



ANALYTICS FOR INDUSTRY

1800 111 672
www.mla.com.au



Analytics for IndustRy



Overview

The meat processing and pastoral industries are rapidly entering the digital age, but not everyone is prepared.

Companies are accumulating large and complex data bases from properties, feedlots and processing systems, but often lack the skills and tools to collate, prepare and analyse these complex databases.

To date pastoral and meat processing companies have employed graduates with a rudimentary knowledge of data handling (largely Excel, with some basic knowledge of other statistical packages) and a passion for agriculture.

It's vital staff improve their data handling skills to cope with the large volumes of satellite, pasture, animal performance and carcass data needing to be delivered back to the processing and production sectors.

The challenge is to integrate this data into the business using appropriate analyses to create value for the business.

The recent push for objective carcass measurement (OCM) will accelerate the need for advanced data processing skills to underpin the development of value based payment systems. Given the sensitive nature of much of this data, it is critical that these analyses be done in-house.

Objectives and outcomes

The Analytics for industRy program delivers statistical training to develop the data handling/analysis skills for staff from the meat processing and pastoral industries. Participants are introduced to modern approaches to data

analysis and visualisation using the statistical programming language, R. The focus is on mastering basic skills so they can undertake these analyses independently.

The program will show participants how to address a statistical problem and break it up into logical steps. The R packages introduced to the participants will then assist them to convert these steps into statistical code, running the code and then producing an output.

The program focusses on handling mapping, pasture, live animal, carcass quality and carcass yield data. It would provide a mechanism for current staff to upskill and better service company needs.

Format

Workshops are delivered in two blocks separated by a week or two. An important component of the program is for participants to perform an analysis on their own data sets. We will begin work on this at the end of the first two days and participants will continue to work on it as "homework" before presenting their results in the follow-up two days. They can email the instructor to get help when they hit any roadblocks along the way.

Why R?

R is a true "data language". The R-Project describes R as a "[programming] language and environment for statistical computing". R was created with statistics and data in mind. In this sense, R is nearly unique among programming languages. It is a language that has been built for statistics and designed for data.

R is open source software, freely available to download and use. The analysis packages that extend R are also open source, free to use and updated by an active community of statisticians and data scientists. There are very active discussion boards and mailing lists where R users can get help from the community.

Among R's many features is the capability to output high quality static plots and figures. Increasingly, there are many interactive visualisations that can also be created with R, including geospatial mapping applications and web-based interfaces so that end-users can query data sets and analysis without needing to know any R programming.



Analytics for IndustRy

Course website: <http://garhtarr.github.io/meatR>

A website has been created with reference material that students will access during and after the program. It also contains more detail about what the course will cover.

Details

Dates: 18 - 20 March 2020 and
31 March - 1 April 2020

Location: MLA Brisbane office,
8/2 Upper Dairy Hall

45 King St, Bowen Hills, QLD

Cost \$2,500 + GST

Instructors

Dr Garth Tarr (University of Sydney)

Garth received his PhD in Mathematical Statistics from the University of Sydney and has held positions at the Australian National University and University of Newcastle. Garth is an expert R user, he has developed three R packages and contributed to a number of others. He is currently a lecturer in statistics and data science at the University of Sydney. His diverse interests include data visualisation, meat science, robust statistics, model selection, econometric modelling (including value based marketing), educational research and biostatistics.

Kevin Wang (University of Sydney)

Kevin is currently a PhD candidate and Postgraduate Teaching Fellow in the School of Mathematics and Statistics at the University of Sydney. Kevin's main research area is in statistical bioinformatics and is developing novel methods brought forward by high dimensional biomedical data. A central focus of his current research focuses on the increasingly popular boutique array platform and its application both as a validation platform for biomarkers for patients in melanoma studies. His other interests include data visualisation, statistical computing, model selection and biostatistics.

Specific topics

Introductory R

Introduction to R and RStudio; package management; importing and exporting data; working with data frames to manipulate, filter and summarise variables.

The tidyverse

Working with dplyr and tidyr for data manipulation, merging and reshaping; stringr for strings; lubridate for working with dates; ggplot2 for powerful data visualisation.

Workflow management

Working with projects in RStudio; data management principles; pipeline development; reproducible reporting with Rmarkdown

Data visualisation

Scatter plots; histograms; box plots; violin plots; spatial heat maps plots; time series; pivot tables; interactive plots; introduction to Shiny web applications.

Statistical methods

Fitting linear regression models; ANOVA; calculating and visualising predicted means; more advanced methods may be presented if there's demand.

