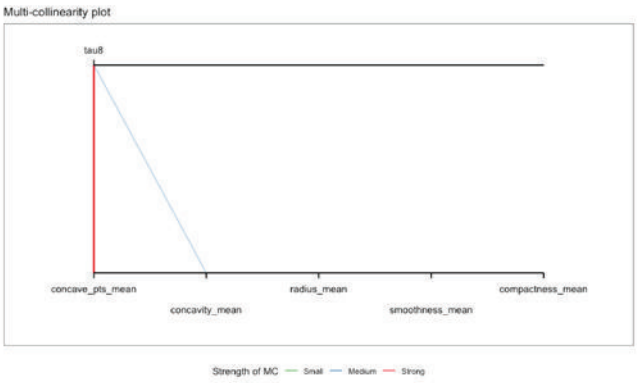
```
x2 = dplyr::select(x, -perimeter_mean, -area_mean)
(mcvis_result2 = mcvis::mcvis(x2))

##      radius_mean texture_mean smoothness_mean
compactness_mean concavity_mean
## tau8        0.03            0             0.01
0            0.23
##   concave_pts_mean symmetry_mean fractal_dim_mean
## tau8          0.72            0                0

plot(mcvis_result2, var_max = 5, label_dodge = TRUE)
```



Repeating mcvis again, we can also find that concave_pts_mean and concavity_mean as another potential source of multicollinearity and as their definition are very similar according to the documentation, we may decide to remove one of these.

Of course, depending on data context, we might not want to simply remove the collinearity causing variables as we have done here. Some common alternatives include averaging collinear predictors or using models that can decorrelate the predictor variables (e.g. random forest). Nonetheless, the diagnosis of multicollinearity is an important part of choosing which modelling strategy one should take.

Conclusion

The linear regression model is arguably the most powerful tool in statistics. However, using it in practice may require extra checks and practical considerations with one often ignored issue being multicollinearity. The R package mcvis provides tools that can help statisticians to identify different sources of multicollinearity to better decision making and the application of linear regression model.

Reference

Belsley, D. A., Kuh, E. & Welsch. R. E. Regression Diagnostics. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1980.

Lin, C., Wang, K. & Mueller, S. mcvis: A new framework for collinearity discovery, diagnostic and visualization. Journal of Computational and Graphical Statistics In Press, (2020). DOI: 10.1080/10618600.2020.1779729

STRengthening Analytical Thinking for Observational Studies (STRATOS)

Introducing the Visualisation Panel (SP)

Marc Baillie (1) and Marc Vandemeulebroecke (1); on behalf of the Visualisation Panel (1) Biostatistical Sciences and Pharmacometrics, Novartis Pharma AG, Basel, Switzerland

Previous issues of this Bulletin have introduced the nine Topic Groups (TGs) and the simulation panel of the STRATOS initiative. In this note, we introduce the newly formed visualisation panel (https://stratos-initiative.org/node/61). The aims of the visualisation panel are to promote the use of good graphical principles (https://graphicsprinciples.github.io/) for effective visual communication, providing guidance and recommendations covering all aspects from the design, implementation and review of statistical graphics.

Why do we need a visualisation panel? The use of appropriate statistical graphics is essential throughout a project, from formulating the research question, initial data analysis, execution of the analysis plan, through to communicating results, recommendations and conclusions. Researchers must not only "get the question right" (understand contextual subject matter) and "get the methods right" (technical expertise) but also "get the message right" (clear reporting). In other words, effective visual communication is a core competency for the applied researcher [1].

What is effective visual communication? Scientific influence relies on clear, transparent and effective communication. Reporting guidelines for example promote accurate and transparent written scientific documentation, e.g. see the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network that acts as an umbrella for reporting guidelines (https://www.equator-network.org/). Visual communication is one of the most effective channels for displaying quantitative information, and as with written communication, it is also important to be clear and accurate. Effective visual communication means using the visual channel to deliver the right information or messages clearly and concisely. By following the right graphical principles, we can better understand data, highlight core insights and influence decisions toward appropriate actions. Without it, we can fool ourselves and others and pave the way to wrong conclusions and actions.

Why is this important to researchers? The role of the "pragmatic statistician" in this process is to ensure that relevant information (concepts, assumptions, patterns, trends, signals, and conclusions) is clearly described and easy to interpret [2]. For this, we must understand the laws and principles of effective visual communication, such as the grammar of a (visual) language [3]. Visualisation is more than "plotting data"; it can lead to a deeper understanding and inform next steps.

A lot of ground has been covered on this theme from Tukey [4], Tufte [5] and Cleveland [6], collaborative initiatives such as

CTSpedia (https://www.ctspedia.org/do/view/CTSpedia) and PSI SIG VIS (https://www.psiweb.org/sigs-special-interest-groups/visualisation) to guidelines and recommendations [7] and [8], through to flexible tools to support statistical graphics (https://ggplot2.tidyverse.org/). However, traditional university and professional training curricula have not placed a lot of focus on effective communication [9]. Many researchers have to learn on the job through trial and error. This often leads to poor practice [10] or the avoidance of graphics [11].

