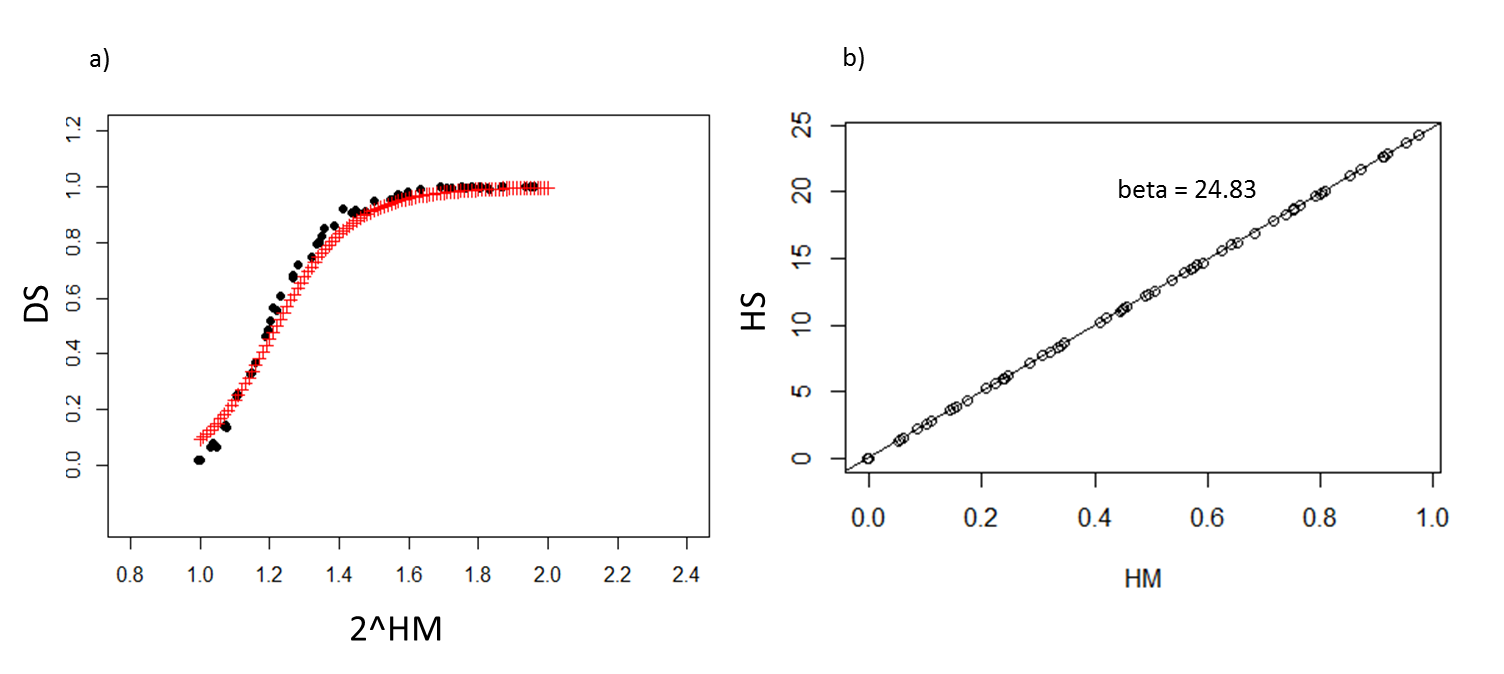
# The relationship between HS and HM

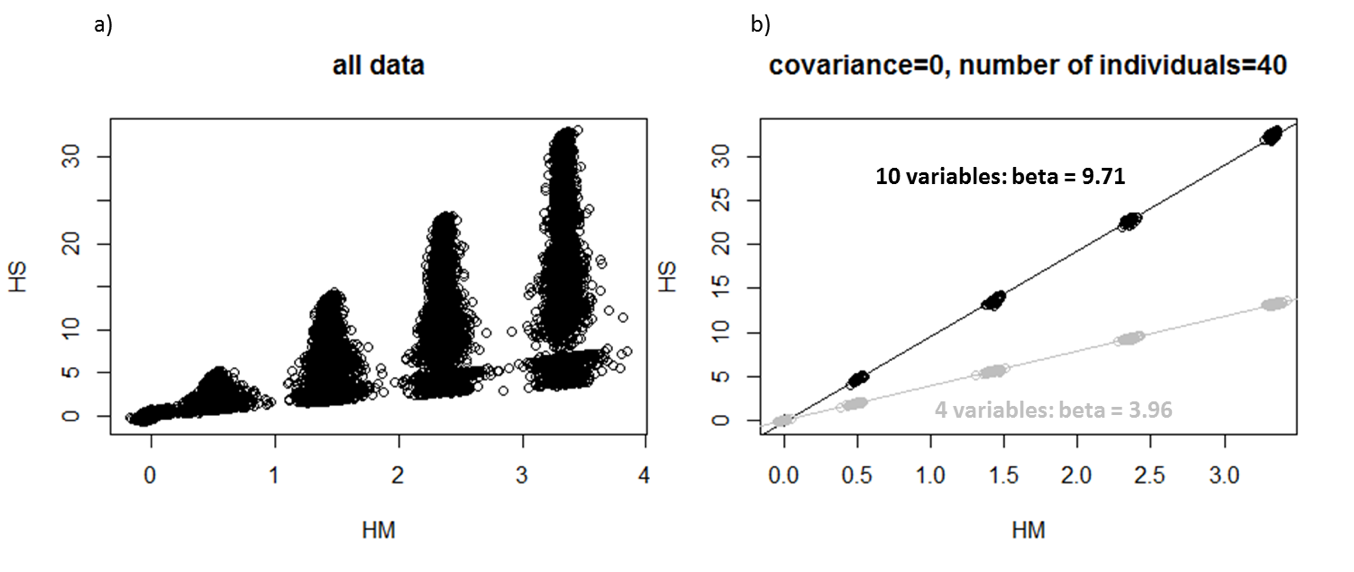
Searby and Jouventin (2004) suggested that HM should be favored over HS in case of modulated signals. However, they only presented the relationship between HM and discrimination score and did not directly compare the performance of HS and HM. Moreover, their artificial dataset was limited in the number of manipulated parameters. We were therefore interested in better understanding the relationship between the HM and HS in our data and whether HM could have advantages over HS.

We first tried to replicate results of Searby and Jouventin (2004) but included the calculation of HS. The process and functions necessary to replicate the data are described in the file “replicate Searby Jouventin 2004.R” within the provided R project (Linhart, 2018). We were able to replicate the results of Searby and Jouventin (2004) quite well (Figure S6.1a). HS was almost identical to HM \* 25 (Figure S6.1b). Because the dataset comprised 25 identity variables, HM could be potentially seen as the average HS per identity variable. We further tested this possibility in our datasets (see below). We found no advantage of using HM over HS for modulated signals as the two metrics are perfectly correlated in the Searby and Jouventin (2004) replicated dataset (R2 = 1.00, Figure S6.1b).



