Ken Perry

TURF ANALYSIS
(instructions)

Systems Design Lab

121 8th Street Altizer
Huntington, W. Va. 25705
Phone (304) 525-8932

TURF ANALYSIS

Your TURF ANALYSIS program uses a very sophisticated statistical tool called Multiple regression. In this type of analysis, historical or past information is used to predict future events. As time passes, the power and accuracy of this type of analysis will become self-evident. If you are not familiar with statistics, don't be discouraged, the concepts described in this manual will give you the necessary guidelines to make confident and accurate predictions. As with any new concept or program, the more you experiement the better you will become at making sound predictions.

WHERE TO FIND INFORMATION

Nearly all the information needed to operate this program can be obtained from the Daily Racing Form. This publication will generally give you a complete and comprehensive background of all the horses racing on a particular day. If you have never used a racing form, then you are encouraged to get one and become familiar with its contents.

INSTRUCTIONS

The best approach in learning how to use TURF ANALYSIS is to go through the program step by step using this manual as you go along. By the time you finish, you should be ready to start designing your own analysis.

LOADING PROGRAM FROM CASSETTE

To load your program, use the standard approach. That is, rewind your tape, type the word LOAD without hitting the return key and start your recorder by depressing the PLAY button. Then hit your RETURN key and wait until you hear two beeps. If your program did not load, repeat this procedure. If your program is loaded, you may start the analysis by typing the word RUN.

If everything has been done properly, you should be looking at the SYSTEMS DESIGN LAB company screen identifying your program as TURF ANALYSIS. When you are ready to continue, hit the RETURN key.

PARAMETER IDENTIFICATION SECTION

In this section you will define the parameters of your analysis. Assume you are handicapping a horse called JO TOBIN in tomorrows race. Further, JO TOBIN will be running on a 7 furlong track, carring 116 lbs, and will be riden by Shoemaker, who wins about 21% of his races.

In order for you to handicap this horse in tomorrows race, you will need to look at what he has done in his past races. For simplicity in this example, only four variables will be used, although you could use more. These four variables will be called DISTANCE, WEIGHT, JOCKEY, AND TIME.

The analysis will be structured such that JO TOBIN's TIME around the track will depend on the DISTANCE, the WEIGHT he is carring, and the expertise of the riding JOCKEY. That is, JO TOBIN's TIME in each race will be different as DISTANCE, WEIGHT, and JOCKEY changes.

Below, a Historical Data Table has been constructed. Assume the information in this table has been taken from a Daily Racing Form:

	DISTANCE (furlongs)	WEIGHT (lbs)	JOCKEY (% wins)	= TIME (seconds)
Race 1	6	119	.21	70
Race 2	6.5	116	.18	76
Race 3	7	117	.22	81
Race 4	6	117	.12	71
Race 5	8	116	.21	91

Remember, in tomorrows race JO TOBIN will be running a DISTANCE of 7 furlongs, WEIGHT 116 lbs, and the JOCKEY's expertise is .21% wins. The only unknown element is the TIME it will take JO TOBIN to run the 7 furlong distance. With the above information, you may procede with the analysis to determine JO TOBIN's

TIME. Answer the following questions using the Historical Data Table on page 2:

```
NAME OF HORSE?  JO TOBIN
NUMBER OF VARIABLES FOR JO TOBIN?    4
NUMBER OF RACES FOR JO TOBIN?        5
```

You should have typed the name of your horse as JO TOBIN, the number of variables 4 (DISTANCE, WEIGHT, JOCKEY, AND TIME), and the number of races as 5 (5 historical races).

VARIABLE IDENTIFICATION SECTION

```
NAME OF VARIABLE X(1)?  DISTANCE
NAME OF VARIABLE X(2)?  WEIGHT
NAME OF VARIABLE X(3)?  JOCKEY
NAME OF VARIABLE X(4)?  TIME
```

DISTANCE, WEIGHT, and JOCKEY are called independent variables and are always entered first. TIME is called the dependent variable and is always entered last. The dependent variable is the one you will always try to predict.

