

The Exploration of the Possible Relationship Between the Win Odds of the Favorite Horse and the Number of Times it Actually Wins the Race

**Information Gathering
&
Statistical Research**

**Prepared for
Lois (DELETED)
Statistical Analysis Tutor
(DELETED)
(DELETED)**

**Statistical Analysis Student
23 February 1992**

(DELETED)
(DELETED)
(DELETED)
(DELETED)

(ii)

(DELETED)

Lois (DELETED)
Statistical Analysis Tutor
(DELETED)
(DELETED)

Dear Ms (DELETED)

Information Gathering And Statistical Research

Thank you for giving me the opportunity to compile this report, which researches the possible relationship between the win odds of the favorite horse and the number of times it actually wins the race.

The report was compiled after randomly selecting a hundred races from 10,865 races and then separating the races where the favorite had won the race. After arranging the data in a histogram table and then a two-way table for a chi-square test to be done, it could be said that there is evidence that there is a relationship between the win odds of the favorite horse and how often it actually wins the race.

Both the histogram and the stem plot show some evidence of the fact that favorite horses with low win odds have a higher win strike rate over those with higher win odds.

The box plot and the Interquartile Range Criterion for Outliers showed that there were no outliers in the sample gathered.

With these findings, it could be said that the evidence of some relationship between the win odds of a favorite horse and how often the horse actually wins could be used to better the chances of winning a race. The results after exploring the data shows that a lower priced favorite has a higher probability of winning the race over a higher priced favorite.

Please do not hesitate to contact me if you require any further information about this report.

Table Of Contents

Letter of transmittal	(ii)
Executive Summary	(iii)
1. Introduction	3
2. Gathering Horseracing Information	4
3. Exploration and Analysis of the Data	5
3.1 Stem Plot and Box Plot	6
3.2 Frequency Table, Observed Odds and Expected Odds Tables	7
3.3 Histogram	8
3.4 Chi – Square Test	9
4. Conclusion	10
Appendix A (Some important horseracing terms)	11
Bibliography	12

Executive Summary

(iii)

Statistics is the science of collecting, organising, and interpreting numerical facts, which are called data. It has been used as the main tool to interpret winning averages for most gambling ventures such as casinos and lotto.

In this report some statistics principles are used to gather and explore horseracing data. The primary aim of the exploration is to find out if there is any relationship between the win odds of the favorite horse and how often it actually wins the race.

Random samples of a hundred races were analysed. A Stem plot and Box plot were constructed to check for any outliers. The data was then sorted into a frequency table with the odds arranged in intervals of 0.5. A histogram was made using this information.

A Chi-square Test showed evidence of a relationship between the win odds of the favorite horse and how often it actually won the race. With this evidence and by using the information from the histogram, it could be said that the favorite horse with win odds of below 2.5 have a higher chance of winning the race than those with above that amount.

With gambling being an increasing social problem, this report could benefit those who are addicted to horseracing by helping them select more winners instead of burning a long-term hole in their pockets.

Introduction

3

This report looks at how statistics could help better understand yet another area of gambling, horseracing.

Although the report shows evidence of a relationship between the win odds of the favorite horse and how often it actually wins the race, it is worth noting that there are many other factors affecting the outcome of any horserace. However, since all the sample races were selected randomly, it could be safely said that all the other factors affecting the outcome of a race were distributed evenly among the samples.

The information in this report should benefit anyone interested in seeing how statistics could be used to benefit us in our everyday life and also how it could be used to better our chances of winning at the races.

NOTE: all odds stated are in decimal format.

2. Gathering Horseracing Information

4

10,865 races were randomly selected from (DELETED) for the year 1991 results. (DELETED) has full information on past race results. They clearly show which horses were the favorite just before the race began. The results then show which horse won the race at what odds. The vital information needed for my presentation is; the horse that was the favorite just before the race started and the odds it was paying for a win, and the winner of the race. All this information are readily available at (DELETED) and they have given permission to use it any way I like.

The selection sample was further narrowed down to a fixed number of races per month from January to October in order to give a more randomly spread data over the year. Each month has over a thousand races so Table B from the Set Text was used to randomly select races for each month. Although the races are separately numbered from number one to twelve for each meeting, the entire race sample had to be renumbered from number one to the last race of the month in order to apply the Random Table B. All the races used in the sample had the odds given in dollars.

A favorite horse is the horse that most of the betting public counts on winning the race and the more money is put on it to win, the lower the odds becomes. However, usually any horse within the close vicinity of being able to win the race could be called a favorite. So for this reason the best way to class horses as a favorite for the purpose of this report is to rank them in order of their win odds. The horse with the lowest win odds is classed as the first favorite; the horse with the next lowest win odds is the second favorite and so on. This report is based on finding out if there is a relationship between the win odds of the first favorite and how often it actually wins the race.