While numerous literature, guidance and solutions exist how do we put this in practice in an accurate, transparent and reproducible way? This brings us back to the aims of the visualisation panel: to promote the use of good graphical principles for effective visual communication, providing guidance and recommendations. An additional aim is to support a cultural change in the adoption of these principles within applied research.

Some of these topics will be addressed in the near future. For example, after the 2nd general meeting of the STRATOS members at the Banff International Research Station in 2019 (https://www.birs.ca/events/2019/5-day-workshops/19w5198), the visualisation panel began working with other topic groups. One initial focus is to look at appropriate graphical and analytical tools to enable a researcher to perform initial data analysis (topic group 3) for regression modelling (topic group 2) in order to avoid misinterpretation, poor presentation and analysis errors. These necessary preparations are often forgotten even by experienced data analysts.

The panel is presently chaired by Mark Baillie, and Tim Morris, Rachel Phillips, Willi Sauerbrei, Svenja Seide and Marc Vandemeulebroecke are members. The panel is actively looking for another co-chair and members interested in contributing. Please reach out to learn more.

References:

[1] Vandemeulebroecke M, Baillie M, Margolskee A, Magnusson B. Effective Visual Communication for the Quantitative Scientist. CPT Pharmacometrics Syst Pharmacol. 2019;8(10):705-719. doi:10.1002/psp4.12455

[2] Chatfield, C. (2002), Confessions of a pragmatic statistician. Journal of the Royal Statistical Society: Series D (The Statistician), 51: 1-20. doi:10.1111/1467-9884.00294

[3] Wilkinson, L. The Grammar of Graphics (Springer, Berlin, Germany, 2005).

[4] Tukey, J. Exploratory Data Analysis. (Addison-Wesley, Reading, MA, 1977).

[5] Tufte, E.R. The Visual Display of Quantitative Information (Graphics Press, Cheshire, CT, 1983(2nd ed. 2001)).

[6] Cleveland, W.S. The Elements of Graphing Data. (Chapman and Hall, New York, NY, 1985).

[7] Guidelines for Reporting of Figures and Tables for Clinical Research in Urology. Vickers, Andrew J. et al. European Urology, Volume 78, Issue 1, 97 - 109

[8] Morris TP, Jarvis CI, Cragg W, et alProposals on Kaplan–Meier plots in medical research and a survey of stakeholder views: KMunicateBMJ Open 2019;9:e030215. doi: 10.1136/bmjopen-2019-030215

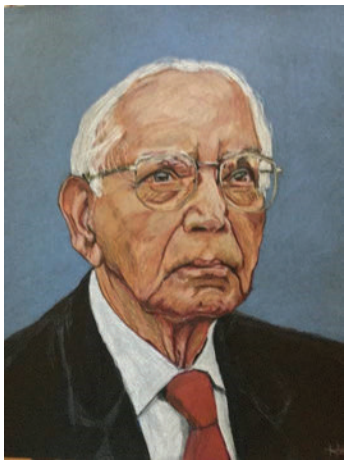
[9] Doumont, J.-L. Trees, Maps, and Theorems: Effective Communication for Rational Minds. (Principiae, Brussels, Belgium, 2009).

[10] Gordon, I. & Finch, S. Statistician heal thyself: have we lost the plot? J. Comput. Graph. Stat. 24, 1210–1229 (2015).

[11] Gelman, A., Pasarica, C. & Dodhia, R. (2002) Let's Practice What We Preach –Turning Tables into Graphs, The American Statistician, 56:2, 121-130, DOI: 10.1198/000313002317572790

Past President and Honorary Life Member of IBS Dr. Calyampudi Radhakrishna Rao (C.R. Rao) Turns 100

By: T. Krishna Kumar, Rockville Analytics - 13201 Carriage Court, Rockville, MD 20850



The International Biometric Society must be very proud that among all statisticians of eminence there are only two who became Centenarians, and both are past presidents of the society, and its honorary life members. These are D.J. Finney and C. R. Rao. Sir R.A. Fisher, the doctoral supervisor of C.R. Rao, founded the Biometric Society in 1947 (when Rao was his Ph.D. student) with five regional Chapters, comprising of Britain, France, Australia, the United States and India. The Indian Region of the Biometric Society was established with Prof P. C . Mahalanobis (C.R Rao's professor in India) as its President, and C.R. Rao as the Secretary. Rao was one of the founding members of the Indian Society of Medical Statistics. Rao was accepted as a student by Fisher only on the condition that he would work in his genetics laboratory to map chromosomes of mice. Dr Rao's earliest statistical work was on application of multivariate analysis to anthropometrics and biometrics. For our training in advanced statistics at ISI in 1959-61, we used extensively Rao's book "Advanced Statistical Methods in Biometric Research" (Wiley (1952)).

Dr. Rao was born on September 10, 1920 in *Huvina Hadagali, now in Karnataka state of India. (for an earlier account of Dr. Rao's achievements see 2015 ISMS Bulletin: http://isms-ind.org/ISMS%20Bulletin%20Septmber%202015.pdf, page 61, and for a more detailed*