Statistical Analysis Report
Greyhound water consumption

Em Prof D Brynn Hibbert AM FRSN FRACI CChem FRSC

Sydney, 10 July 2021

Introduction

Instructions were received from Dr K Hunter to analyse data collected as part of the study of water consumption. The data was contained in an Excel Spreadsheet (GWIC.2019.1 - study data - UPDATED 2021.05.12.xlsx), which was updated with the total number of races at each meeting (GWIC.2019.1 - study data - UPDATED 2021.05.19.xlsx).

Although the focus is on water consumption no constraints were imposed on the statistical approaches taken or data included.

Within the spreadsheet, data were extracted in a form that could be read by statistical analysis software (Matlab™ R2019b). Summary statistics were calculated in the spreadsheet. All other calculations were made as an interactive Matlab 'Live code' program. The spreadsheet and Matlab script and output have been made available to the project leader.

Analysis

Variables studied

Water consumption (pre-race, post-race and total) is the primary quantity of interest in the study. Place in the race is also an outcome that could be affected (among others) by water consumption. Inherent characteristics of each dog: sex, age and weight are known. Race day information includes: track (code), distance to track, race number, race distance, kennel temperature and humidity, air temperature and humidity, and whether hydration products or pre-race water was given by the trainer prior to the race meeting commencing (i.e. *en route* to the track or on arrival at the track prior to kennelling). Information that was not used was the trainer (anonymised identification number), box number, details of the origin of the journey, whether hydration products were used after the race meeting, and if the greyhound drank from the wash bay hose after racing. Five greyhounds were enrolled twice, but this information was not used. Each result was included as an independent entry.

Data set

Seventy dogs were chosen for the study, 6 did not complete the study for different reasons and 54 had measurable water consumption (without spillage). Out of these 14 did not drink before their race.

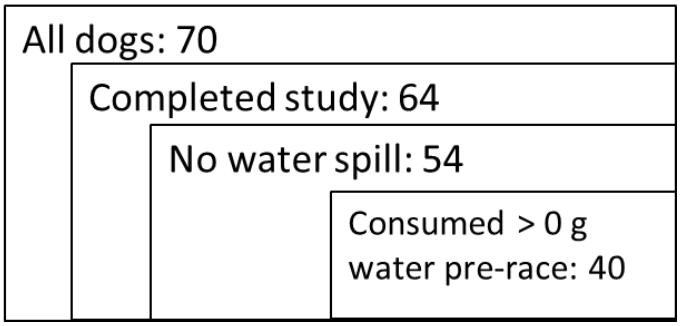

Figure 1: Data sets used for analysis

Temperature and humidity in the kennel and air were measured for each greyhound before racing, at the mid-point of the meeting and at the end of the meeting. Values used in the analysis were linearly interpolated for the intermediate races.

Analysis

Figure 2 shows the variables modelled, with significant effects shown in red. Continuous variables (e.g., distance to track) were fitted by a linear model, while categorical variables (e.g., sex) were modelled by one-way ANOVA. The test of significance was that the probability of an F-test of 'no effect' was less than 0.05, and so a significant effect of a prediction variable can be reported as "significant with a confidence better than 95 %".