3. Explorations and Analysis of the Data 5

The hundred samples were sorted by separating the races where the favorite had won from the races where the favorite had lost. The data was then sorted in ascending win odds order and divided in to six groups of interval with each interval separated by 0.50 from the next. A frequency table was constructed using the six lots of intervals by counting the number of times the favorite horse won having the win odds of the different intervals. A histogram graph was then constructed using the information from the table.

The sorted information of the races where the favorite horse had won the race was also used to detect any outliers. The information extracted is as follows:

Median: 2.35 **Quartile 1:** 1.8 **Quartile 3:** 2.85 **Interquartile Range:** 1.05

The 1.5 x IQR Criterion for Outliers: $1.5 \times 1.05 = 1.575$

$1.8 - 1.575 = 0.225$, $2.85 + 1.575 = 4.425$ thus the **outlier range is** 0.225 to 4.425 with our sample having no outliers. **The five number summary is:** 1.3, 1.8, 2.3, 2.85, 3.9

A step plot was also constructed to check for any outliers. The five number summary was used to make a box plot.

The win odds range data was also used to make an observed odds table and an expected odds table. The information from the tables was used to do a Chi-square Test, which tested to see if there was any evidence for or against the null hypothesis. The null hypothesis for this experiment stated that there was no relationship between the win odds of the favorite horse and its chances of winning the race.

A confidence interval test was also carried out using the mean odds and standard deviation, which were both worked out using Lotus 1-2-3.

df = 31, Mean of sample = 2.38 S = .7315, 95 % confidence interval is:

Mean of sample (Plus, Minus) (t*) (S/Square root of n) i.e. - 2.38 plus, minus (2.042) .7315/
 $5.656 = 2.38$ plus, minus 0.264.

We could be 95 % confident that a winning favorite chosen at random would have a win odds between 2.11 and 2.64.

- 3.1 Stem Plot and Box Plot** **6**
(Attached as graph paper -- Exhibit A)
- 3.2 Frequency Table, Observed Odds and Expected Odds Tables** **7**
(Attached as table paper -- Exhibit B)
- 3.3 Histogram** **8**
(Attached as graph paper -- Exhibit C)

3.4 Chi- Square Test **9**

$$X^2 = \text{SUM} \frac{(\text{Observed Count} - \text{Expected Count})^2}{\text{Expected Count}}$$

4. Conclusion

10

The hypothesis test of the two way table are as follows:

The null hypothesis H_0 states that there is no association between the row variable and column variable which in this case is the win odds and the number of times the favorite horse actually wins the race. The X^2 value was 24.728 with the degrees of freedom of 5, which gave $P < .0005$. With the probability so small there was enough evidence to reject the null hypothesis and to conclude that there is a relationship between the win odds of the favorite horse and how often it wins the race.

The step plot and box plot showed that there were no outliers in the sample making the spread of the data more consistent.

The histogram when used with the chi-square test showed some very interesting information. It revealed that horses with win odds of below 2.5 had a greater chance of winning the race over those with odds of more than 2.5. Out of the total of thirty-two races, which were won by the favorite, twenty-two of them were won by favorites paying less than 2.5 in win odds.

The confidence interval test showed that we could be 95% confident that the winning favorite horse chosen at random would have win odds between 2.11 and 2.64.

In summary it could be said that the favorite horse paying odds less than 2.5 for a win have more chances of winning the race and there is evidence of a relationship between the win odds of the favorite horse and how often they actually win the race.

As discussed, in Assessment tasks 2 & 3, I will be analysing the variables that affect the favorite horse from winning its race. This will be tested against the degrees of freedom and X^2 value to increase the winning chances even further.

Some important horseracing terms

Horse Number: Every horse in a race is allocated a number. It starts from the number 1 and goes up including all the horses in the field.

Field: Field refers to all the horses at a particular race. It does not refer to the horses in the entire racecourse. An example could be Saturday race number four with ten horses being a field. Obviously a larger field would consist of more horses when compared to a smaller one.

Barrier or Stall Number: Horses are allocated barriers/stalls/lanes, which they stand in just before the race. You could think of it as lanes used in sports running events such as the 100-meter sprint.

Form: Form is information about a horse's past performance. Numbers 1, 2 and 3 are used to represent first, second and third when showing past results.

Distance Of Race: As the wording states, this is the distance over which the race is or was run.

Weight: This refers to the weight the horse has to carry.

Track Condition: Track condition is the state of the track. It is usually classed as "firm" (F), "easy" (E) or "soft" (S). Sometimes the term "fast" (F) is used instead of firm but both words refer to similar track conditions. "Heavy" (H) is used to describe extremely soft tracks or at times water bogged ones.

Favorite: This is simply the most favored horse to win by the betting public in a particular race. Favorites usually pay lower odds when compared to the rest of the field.

Bibliography

12

Moore George and David Labara, (1989) *Introduction to Variable Statistics*. H. Freeman and Company New York.