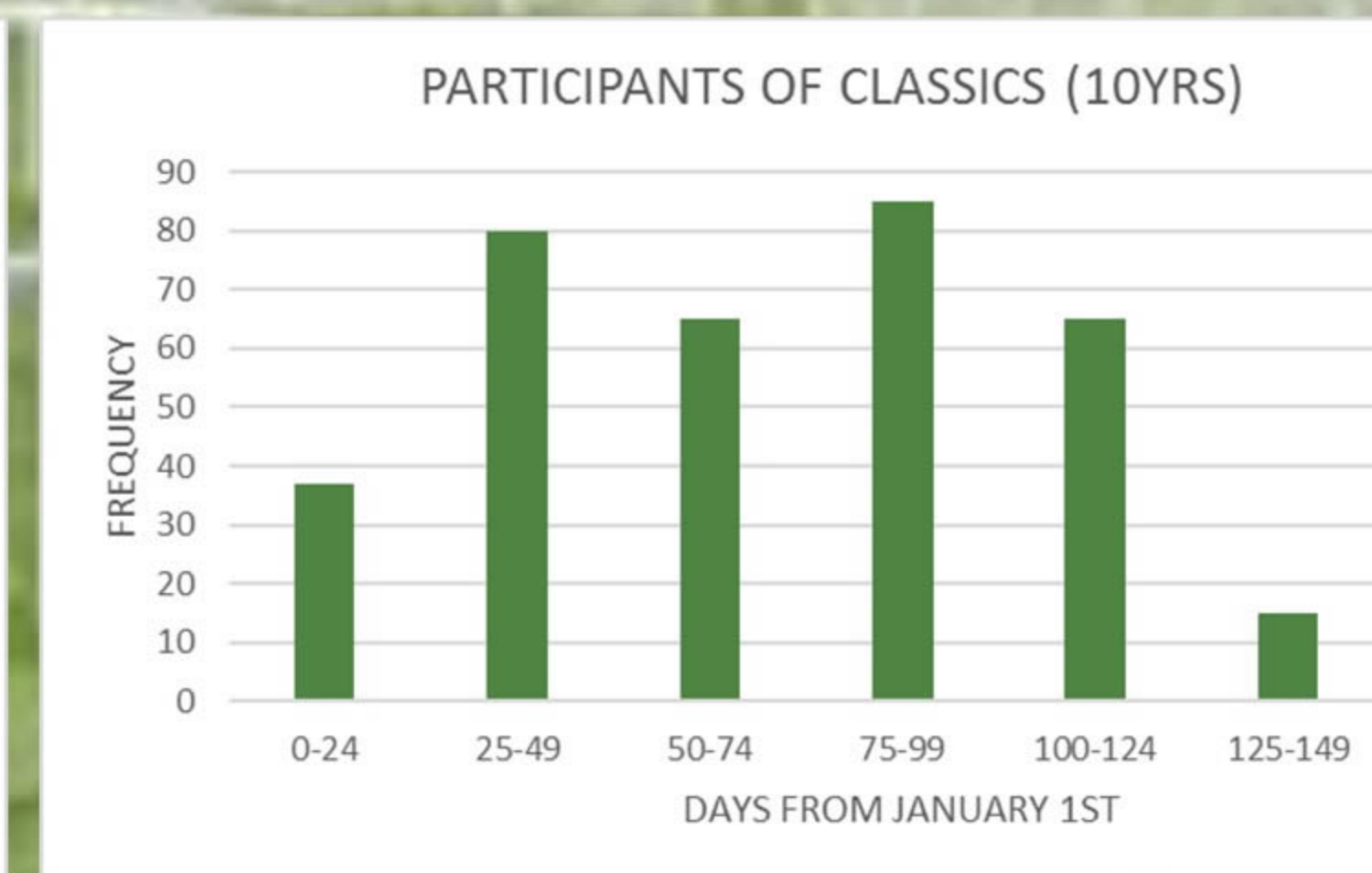
In order to do this we accumulated the names and dates of births of all the thoroughbred's which have won the Irish Classics (1000 Guineas, 2000 Guineas, Oaks, Derby) since 1966, that equates to fifty years worth of winners, as well as a list of all horses which have run in the Classics since 2007. It was important to go back 50 years so that our results would not favour newer technologies which now allow for artificially induced fertility and more successful selective breeding. It was also necessary to disregard National Hunt racing as the races have no age restrictions and hence it was impossible to accurately compare and contrast the dates of births of competitors.