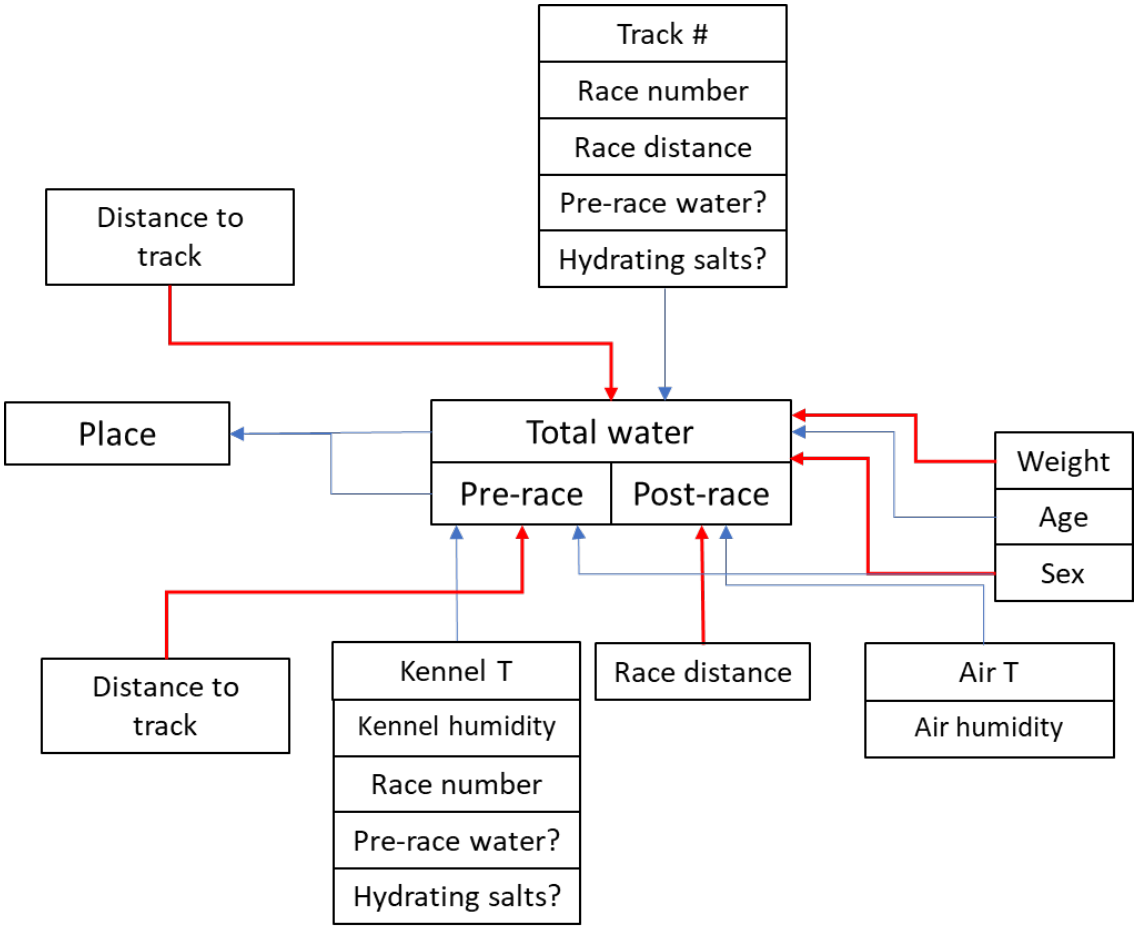

Figure 2: Variables modelled in study. Blocks of variables were modelled individually. Blue arrow = no significant relationship to single variable. Red arrow = significant relationship to single variable. (See text for tests applied and meaning of 'significant')

Several multivariate models were considered but the data were mostly not sufficient to discover meaningful complex interactions and higher-order effects.

Distributions

As many tests assume a normally-distributed variable, the statistical distributions of water consumption (pre-, post- and total) were determined. Pre-race water consumption was skewed by the 14 dogs who drank nothing, and a number of low consumptions. A normal distribution could not fit the data, but consumption greater than zero did fit a lognormal distribution (see Figure 3). The implication for modelling is that the logarithm of the pre-race water consumption is a more appropriate variable. It will be seen that graphs of analyses of pre-race water consumption are mostly on a log scale. As a lognormal distribution does not contain zero, only the 40 dogs that had measurable, non-zero, pre-race intake could be modelled. The 14 dogs who did not drink before the race were considered separately.

