RM-ANOVA on RAUs vs mixed effects logistic regression: a ruling of the high court for statistics.

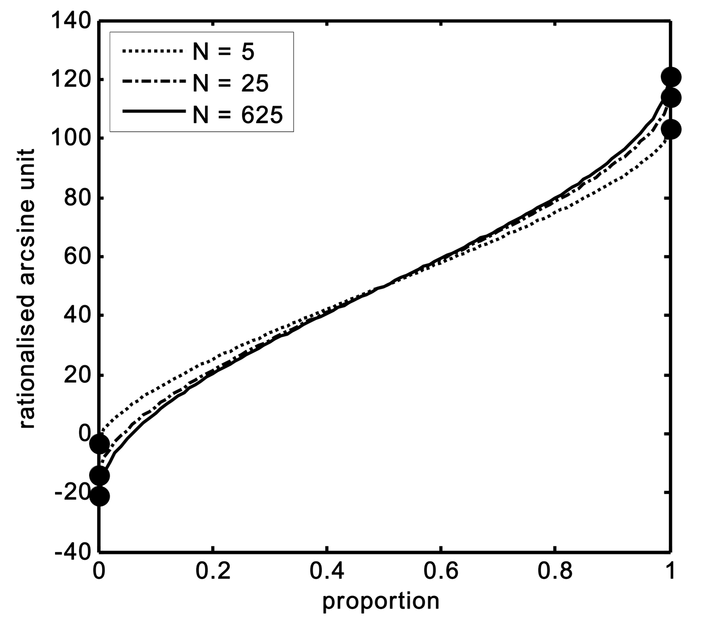
Gaston Hilkhuysen,
Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille, France

INTRODUCTION

Studebaker (1985) introduced the 'rationalized' arcsine (RAU) transform as a method to make proportions suitable for statistical analysis. It produces values numerically close to the originating percentages. Thirty years later, the transform is still widely applied[1] (49 citations in 2014 according to ISI-knowledge), especially by researchers of speech in noise, and usually before the application of analysis of variance for repeated measurements (RM-ANOVA). This approach has several shortcomings. Mixed-effects logistic regression overcomes these limitations.

RM-ANOVA on RAU

The RAU transform: $r = -23 + 46 \cdot (\sin^{-1}\sqrt{pN/(N+1)} + \sin^{-1}\sqrt{(pN+1)/(N+1)})$



where r denotes the RAU score; p the proportion and N the number of observations on which the proportion is based.

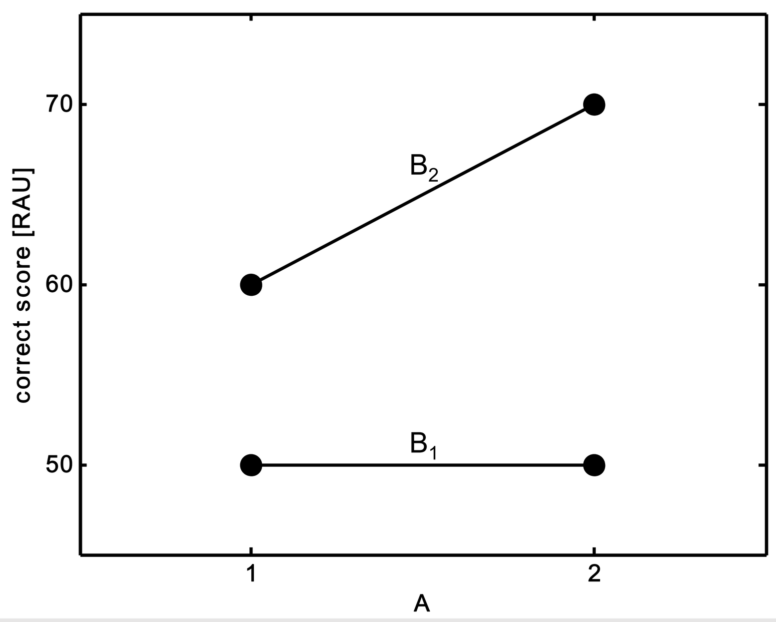
Simple 3-factor ANOVA model:

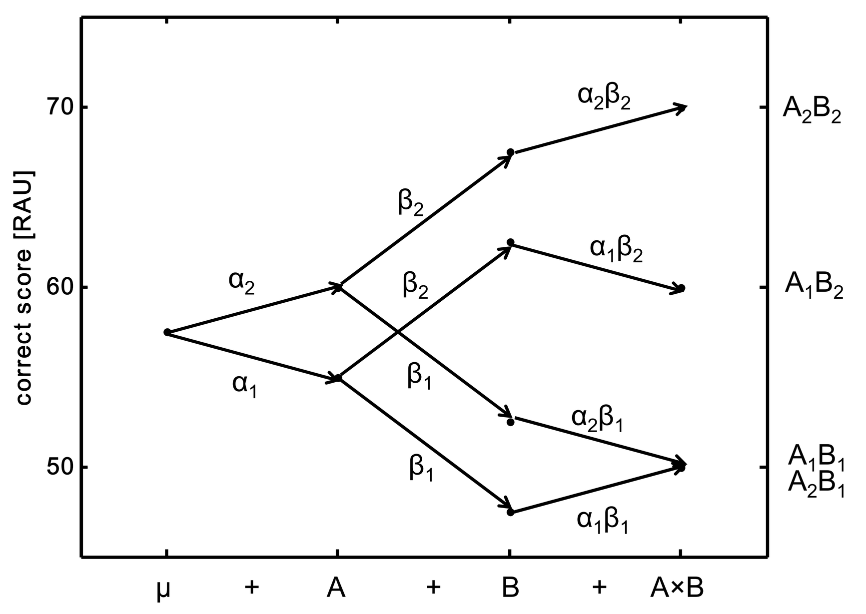
$\hat{r}_{ijn} = \mu + \alpha_i + \beta_j + \gamma_n + \alpha_i\beta_j + \alpha_i\gamma_n + \beta_j\gamma_n + \alpha_i\beta_j\gamma_n$

where:
$\hat{r}$ denotes the estimated RAU score; μ the global mean score; α and β the effects of factors A and B; γ the subject effect; *i and j* the levels within A and B; and *n* the various subjects;

Obscuring effects

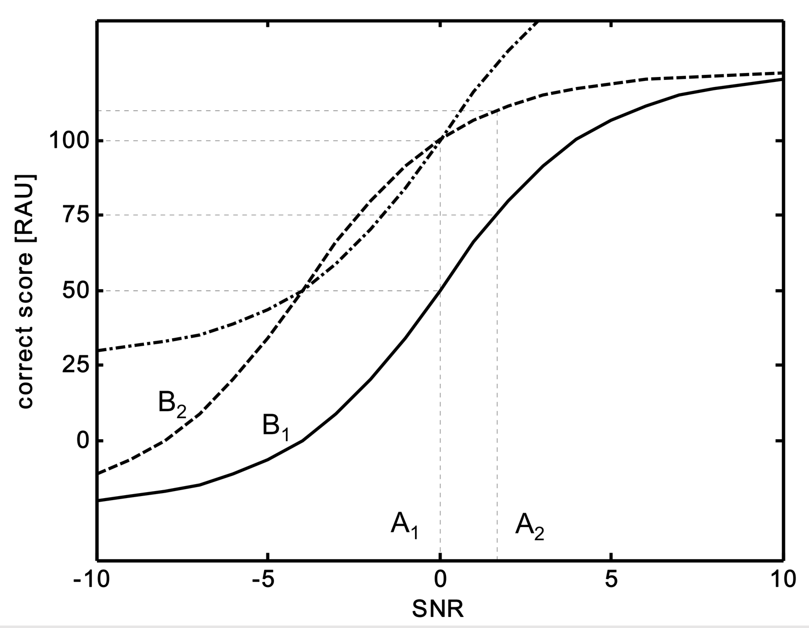
Consider the situation where the outcome is constant across the conditions A_1B_1 and A_1B_2, but augments across the conditions A_2B_1 and A_2B_2. ANOVA will report significant main effects of A, B as well as their interaction A × B. It appears more elucidating to state that changes in B has no effect for A=1, but are present for A=2.

