Making a group distribution from individual distributions

Denis Cousineau, Université de Montréal
Yves Lacouture, Université Laval

- This talk presents a quantitative method to let the data speak

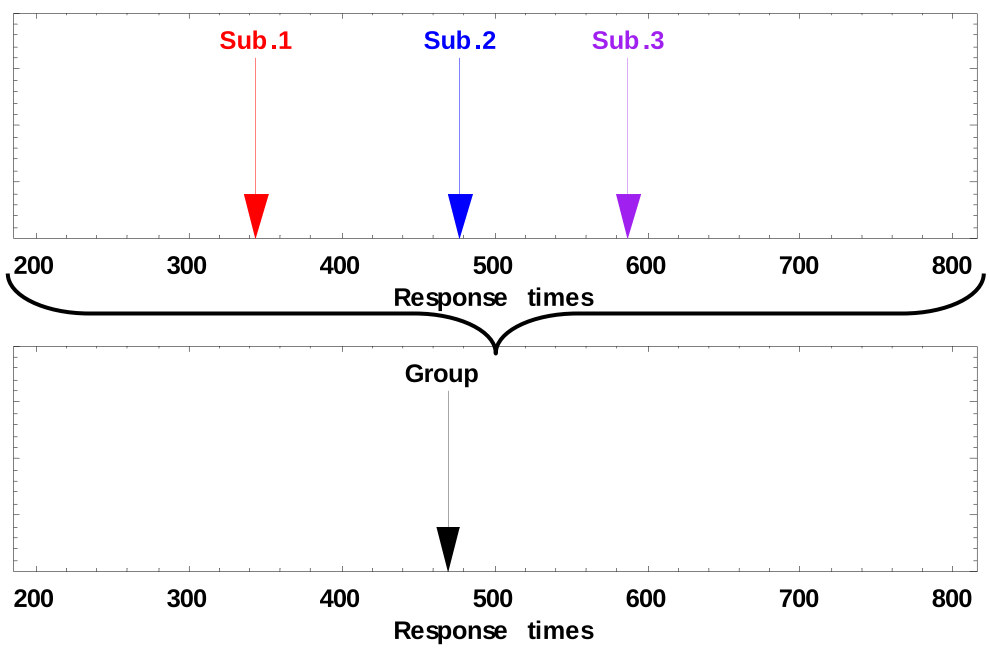
- Speak what language?

- The "average" language
  - to obtain a group picture, where the group is the average of the subjects
  - to allow for inferences on <u>a group</u> of subjects.

Various aspects of the data (1/3)

- The group position (central tendency) is the average of the individual subjects' positions