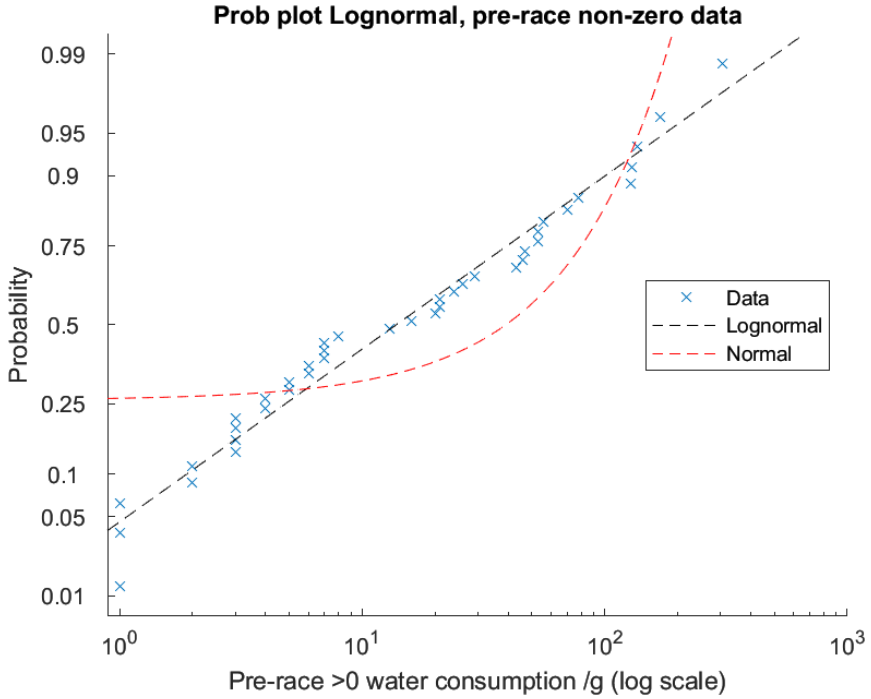

Figure 3: Probability plot of pre-race water consumption.

Post-race water consumption was, for the majority of greyhounds, greater than pre-race consumption, and was uncorrelated with pre-race consumption. The distribution was best fit by a normal distribution, in contrast to the lognormal distribution of pre-race water (see Figure 4).

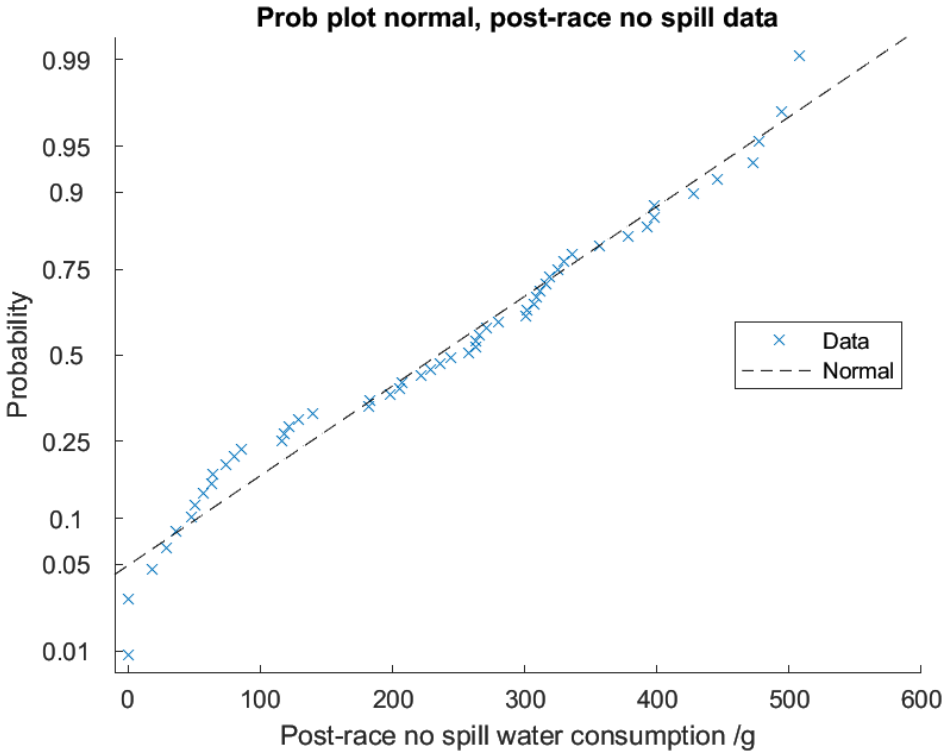

Figure 4: Probability plot of post-race water consumption.

The total water consumed better fitted a normal distribution with a deviation at low consumption from the pre-race data (Figure 5) and was treated accordingly.