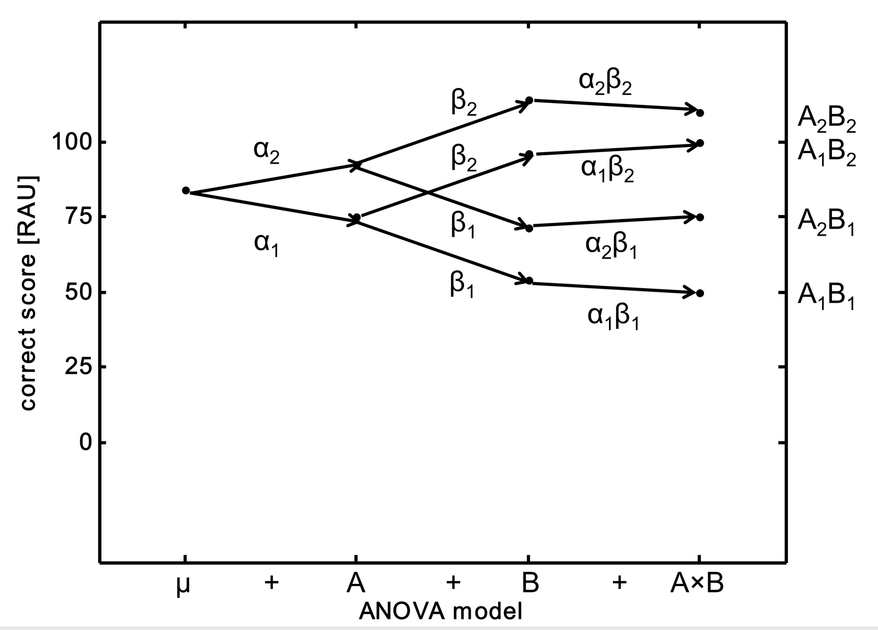




Trivial hypothesis

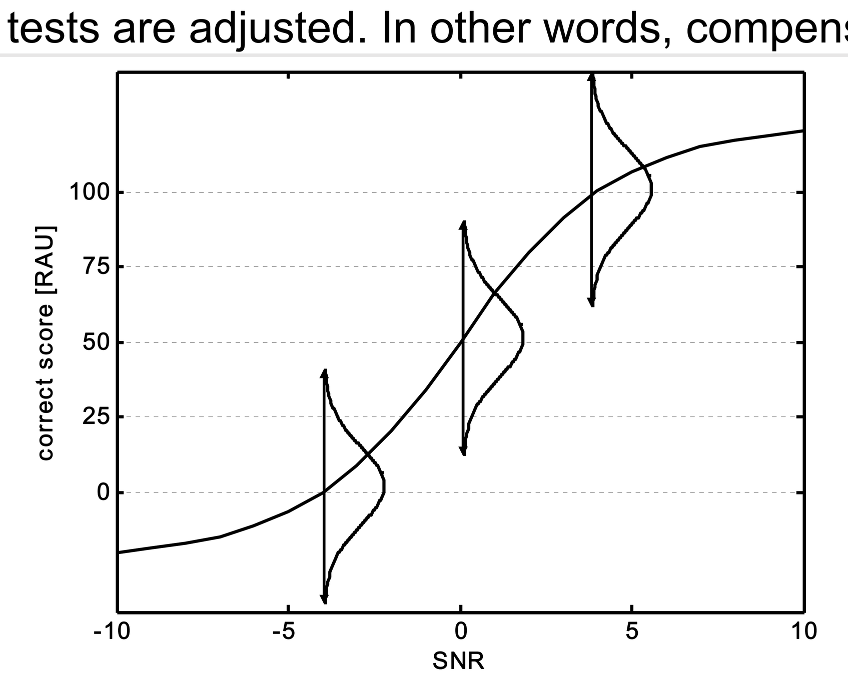
Many studies express outcomes in terms of RAU scores as a function of speech-to-noise ratio (Factor A). The resulting psychometric functions (PMF) are evaluated for different groups of listeners or types of processing (Factor B). Significant main effects of B, and its interaction effect with A, are falsely interpreted as reflecting effects on the PMF's position and shape, respectively. Because ANOVA models without A × B predict an impossible shift of the psychometric function along the vertical axis, the interaction becomes significant when B shifts the PMFs along the horizontal axis.