n = 3 in this example



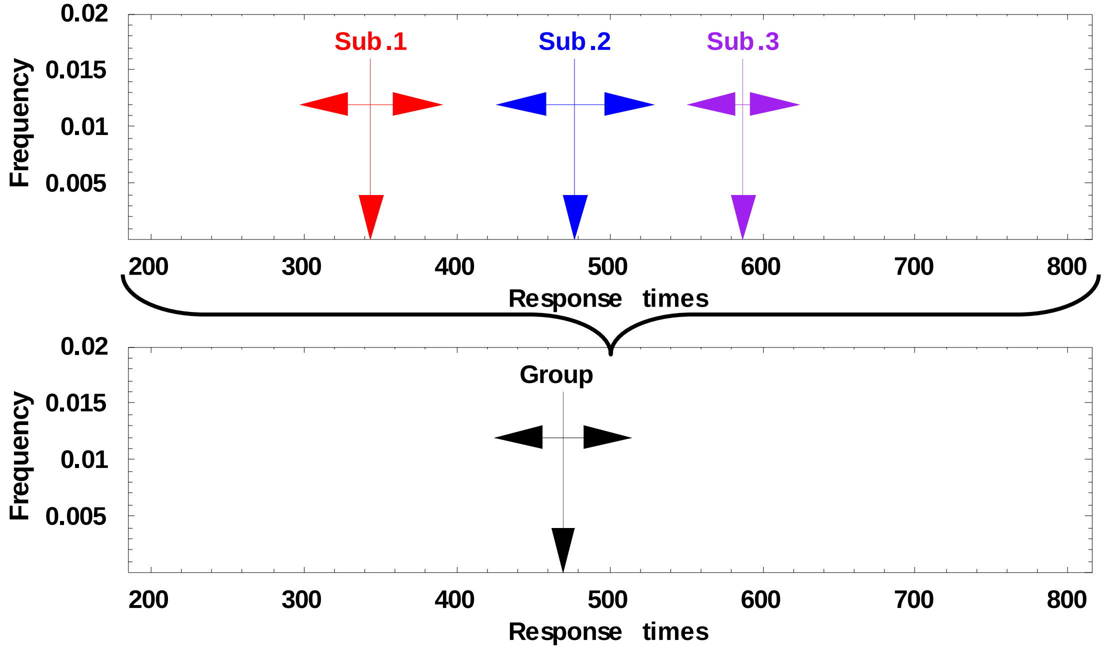
Here, "average" is the **<u>mean</u>**

Why look at the average position?

- The dependant variable could be response times (RTs)
- Visual attention: examine the mean RTs as difficulty is increased
- Language processing: look at the change in mean RTs as the number of letters increases
- etc, etc, etc, etc, etc, etc, etc, etc, etc, etc

Various aspects of the data (2/3)

- The participants have a standard deviation (SD).
- The group SD is the ____?___ of the individual SDs



Here, "average" is **not** the standard deviation or the mean

Why look at standard deviations?

- A theory of visual search (serial self-terminating model: Cousineau & Shiffrin, 2004) makes a strong prediction on the ratio:

$$\frac{\Delta SD}{\Delta MN} = 0.60$$

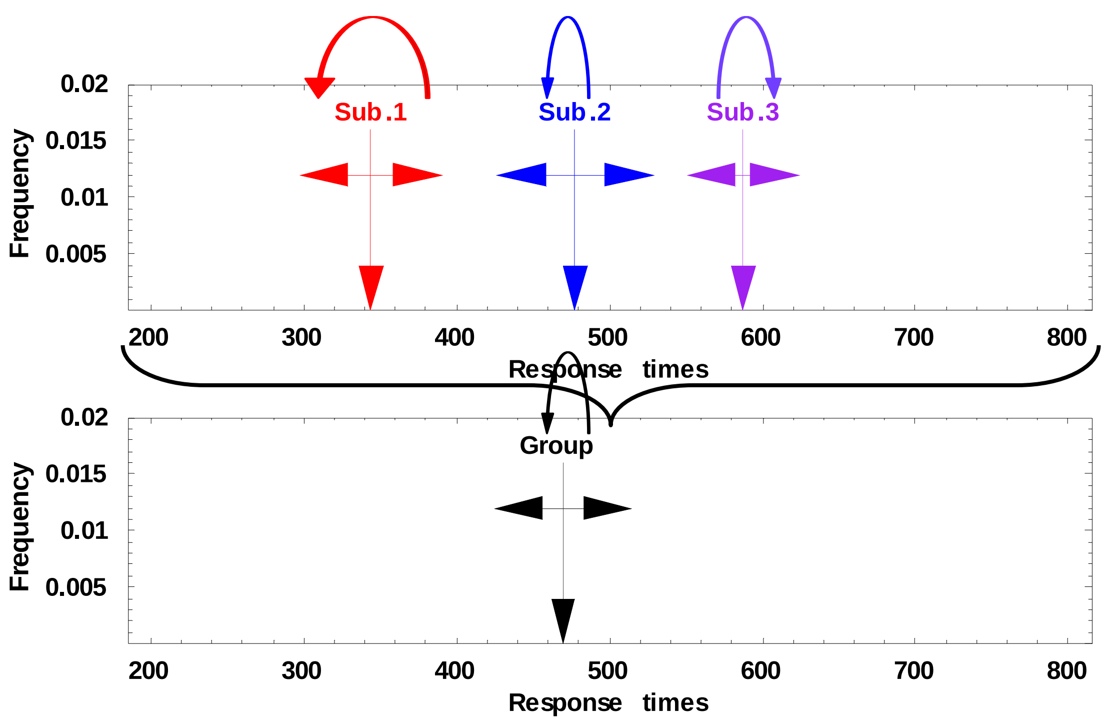
- A theory of language fluency (Segalowitz, 1998) predicts that:

$$CV = \frac{SD}{MN} \text{ is linear}$$

Various aspects of the data (3/3)

- The participants may have asymmetrical distributions of RTs; this is called the skew.

- The group skew is the ___?___ of the subjects' skew.



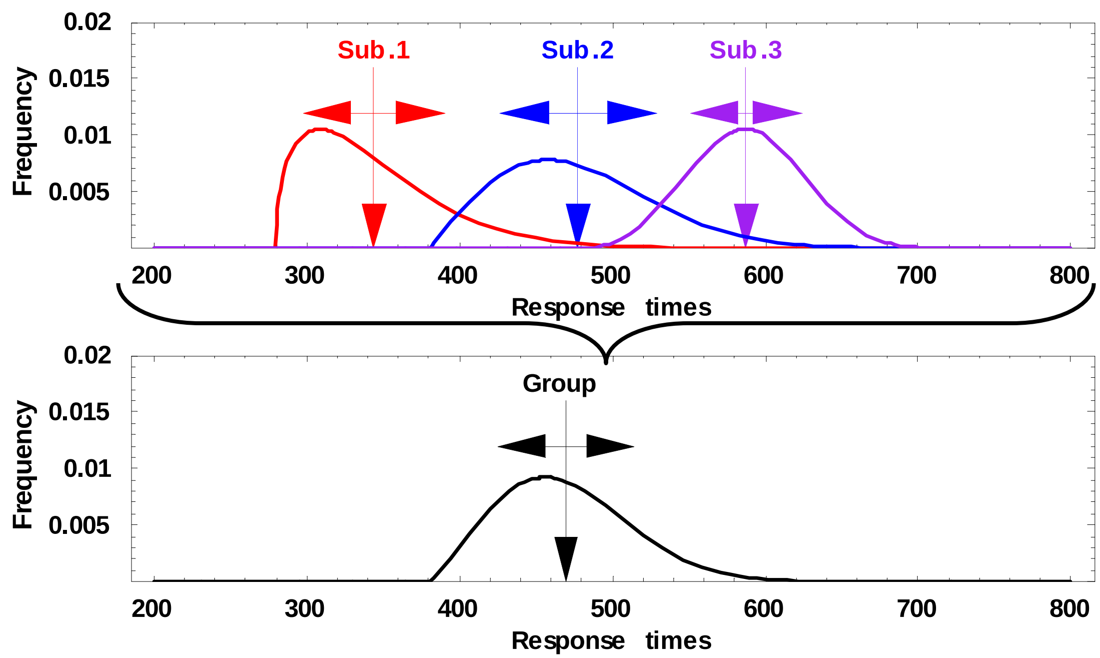
Why look at the skew?

- Theory of automaticity (Logan, 1988):
  - skew should be a constant throughout sessions of practice

- Race model (Cousineau, Goodman, & Shiffrin, 2003):
  - Skew is the “signature” of the neural architecture.
  - it should be a constant throughout sessions
  - its value indicates the properties of the network.

The solution?

- Instead of averaging summary statistics (mean, SD, skew), **<u>average the distributions</u>**
- then compute the summary statistics on the group distribution.



?

How do we average distributions? (1/2)

- According to Thomas and Ross' (1980) theorems:
  - pooling all the RTs won't do
  - vincentizing won't do either (except in restricted cases).