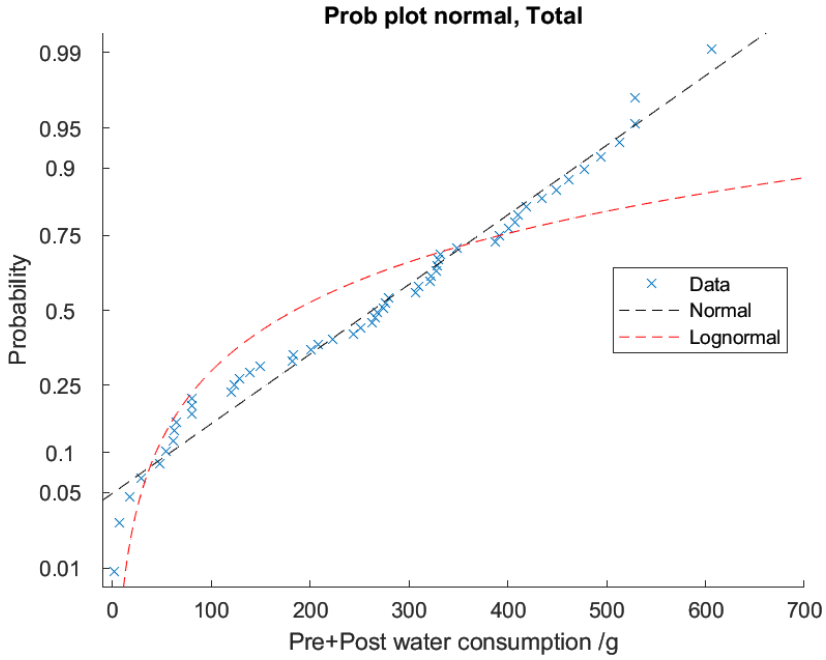

Figure 5: Probability plot of total water consumption.

Each distribution (and alternative for pre-race and total consumption) was tested by a one-sample Kolmogorov-Smirnov test (see Matlab output for results), which showed clear support for the chosen distribution.

Results

Summary of data

The Tables below shows summary statics of the data. For continuous variables the mean, median, standard deviation, maximum, and minimum are tabulated in Table 1. Categorical variables are counted in Table 2.

Table 1: Summary statistics of continuous data

Data	***n*** *	**mean**	**standard deviation**	**median**	**maximum**	**minimum**
Race number	64	6	3	6	12	1
Race Dist /m	64	385.3	87.2	411	618	259
Place	64	4.3	2.0	4	8	1
Weight /kg	64	28.8	3.1	28.1	36	22
Age /y	64	2.9	1.0	2.7	6.2	1.7
Dist to track /km	64	49.4	43.3	29.2	150	3.1
Pre-race water /g	54	28.9	54.6	5.5	305	0
ln(pre-race water /g)	40	16.2	1.6	2.7	5.7	0
Post-race water /g	54	234.2	141.8	251	508	0
Total Water /g	54	263.1	159.3	271	606	2
T (kennel) /°C†	64	18.9	1.3	18.5	21.4	16.3
T (air) /°C†	64	23.3	3.7	22.3	29.7	18.4
Humidity (kennel) %†	64	65.8	7.1	64.4	83	54
Humidity (air) %†	64	54.6	18.6	54.3	97.7	13.3

* Dogs completing study *n* = 64; Dogs not spilling water *n* = 54; Dogs drinking >0 g pre-race *n* = 40.
† Interpolated temperature and humidity data.

Table 2: Count of categorical data

Data	***n*** *	***N*(TRUE)**	***N*(FALSE)**
Sex = Female	64	36	28
Trainer gives water prior to race meeting commencing (i.e. *en route* to track or on arrival prior to kennelling)	64	38	26
Trainer gives hydration products prior to racing	64	26	38
Dog drank from hose bay after racing†	64	59	5
Bowl refilled†	64	46	18

* Dogs completing study *n* = 64;
† Not used in this analysis.

Non-significant effects

Significant effects will be described in detail below. Some highlights of the non-significant effects are:

- Placing in a race was not correlated with pre-race or total water consumption (see Figure 6).
- Provision of water or hydration products by the trainer prior to racing was not correlated with water consumption.
- Kennel temperature or humidity was not correlated with pre-race water consumption. It is noted that the range of temperatures in the kennels was only ~5 degrees.
- Air temperature or humidity was not correlated with post-race water consumption.
- Post-race water consumption did not correlate with pre-race water consumption but tended to be greater.

A quadratic model of pre-race water and place was also not significant. This allowed for evaluation of the scenario in which some water may help improve placing, but too much water may lead to a deterioration in performance. No such effect was observed.

