OZCHASE BOX DRAW DATA AUDIT – 2016

Terms of Reference

The OzChase box draw data audit was completed by Associate Professor Berwin A Turlach of the Centre for Applied Statistics at The University of Western Australia, at the request of the then RWWA Manager of Greyhound Racing, Mark Bottcher, on behalf of the OzChase User States with data extracted from the OzChase database by RWWA OzChase Support staff.

The terms of reference for the OzChase box draw data audit were as follows:

- RWWA to provide a csv file of racing data for all OzChase User States from the date that the User State commenced grading their fields on the system. Draw data will only be provided where the box draw was completed automatically by the system, i.e. manually completed draws replicated into OzChase will not be included in the data for auditing.

- The draw data to be provided will include details of the owner(s) and trainer of each greyhound at the actual date and time when the automatic box draw was completed in the system, rather than at the date and time of the actual race itself.

- RWWA would like a similar analysis performed as in the original algorithm analysis completed during the system development, concentrating on draws for the following entities where the auditor believes the sample size of their box draws in the data period is of statistical relevance to be considered:
  - Individual trainers
  - Individual owners
  - Group ownerships, be they either an Owner Group or Syndicate (specified within the data file)
  
  NOTE: There is no requirement to analyse to an individual owner level within the group ownerships.

- Aside from analysing these entities’ box draws themselves, RWWA would like analysis to be performed on the situation where trainers, with multiple starters in a race, are drawn side-by-side; this being one of the concerns continually raised by industry participants. RWWA is happy to take advice from the auditor on the number of multiple starters and the field size to consider auditing, especially given the potential complexity involved and the fact that auditing of data in relation to perceived ‘patterns’ in the box draw data may become a never-ending task.

- All analysis should be completed to a State level only, there is no need to consider owners and trainers across borders due to the additional complexity of the analysis and reporting.

- When analysing box draws, the auditor must make allowance for races where the number of runners drawn are less than 8. The rule of racing related to this matter is detailed overleaf:
R 22(3) Where there are less than 8 greyhounds eligible to compete in an Event at the time when the box draw is to be carried out, the following boxes shall be left vacant:

<table>
<thead>
<tr>
<th>NUMBER OF ELIGIBLE GREYHOUNDS</th>
<th>BOXES TO BE LEFT VACANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3 and 6</td>
</tr>
<tr>
<td>5</td>
<td>3, 5 and 7</td>
</tr>
<tr>
<td>4</td>
<td>2, 4, 6 and 8</td>
</tr>
<tr>
<td>3</td>
<td>2, 4, 6, 7 and 8</td>
</tr>
<tr>
<td>2</td>
<td>2, 4, 5, 6, 7 and 8</td>
</tr>
</tbody>
</table>

**Automatic Box Draw Process**

The automatic box draw process in OzChase was established to, in effect, replicate the concept of a manual box draw often completed on a race night, or elsewhere, for feature events around Australia.

The automatic box draw process sorts the list of greyhounds drawn in the race into a Pick Order on one hand and separately sorts the list of boxes which will be utilised by those greyhounds on the other, taking into account the field size and any box numbers to be left vacant, as required. The box for a particular greyhound is then matched up from the sorted list of box numbers using the Pick Order created for that list of greyhounds. This effectively provides two levels of randomness, just the same as a person(s) selecting a greyhound name at random from one ‘container’ and then selecting its box at random from another ‘container’, both of which have been shuffled or shaken; being a fair summation of the usual manual draw process.

The list of greyhounds to be box drawn in a particular race is a list of numerical Contestant ID’s which has no association whatsoever with the greyhound itself, nor any other data related to the greyhound such as its name, sire, dam, breeder, owner or trainer. The Contestant ID is allocated to a greyhound when the system or the Grader actually places a greyhound into a particular race. The ID is a unique sequential number allocated to every runner in all races created in OzChase.

A pictorial summary of the OzChase automatic box draw process is detailed overleaf. The example shown has a Pick Order of the greyhounds as 8, 7, 6, 5, 4, 3, 2, 1 and the boxes were sorted as 2, 4, 6, 8, 1, 3, 5, 7. The resulting box draw then has the greyhound with Pick Order 1 allocated to the first box in the list – box 2, the greyhound with Pick Order 2 allocated to the second box in the list – box 4, the greyhound with Pick Order 3 allocated to the third box in the list – box 6, etcetera.

A similar principle applies to the allocation of the reserve runners as first or second reserve (numbers 9 or 10), in OzChase User States where the reserves in standard (non-feature final) races are not seeded. The pictorial summary is referenced as ‘non-NSW – non-Feature’ given that NSW seed their reserves in all races and all feature event finals throughout Australia have seeded reserves, making the Reserve Runner Draw not applicable.
Audit Results – TAS Races

The results provided by the auditor in the box draw data audit report (in relation to TAS races only) are detailed below in full.

3 Results

3.6 Races in Tasmania

The provided spreadsheet contained information on 5054 races held at 491 meetings between 13 August 2012 and 22 September 2015. There were 2322 unique greyhound IDs, 893 unique owner IDs and 229 unique trainer IDs.