Then the arguments used in the proof of Theorem 1 show that, if Eqs. (5), (19), and (20) hold,

$$g_p(a, b) = a + c_p b.$$

The results follow on defining Φ by the equation

$$\Phi[h^{-1}(c_p)] = p.$$

Related to generalized Vincentizing is the *generalized Q–Q plot*, which is defined as the plot of $h(Q_{jp})$ against $h(Q_{ip})$. If we define z_p as the solution of $\Phi(z_p) = p$, it can be seen from Eq. (21) that

$$h(Q_{ip}) = h(z_p)\, h(\beta_i) + h(\alpha_i),$$

which implies that

$$h(Q_{jp}) = \frac{h(\beta_j)}{h(\beta_i)} h(Q_{ip}) + h(\alpha_j) - \frac{h(\beta_j)}{h(\beta_i)} h(\alpha_i).$$

h is a transfo that linearize Q-Q plot h(RT) vs h(RT)

That is, $h(Q_{jp})$ and $h(Q_{ip})$ are linearly related. The argument goes the other way too, as in the proof of Theorem 2, to establish:

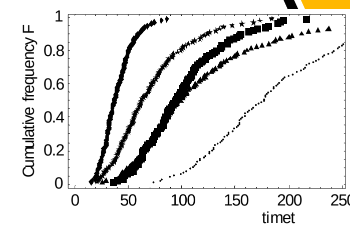
THEOREM 4. *Let h be a continuous, monotonic function. The plot of the transformed order-p quantile, $h(Q_{ip})$, from the ith distribution against $h(Q_{jp})$, as p varies from 0 to 1, is linear for all $i, j \leqslant n$ if and only if the ith d.f., $F_i(t)$, is of the form*

$$F_i(t) = \Phi\left\{h^{-1}\left[\frac{h(t) - h(\alpha_i)}{h(\beta_i)}\right]\right\}.$$

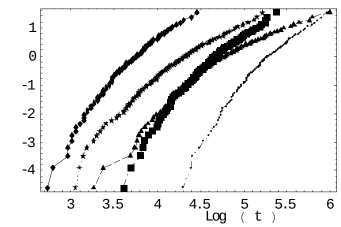
In practice, h is likely to be unknown. One can search for a monotonic transformation such that the generalized Q–Q plots are linear and, if the search turns out to be successful, one would have established the validity of Eq. (21). To make such a search feasible,

How do we average distributions? (2/2)

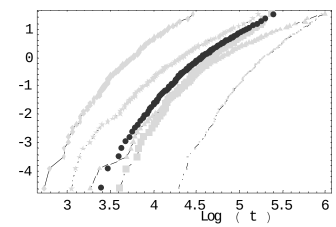
① **Obtain the individual cumulative distributions (CDF) *shifted to zero*;**



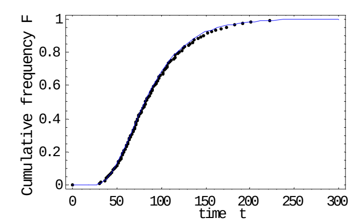
② **Linearize them (using log);**



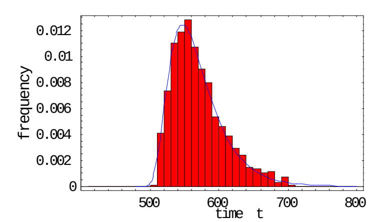
③ **Find the average of the lines;**



④ **Undo the linearization;**



⑤ **Sample the obtained distribution.**



- This graphical method is based on analytical arguments (I spare you the details…)

- They show that:
  - The group mean is the mean of the individual means
  - The group SD is the <u>geometric mean</u> of the individual SDs
  - The group skew is the mean of the individual skews.

- Use the above as shortcuts if you are not interested in the group distribution.

- Simulations where required…
  (I spare you the details)

  because the difficulty is to shift the individual distributions by the right amount.

- They show that even poorly-estimated shifts do not distort the estimated group distribution

  ➔ the method is robust.

- Thank you.

  - This talk is available at http://mapageweb.umontreal.ca/cousined