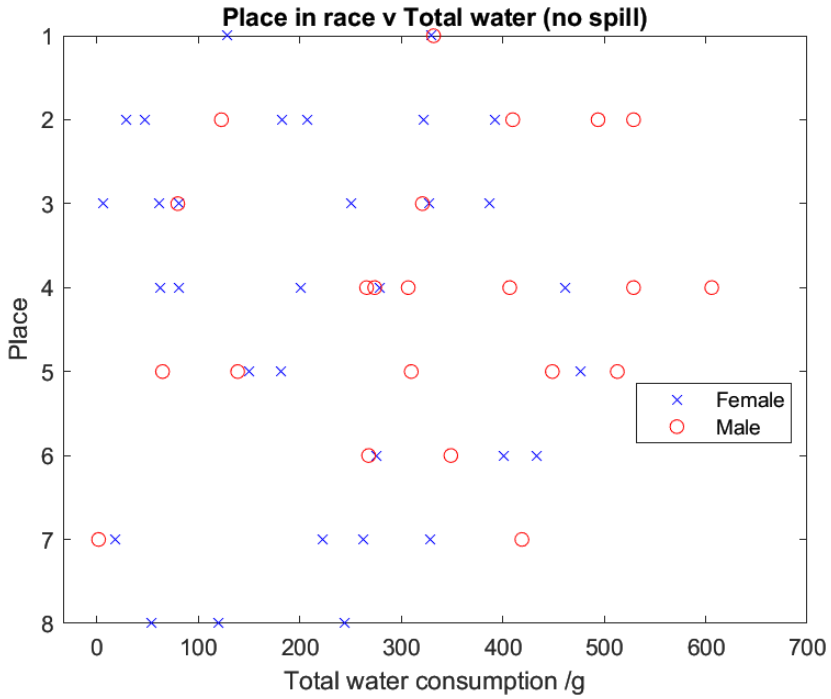

Figure 6: Place in race v total water consumed.

Significant effects

Significance of a model is assessed by calculating the probability of a constant model (i.e., the variables being studied have no effect). If this probability (α) is very small we reject the constant model and accept that the variables do indeed have a significant effect, "with 100(1 – α) % confidence". A threshold commonly used is $\alpha < 0.05$, or significance with 95 % confidence. In this section we give both α and the percentage confidence.

Distance to track correlated more strongly with pre-race water consumption [$p = 0.018$ (98.2 %), see Figure 7], not with post-race water consumption, and less strongly with total consumption [$p = 0.047$ (95.3 %)]. However, the effect is not great – about 2 g of water more, on average, for 10 km greater distance at the median distance of ~30 km.

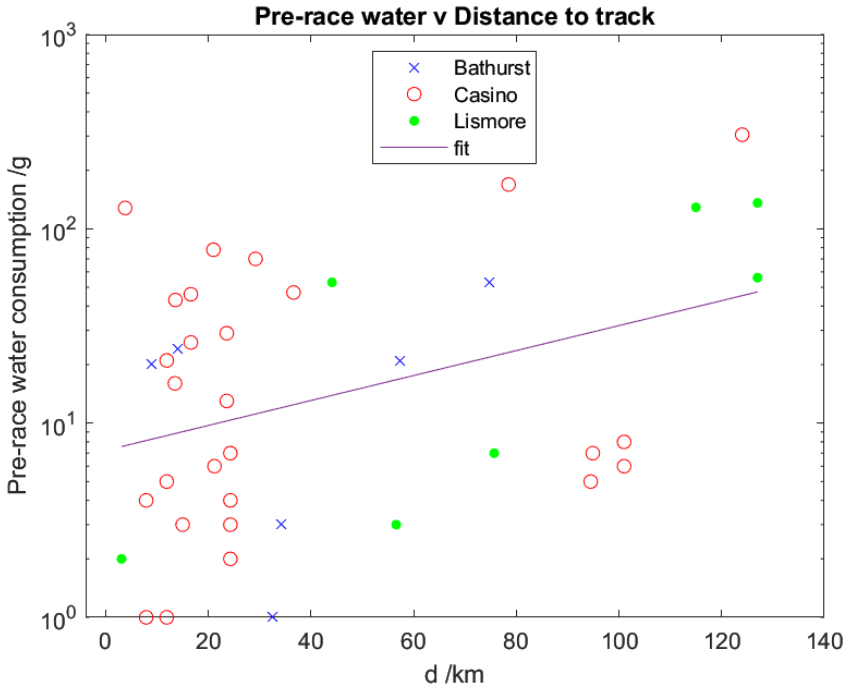

Figure 7: Pre-race water consumed (log scale) v distance travelled to track.

The race number (∝ time spent in the kennel) of itself does not correlate with pre-race water consumption, but together with distance to track gives a significant correlation [p=0.004 (99.6 %)]. The output of the linear model is shown below in Table 3 and Figure 8. Note that consumption increases with distance to track (edge along y-axis) and race number (edge along x-axis), but the effect is not additive.