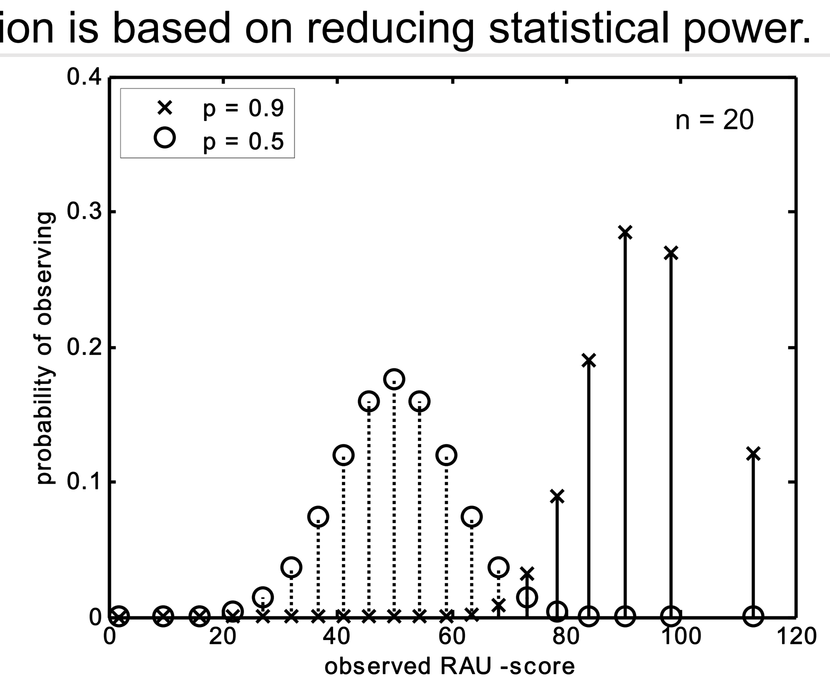




Invalid error distribution

Hypotheses testing in ANOVA is based on the assumption that the error distribution is constant and symmetric for different values of the outcome measure. Such is not the case for RAU scores or percentages that show highest variance around 50 % and less towards the extremes. For RM-ANOVA, assumptions are more stringent, i.e., the differences between the repeatedly measured scores should have constant variance. This usually checked with Mauchly's test for sphericity, without considering its statistical power or verifying its assumptions. (The latter would lead to a Droste effect.) To compensate for violations of RM-ANOVA's assumptions, the degrees of freedom in the RM-ANOVA's F-tests are adjusted. In other words, compensation is based on reducing statistical power.

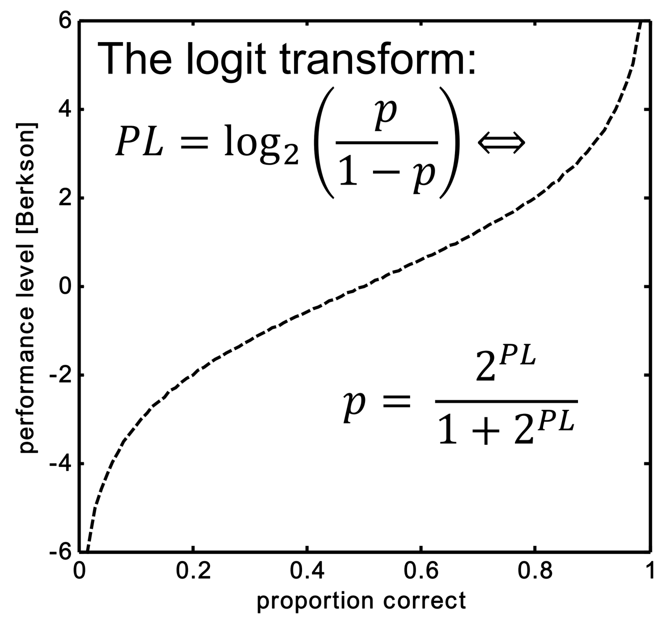




Unable to deal with unbalanced data

When subjects are unable to provide data in all requested experimental conditions, the number of observations will vary across conditions. The global mean will be biased, and the tested hypotheses are rarely of interest.