**Figure S6.1**. Relationship between discrimination score (DS) and 2^HM as reported by Searby and Jouventin (2004) and our replication of the same data (a) – red crosses show the line predicted by Searby and Jouventin (2004, Fig 2b). Relationship between HS and HM in replication of data from Searby and Jouventin 2004 (b).

However, in our dataset, this perfect relationship can became blurred because of changes in covariance (HS values are lower than expected from HM) and insufficient sampling (in cases when there are few individuals, HS is underestimated) (Figure S6.2a). When the covariance is set to zero and with sufficient sampling of individuals (full set of 40 individuals) HS is again almost equivalent to HM \* number of variables (Figure S6.2b). We suggest that HS should be preferred over HM because HM requires additional information about the dimensionality of the data which would need to be assessed separately for non-independent variables.



**Figure S6.2.** Relationship between HM and HS in our all simulated datasets for all data pooled (a) and for selected datasets with covariance=0 and number of individuals=40 (b).

**References:**

Linhart, P. (2018). *pygmy83/Identity-metrics: Identity metrics*. Zenodo. https://doi.org/10.5281/zenodo.1252272

Searby, A., & Jouventin, P. (2004). How to measure information carried by a modulated vocal signature? *Journal of the Acoustical Society of America*, *116*, 3192–3198. https://doi.org/10.1121/1.1775271