Table 3: Output of linear model for ln(pre-race water) v race number x Distance to track

```
Linear regression model:
    ln_Water_pre ~ 1 + Race*To_Track_km

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	0.16419	0.70467	0.23301	0.81708
Race	0.28911	0.09921	2.9141	0.0060975
To_Track_km	0.037833	0.010423	3.6298	0.00087426
Race:To_Track_km	-0.0038911	0.0016	-2.4319	0.020121

```
Number of observations: 40, Error degrees of freedom: 36
Root Mean Squared Error: 1.35
R-squared: 0.305, Adjusted R-Squared: 0.247
F-statistic vs. constant model: 5.26, p-value = 0.00411
```

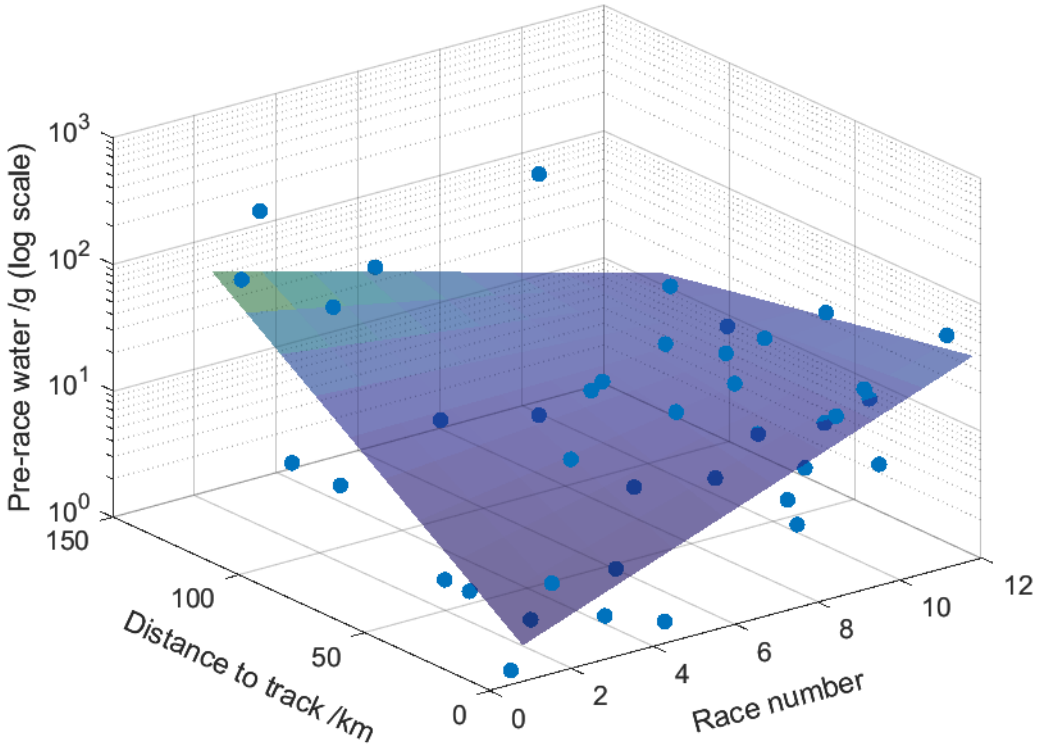

Figure 8: Pre-race water (log scale) v distance travelled to track v race number. Surface is fitted model.

Post-race water consumption correlates with the race distance [p = 0.014 (98.6 %)], Figure 9). Dogs drink more after a longer race by, on average, 50 g for every 100 m extra distance.

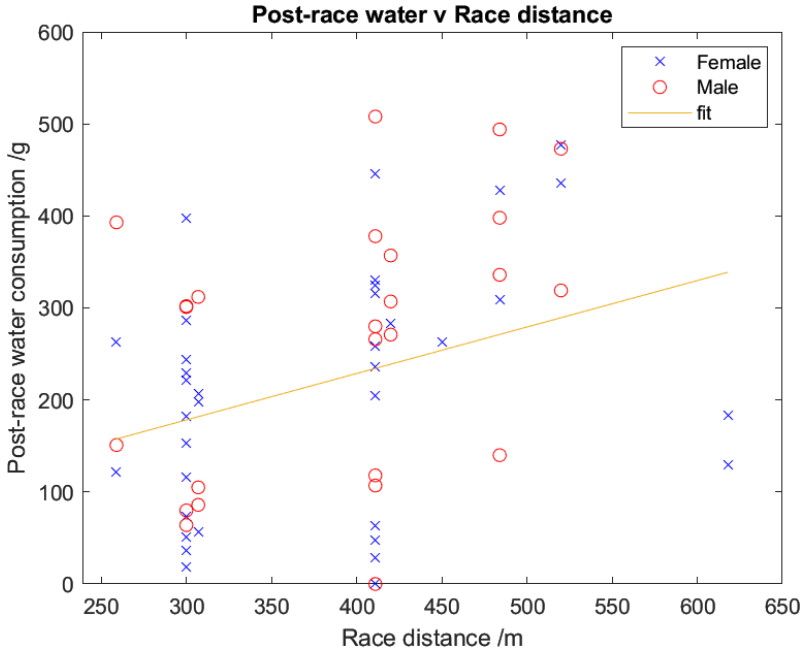

Figure 9: Post-race water consumption v race distance.

Heavier dogs drink more water than less heavy dogs [p = 0.016 (98.4 %), see Figure 10], on average 17 g of water in total per kg of additional weight. This correlation shows up as an

effect of sex (male greyhounds weigh, on average, 5.3 kg more than female greyhounds). See Figure 11, p= 0.013 (98.7 %).