ENTER DATA FOR JO TOBIN'S DISTANCE

```
DISTANCE IN RACE (1)?   6
DISTANCE IN RACE (2)?   6.5
DISTANCE IN RACE (3)?   7
DISTANCE IN RACE (4)?   6
DISTANCE IN RACE (5)?   8
```

ENTER DATA FOR JO TOBIN'S WEIGHT

```
WEIGHT IN RACE (1)?   119
WEIGHT IN RACE (2)?   116
WEIGHT IN RACE (3)?   117
WEIGHT IN RACE (4)?   117
WEIGHT IN RACE (5)?   116
```

ENTER DATA FOR JO TOBIN'S JOCKEY

```
JOCKEY IN RACE (1)?   .21
```



```
JOCKEY IN RACE (2)?   .18
JOCKEY IN RACE (3)?   .22
JOCKEY IN RACE (4)?   .12
JOCKEY IN RACE (5)?   .21
```

ENTER DATA FOR JO TOBIN'S TIME

```
TIME IN RACE (1)?   70
TIME IN RACE (2)?   76
TIME IN RACE (3)?   81
TIME IN RACE (4)?   71
TIME IN RACE (5)?   91
```

Remember, TIME is the dependent variable and was entered last. This is also the variable you will predict in tomorrows race because you already know the DISTANCE(7), WEIGHT(116),and the JOCKEY's expertise(.21).

```
WANT TO MAKE CORRECTIONS?  yes or no
```

If you made a mistake in entering your data, stop and correct it now.

THEY'RE OFF

The computer is now analysing the data you have just entered. The results of this analysis will begin to appear in a few seconds.

REGRESSION EQUATION

```
VARIABLE X(1).............  10.0330085
VARIABLE X(2).............  -.30825705
VARIABLE X(3).............  -2.26480863

CONSTANT..................  47.0707021
```

This is an equation developed by the data you just entered. Written out, the regression equation would look like this (numbers rounded):

Y = 47.07 + 10.03X(1) - .30X(2) - 2.26X(3)

Re-writing the equation with the variable names, it would look like this:

TIME=47.07+10.03(DISTANCE)-.30(WEIGHT)-2.26(JOCKEY)

Hit return and in the next section you can check the accuracy of this equation against your historical races.

ACTUAL VS CALCULATED

JO TOBIN'S TIME IN RACE 1

ACTUAL.....	70
CALCULATED.....	70.110542
DIFFERENCE.....	-.11054218
PERCENT DIFF.....	-1.57934598E-03

Remember our dependent variable TIME in Race (1)? Go back to page 2 and look at the Historical Data Table and find JO TOBIN's TIME in the first race. His actual TIME around the 6 furlong track was 70 seconds. The Regression Equation predicts that his TIME should have been 70.110542 seconds. Using the information in the first race the 70.11 seconds was calculated as follows (numbers rounded):

70.11 = 47.07 + 10.033(6) - .308(119) - 2.265(.21)

Use the return key and continue to review each of the next four races until you come to the next section.

EQUATION STATISTICS

SS REGRESSION.....	294.68943
SS RESIDUAL.....	.110556618
TOTAL VARIANCE.....	294.799987
INDEX OF DET(R-SQ).....	.999624978
CORRELATION COEFF.....	.999812471
STD ERROR OF ESTIMATE.....	.332500553
MEAN SQUARE.....	.110556618
ZERO CHECK ON RESIDUAL.....	-1.1920929E-07
DEGREES OF FREEDOM.....	1 and 3

F RATIO 888.502431

Everything on the Equation Statistics screen will not be covered in this manual. However, some of the more important concepts will be reviewed to establish some working guidelines.

1. SS REGRESSION: This may also be called the "known" or "explained" variance. The greater this number is in relation to the TOTAL VARIANCE, the more accurate your predictions.

2. SS RESIDUAL: This may be called the "unknown" or "unexplained" variance. This is an indication of unknown factors influencing your dependent variable. The smaller this number is in relation to the TOTAL VARIANCE, the more accurate your predictions.