## Hypothesis Test

Our original null hypothesis was that all thoroughbred racehorses which placed 1<sup>st</sup>, 2<sup>nd</sup>, or 3<sup>rd</sup> in the 4 classics ( the Oaks, the Derby, 1000 Guineas, 2000 Guineas) were born in the first 30 days of the year. However after extensive analytical tests we are 95% confident that the average successful thoroughbred's date of birth lies between 63.915 and 78.88 days after January 1<sup>st</sup>. Therefore we can reject our original hypothesis and accept our alternative hypothesis, that they are not exclusively born in the first 30 days.



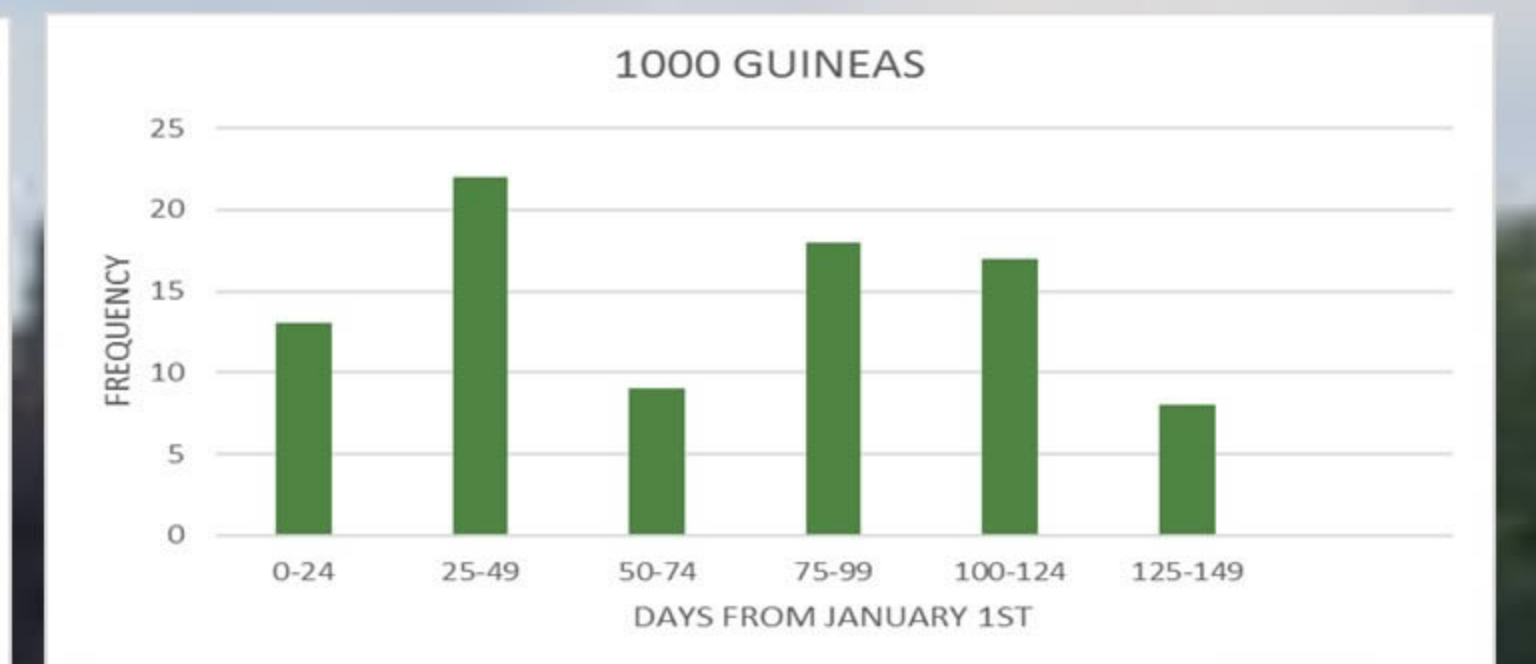
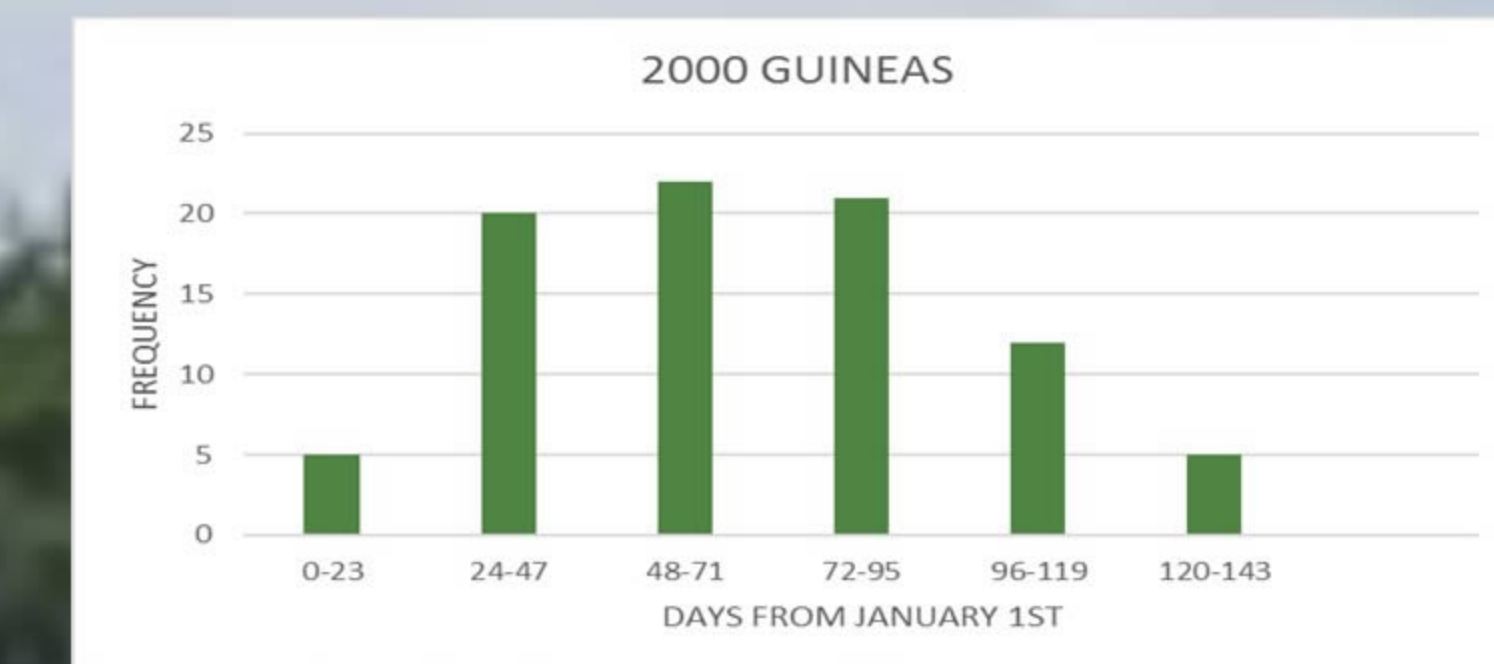
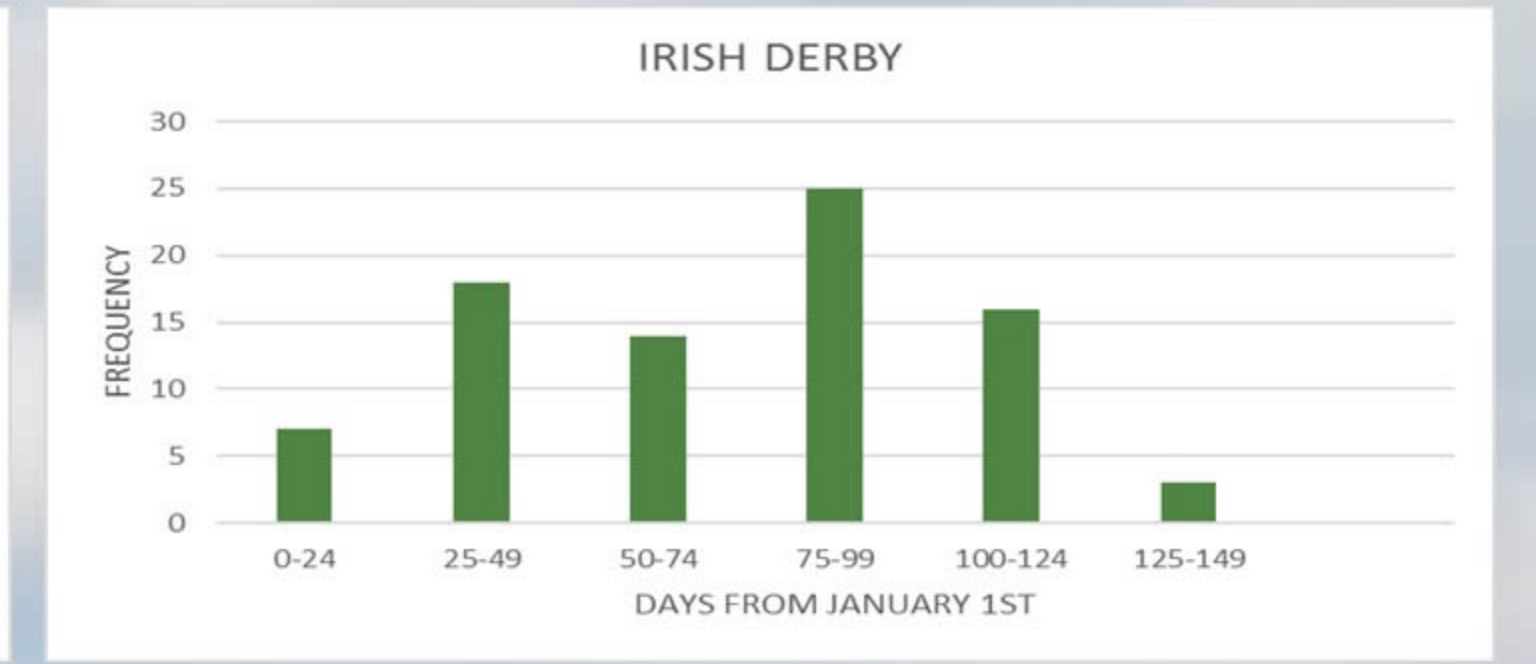
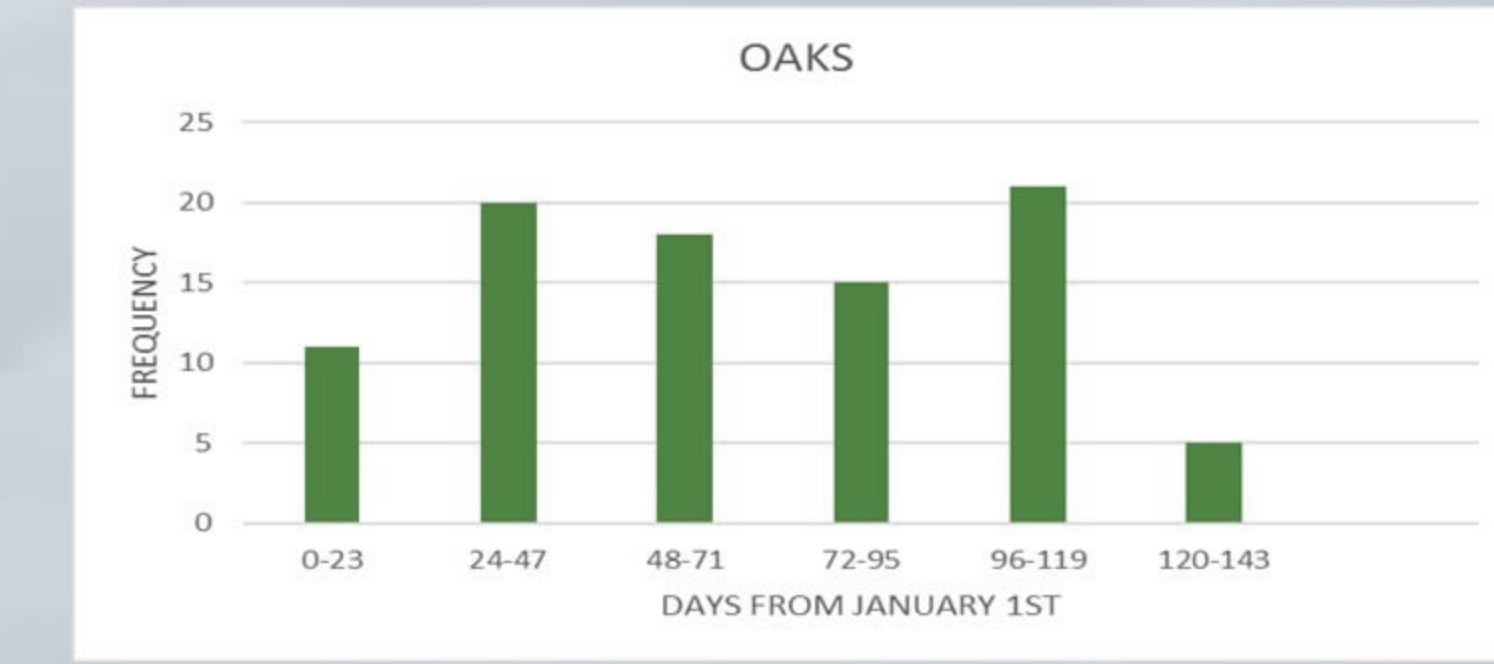
The histogram of 10 years of 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> place clearly suggests normal distribution, although it does appear that the bulk of the horses placing 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> are born between the