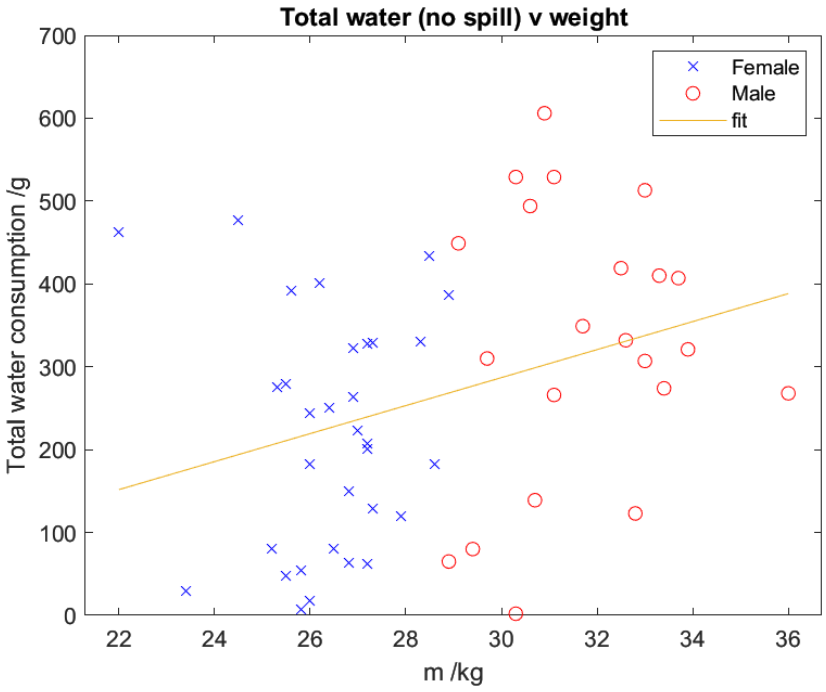

Figure 10: Total water consumed v weight of dog

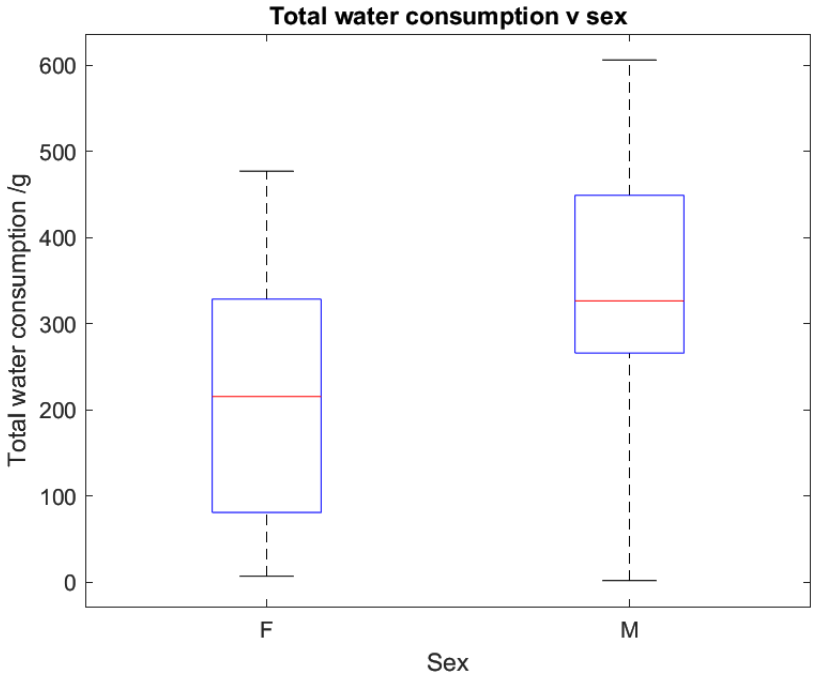

Figure 11: Total water consumed by sex of greyhound

Total water consumption correlates with distance to track, and this can be divided into male and female effects [p = 0.004 (99.6 %)]. Male greyhounds drink proportionately more per km distance travelled than female (Figure 12). A significant two-factor effect, weight and distance to track, on total water consumption was also found [p = 0.006 (99.4 %)]. This mirrors the effect of sex described above.

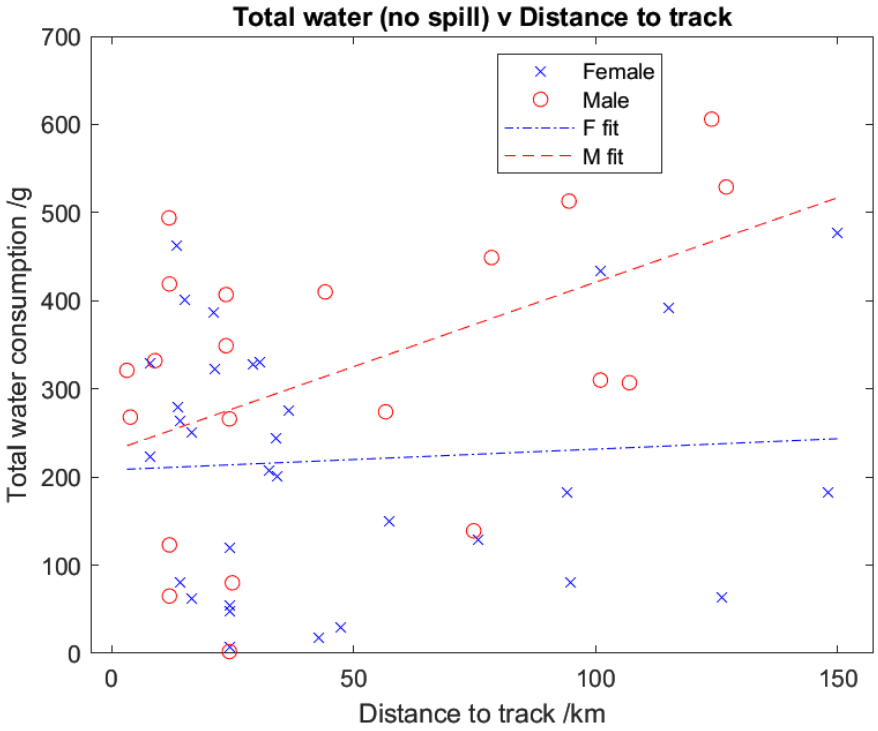

Figure 12: Total water consumption v distance to track fitted by sex of greyhound.

Categorical test

Out of the 54 dogs who had completed the study and had measurable water consumption before racing, 14 recorded zero g. Out of the 14 who did not drink in the kennel before the race, eleven dogs had been given water by the trainer (either *en route* to the track or after arrival), and three had not. A chi-square test gave a probability of 0.063 for the null hypothesis ($\chi^2 = 3.46$) implying a significance of 93.7 % for a correlation between having water before kennelling and not drinking when in the kennel before the race.

Conclusions

The results presented here are a statistical analysis of 40 to 70 sets of numbers. The data were collected from greyhound races, but any conclusions that use the statistical analysis concerning the nature of effects must be based on knowledge of racing and the physiology of greyhounds and understood as being supported by this set of results only. Note also that correlation does not imply causation, and that average effects do not inform us about any particular dog. As with any study, while recognising the effort put into the collection of data, more data will allow stronger conclusions.

Given the above caveats, the results appear to be reasonable, with correlations between water consumption and distance to track, weight, sex, race number (pre-race consumption, with distance to track), and race distance (post-race consumption). Perhaps of interest is that the other variables age, trainer giving water, trainer giving hydration products, kennel and air temperature and humidity, did not show significant effects. Note that more data could show up a smaller effect that is presently hidden in the noise.

It is also of significance that the placing of a greyhound is not correlated with water consumption.

A study of water consumption of the same greyhounds through a season might also yield useful results.