MIXED-EFFECTS LOGISTIC REGRESSION



ML-LR model:
$\widehat{PL} = \beta_{0n} + \beta_{1n}X_1 + \beta_{2n}X_2 + \beta_{3n}X_1X_2$

where $\widehat{PL}$ denotes the estimated performance level, β the regression coefficient; n the various subjects; and X_i the predictor variables.

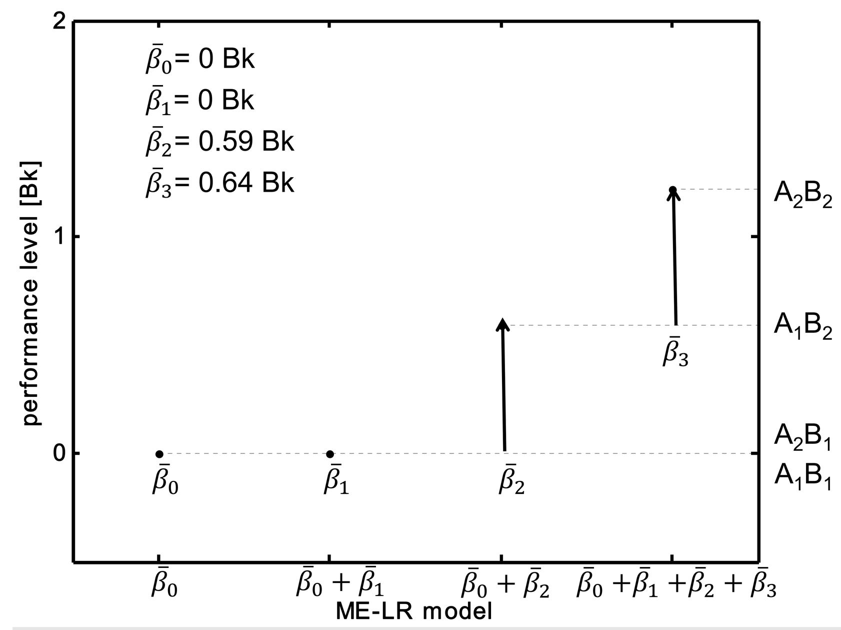
Consequences of dummy coding:

		X_1	
		0	1
X_2	0	A_1B_1	A_2B_1
	1	A_1B_2	A_2B_2

ML-LR has three ingredients: a logistic transform; a regression equation; and a mixed-effects aspect. The logistic transforms gives proportions a range from -∞ to ∞. Consequently, floor and ceiling effects no longer exist, and the proportions are always predicted within the 0 to 1 range. Using a 2-base logarithm facilitates interpretation of the regression coefficients: a one-Berkson unit increase signifies a doubling of correct answers with a fixed number of errors.

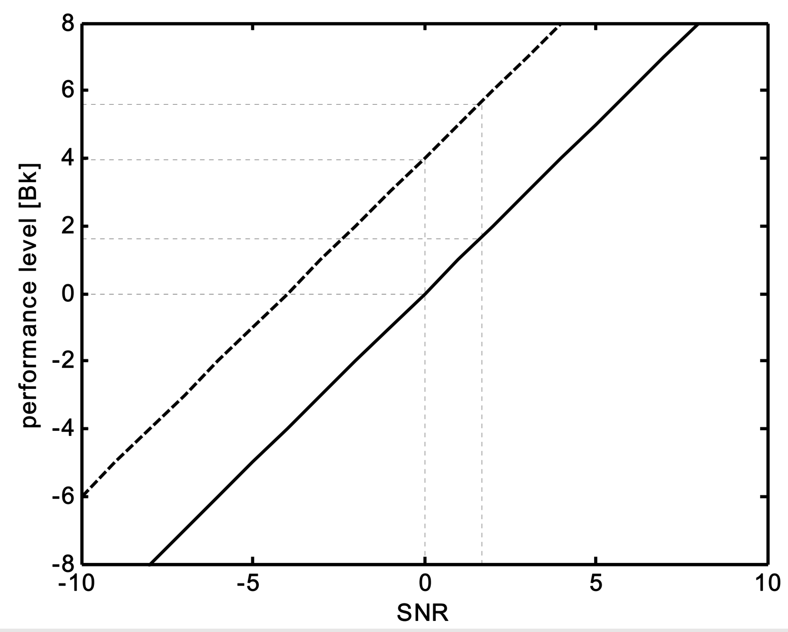
The regression equation estimates performance levels as a linear combination of effects resulting from the experimental conditions. With dummy coding, effects are expressed relative to a reference condition (here A_1B_1).