3.6.1 Allocation of individual greyhounds

The analysis of the actual overall allocation of greyhounds to boxes provided no evidence against the assumption that in every race every greyhound is equally likely to be assigned to any one of the boxes used (pall = 0:537). Analysing each individual greyhound’s allocation to boxes, taking into account that this involves the analysis of 2322 p-values by either using a 5% significance level with a Bonferroni correction or controlling the false discovery rate at 5%, provided no evidence that the box allocation of any individual greyhound was suspicious. Analysing the 253 p-values for the subset of greyhounds that started in at least 40 races yielded the same result, i.e. no evidence was found that the assumption, that each greyhound is equally likely to be assigned to any one of the boxes used in races in which the greyhound ran, is not tenable.

3.6.2 Allocation of greyhounds by owner

When analysing the allocation of greyhounds owned by the same owner to boxes, it was noticed that in 10 cases the information on the owner of the greyhound was missing. The analysis of the actual overall allocation of greyhounds owned by the same owner to boxes provided no evidence against the assumption that in every race every greyhound is equally likely to be assigned to any one of the boxes used (pall = 0:728). Analysing for each individual 11 owner the allocation of his or her greyhound(s) to boxes, taking into account that this involves the analysis of 893 p-values by either using a 5% significance level with a Bonferroni correction or controlling the false discovery rate at 5%, provided no evidence that the box allocation of greyhound(s) belonging to the same owner was suspicious for any of the owners. Analysing the 257 p-values for the subset of owners who had greyhounds starting at least 40 times yielded the same result, i.e. no evidence was found that the assumption, that the greyhound(s) of any owner is/are equally likely to be assigned to any of the boxes used in races in which the greyhound ran, is not tenable.

3.6.3 Allocation of greyhounds by trainer

When analysing the allocation of greyhounds trained by the same trainer to boxes, it was noticed that in 34 cases the information on the trainer of the greyhound was missing. The analysis of the actual overall allocation of greyhounds trained by the same trainer to boxes provided no evidence against the assumption that in every race every greyhound is equally likely to be assigned to any one of the boxes used (pall = 0:601). Analysing for each individual trainer the allocation of his or her greyhound(s) to boxes, taking into account that this involves the analysis of 229 p-values by either using a 5% significance level with a Bonferroni correction or controlling the false discovery rate at 5%, provided no evidence that the box allocation of
greyhound(s) trained by the same trainer was suspicious for any of the trainers. Analysing the 144 p-values for the subset of trainers who had greyhounds starting at least 40 times yielded the same result, i.e. no evidence was found that the assumption, that the greyhound(s) of any trainer is/are equally likely to be assigned to any of the boxes used in races in which the greyhound(s) ran, is not tenable.

3.6.4 Allocation of greyhounds to neighbouring boxes

Of the 5054 races in the spreadsheet, 4875 were races with 8 runners and in 2388 of those at least one trainer had exactly two greyhounds running. Altogether there were 122 trainer IDs involved, 2917 cases in which trainers had exactly two greyhounds running in a race, and 1 of these 2917 cases was attributed to the trainer ID used to code the missing trainer information. The analysis of the overall contingency table provided no evidence against the hypothesis that allocation mechanism works as it is designed to do (pall = 0.680). Analysing the results for individual trainers, taking into account that this involves the analysis of 122 p-values by either using a 5% significance level with a Bonferroni correction or controlling the false discovery rate at 5%, provided no evidence that any trainer has his or her two greyhounds allocated more often to neighbouring boxes than what would be expected under the allocation mechanism. Analysing the 23 p-values for the subset of trainer who have exactly two greyhounds starting in at least 20 races yielded the same result, i.e. no evidence was found that the assumption, that the two greyhounds of any trainer are equally likely to be assigned to any of the 8 boxes, is not tenable.

Audit Conclusion

The conclusions provided by the auditor in the box draw data audit report are detailed below in full.

In his conclusion, the auditor makes reference to a very small amount of missing data in the data extract provided by RWWA. This reference relates to 824 records out of the 470,313 provided (0.18%) with a blank Owner or Trainer field. The blank Owner or Trainer in these records was due to the fact that a box draw can be done any time up to about a week and a half before the actual race, and OzChase Users can process transfers or make corrections to the Owner or Trainer data after the actual Box Draw Date (which can then sometimes show as a blank Owner or Trainer at the time of box draw, as in the extract):

4 Conclusion

RWWA’s box draw algorithm is supposed to allocate, in any given race, the participating greyhounds to boxes in such a manner that each greyhound is equally likely to be assigned to any of the employed boxes. The summary conclusion from all the analyses reported in Section 3 is that there is no evidence in the data to suggest that RWWA’s box algorithm does not work as designed when analysing the data with respect to

- the allocation of individual greyhounds to boxes over all the races that they ran in,
- the allocation of the greyhound(s) of any owner to boxes over all races in which that owner had greyhound(s) running,
- the allocation of the greyhound(s) of any trainer to boxes over all races in which that trainer had greyhound(s) running, and
- the allocation of greyhounds trained by the same trainer to neighbouring boxes, if a trainer had two greyhounds in the same race.
While it is preferable to analyse complete data, for these analyses missing data was re-coded using a unique ID, as described in Section 2 (Methodology), to see whether there are any suspicious patterns in the missingness of the data. As none were found and the proportion of missing data is very small, it stands to reason that the summary conclusion would not change if the data had been complete.

- ends -