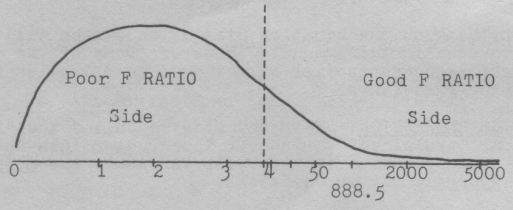
3. INDEX OF DETERMINATION(R-SQ): This measure will range from 0 to 1. The closer this number comes to 1, the better your chances of having an accurate prediction.

4. CORRELATION COEFFICIENT: This measure will range from -1 to 1. The closer this number comes to -1 or 1, the better your chances of having accurate prediction.

5. F RATIO: The F RATIO is an overall test of validity of the equation. The higher this number, the better. Generally, anything over 4 or 5 will be acceptable. This number is used in conjunction with F RATIO tables found in most statistic books. This test can be seen on a graph. The farther out to the right of this graph, the less there is a chance of your regression equation being random or inaccurate. An F RATIO graph is shown on the next page. Your F RATIO number is 888.5 and is shown on the right side.

6. T RATIO: The T RATIO is another test on the independent variables. This will give the importance each variable and is also used in conjunction T RATIO tables. (graph not shown)

F RATIO GRAPH:



If you follow the guidelines above, you should have no problem determining if your analysis is accurate or not. (Hit return to continue)

WANT TO RETURN TO REGRESSION SOLUTIONS? Yes or No

You may return and review everything covered up to this point by typing YES. Try it and then type NO to continue.

WANT TO PREDICT JO TOBIN'S NEXT RACE? Yes

You are now ready to predict JO TOBIN's TIME in tomorrows race.

JO TOBIN'S DISTANCE? 7
JO TOBIN'S WEIGHT? 116
JO TOBIN'S JOCKEY? .21

JO TOBIN'S TIME WILL BE 81.0683338

JO TOBIN's predicted TIME in tomorrows race is 81.06 seconds given a DISTANCE of 7 furlongs, WEIGHT 116lbs, and a JOCKEY rating of .21. The analysis of JO TOBIN is now complete.

Since you know the predicted TIME it should take JO TOBIN to run the 7 furlong distance, you are now ready to analyse the other horses in the race. After you have completed this, rank the horses from the fastest to the slowest. The fastest horse is the predicted

winner. The two fastest horses are your best EXACTA pick.

Hint: If you are able to confidently eliminate those horses having the least chances of winning by just looking at the Racing Form, then do so and use this program to analyse only the top contenders. This will save you time in analysing each race. The example you just completed would normally take about 5 minutes.

The example used in this manual is only one of many different approaches to analysing a horse race. Some people have successfully used this program in analysing pace, closing strength, and track variance. The system best suited for you will ultimately be a combination of the computer analysis and your knowledge of horse racing and the existing conditions at race time. For instance, if a horse was in a recent accident, you should not bet on a 3 legged horse even if the computer predicts it to win.

A LAST WORD ON PROGRAM RESTRICTIONS

1. Number of Variables: You may use as many variables as you like, but you must use at least 3. This is a multivariate analysis requiring at least 3 variables. Less than 3 variables is called a simple linear regression such as the program called "The Forecaster" by Systems Design Lab.

2. Historical Data Sets: There must be at least 1 more set of data than the number of variables or else there will be insufficient data to make an analysis. In our example we used 5 Historical Data Sets and 4 Variables.

3. Data: When entering data for any variable, the data must change at least once. For instance, if you are entering 5 Historical Data Sets for the variable Distance, you cannot enter 6 furlongs for all races. This would no longer be a variable, but instead, it would be a constant (6).

4. Very rarely you may enter a set of data to where an internal calculation Matrix Invert cannot be

calculated. This would be a very rare occasion, however, you would not be able to complete the analysis without starting over and changing something.

The computer will automatically check for these four restrictions and give you an error message whenever violated.

We at SYSTEMS DESIGN LAB thank you for purchasing our software and encourage you to call if you ever need assistance.

****

Apple II is a registered
trademark of Apple Computer Inc.