dates January 26th and March 26th. A random selection of foals also indicates that the bulk are born between mid February and the beginning of April.



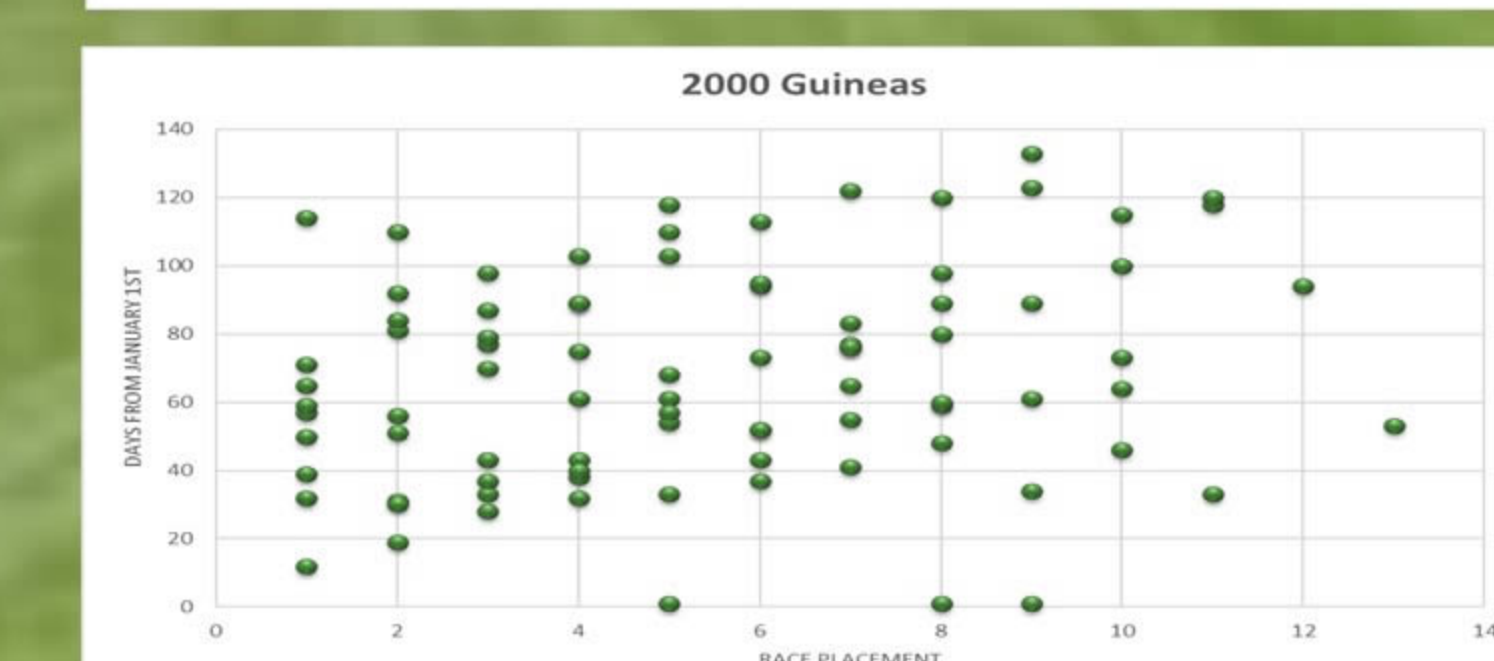
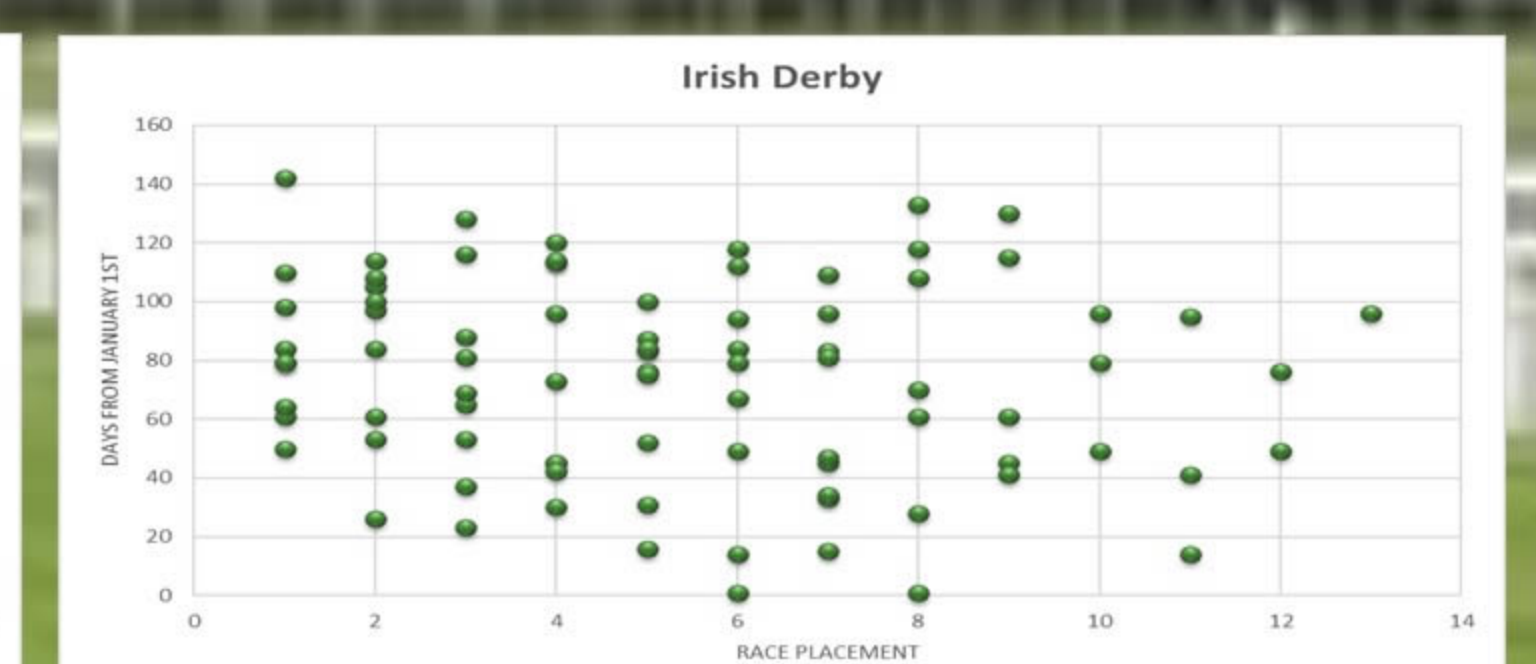
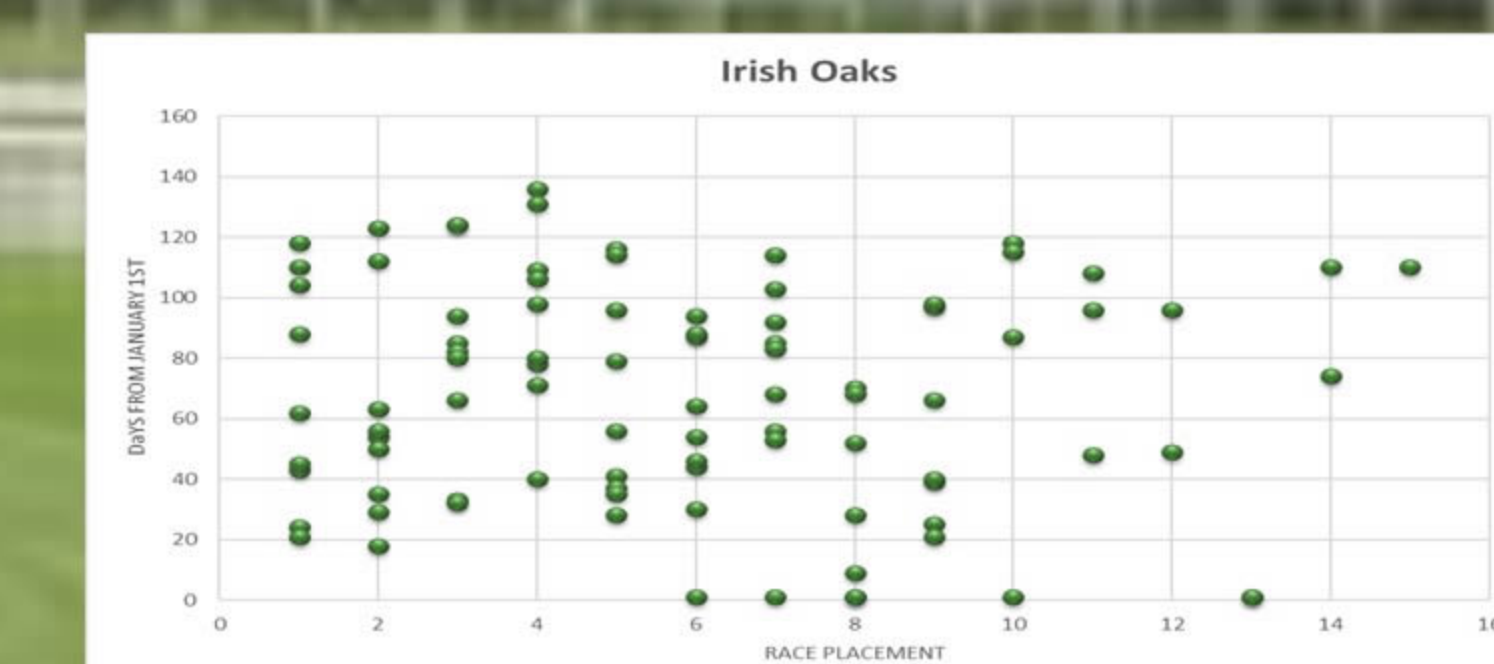
The frequency distribution chart clearly suggests that the majority of the winning horses throughout the last 50 years were born between day 40 and 120, which equates to February 9th and April 30th. We can further refine this to day 60 to 80, a period of 20 days between March 1st and 20th.