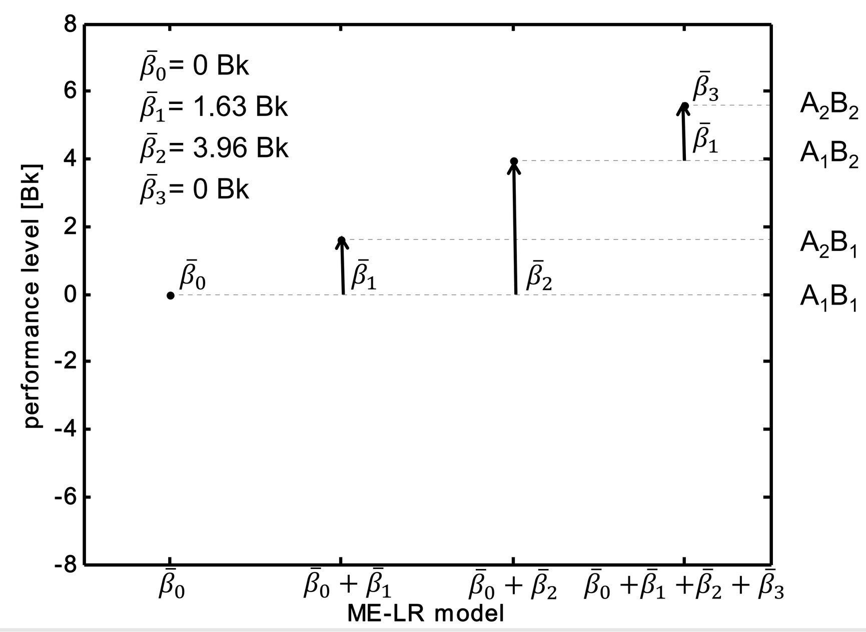
The n subscript in β_{in} indicates that this regression coefficient varies across subjects, and represents the random effect. Commonly a normal-distribution is assumed, and the mean ($\bar{\beta}_i$) and variance of this distribution are estimated. If this variance does not differ significantly from zero, the across-subject variance can be eliminated from the regression equation. Besides subjects one can include additional random factors, e.g. the speech items, to compensate and inspect item difficulty.



Correct interactions

Assuming that the PMF has a logistic shape, this function becomes linear after transform. Consequently, shifting the PMF among the horizontal axis can be translated into a shift along the vertical axis.





Significance testing

Each set of β_{in} parameters leads to a set of predicted performance levels $\widehat{PL}$. Knowing that the data follow a binomial distribution, one can calculate the probability of finding the data given the $\widehat{PL}$s, a quantity known as likelihood. The β_{in} parameters are estimated by maximizing this likelihood. Their significance can be tested by bootstrapping; by Bayesian methods; or by looking at changes in the likelihood while adding or removing predictors variables from the regression equation. Significant predictors should be kept.

Further reading

Demidenko, E. (2013) Mixed Models: Theory and applications with R. (John Wiley and Sons, Hoboken, NJ).
Goldstein, H. (1995). Multilevel Statistical Models. (Arnold, London, UK).
Hilkhuysen, G., Gaubitch, N., Brookes, M., and Huckvale, M. (2012). "Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios," J. Acoust. Soc. Am. 131, 531–539.
Hox, J. (2010). Multilevel Analysis: Techniques and Applications. (Erlbaum Associates, Mahwah, NJ).
Max, L., and Onghena, P. (1999). "Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research," J. Speech Lang. Hear. Res. 42, 261–270.
Quené, H., and van den Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: a tutorial," Speech Commun. 43, 103–121.
Rasbash, J., Steele, F., Browne, W., and Prosser, B. (2005) A user's guide to MLwiN. (Centre for Multilevel Modeling, Bristol, UK).
Wu, L. (2010) Mixed Effects Models for Complex Data. (CRC Press, London, UK).