ALL CLASSIC PARTICIPANTS (10YRS)	
Mean	68.04201681
Standard Error	1.941019609
Median	69
Mode	1
Standard Deviation	36.67448558
Sample Variance	1345.017893
Kurtosis	-0.947828654

Skewness	-0.05919824
Range	141
Minimum	1
Maximum	142
Sum	24291
Count	357
Largest(1)	142
Smallest(1)	1
Confidence Level(95.0%)	3.817306204



The frequency distribution charts of our data clearly suggests normal distribution, with the bulk of the participants being born between days 23 and 99, which equates to January 23rd and April 9th



Our scatterplot analysis of the data suggests no correlation existed between placement and date of birth. The correlation coefficient confirms this, as it was quite close to zero implying no correlation.

## Conclusion

- Following a thorough analysis of our data it is evident that although there exists no clear bias towards a certain date, or any increased chance of success based on birthdate, it is strikingly apparent that based on our sample selection all winners are born in the first 150 days of the year.
- Our central tendencies also suggest that the bulk of all successful flat thoroughbred horses do appear to be born in February and March, suggesting there may exist an advantage to horses born closest to the beginning of the year.
- However on further analysis we also discovered that part of this result mirrors natural breeding patterns, as mares are long day breeders with a gestation period of 11 months and hence foal in the March and April.