

Computational Cognitive Science

Lecture 7: Aggregation across Participants

Frank Keller

12 October 2026

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Slide credits: Chris Lucas, Benjamin Peters

The Effect of Averaging

Modeling and Data Aggregation

Fitting Aggregate Data

Fitting Individual Participants

Modeling Differences between Participants

Reading: Chapter 5 of [Farrell and Lewandowsky \(2018\)](#).

The Effect of Averaging

The Effect of Averaging

Admission data from UC Berkeley in 1973:

- 8,442 men applied; 44% were admitted.
- 4,321 women applied; 35% were admitted.

At the aggregate level, women appear substantially less likely to be admitted.

Does this mean that individual departments discriminated against women?

The Effect of Averaging

Admission data for four UC Berkeley departments:

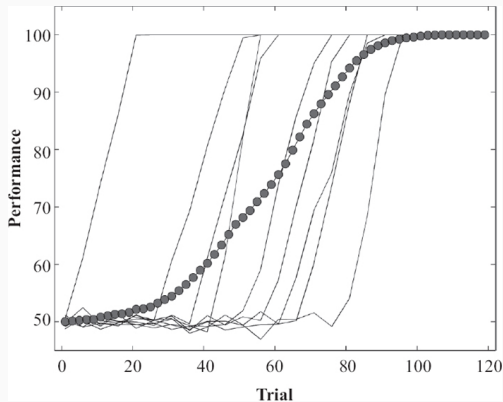
Department	Men		Women	
	Applied	Admitted	Applied	Admitted
A	825	511 (62%)	108	89 (82%)
B	560	353 (63%)	25	17 (68%)
C	325	120 (37%)	593	225 (38%)
D	191	53 (28%)	393	114 (29%)

Within each department, women were admitted at a similar or higher rate than men.

Simpson's paradox: aggregation can reverse the pattern present in the underlying groups.

The Effect of Averaging

Averaging of learning curves generated by a model:



- not a single individual learns in the manner suggested by the average curve
- individual learning is abrupt/linear, but begins at different times; averaging creates a smooth nonlinear curve.

Modeling and Data Aggregation

Modeling and Data Aggregation

When building models, we need to decide whether to aggregate the data. We can assess and fit models with:

1. Summary statistics (like averages)
2. Data merged across individuals
3. Individual-level data
4. Groups or clusters of individuals

Why aggregate?

- Less work for you
- Less computationally expensive
- Usually implies simpler models
 - Less risk of overfitting with small data sets
 - Easier to communicate
- Sometimes the only realistic option (e.g., few data points per participant)

Why avoid aggregating?

People aren't the same:

- Different strategies
- Different expectations
- Motivation, memory, ...

Pretending they are the same can:

- Mask interesting patterns
- Lead to spurious conclusions

Groups: Splitting the difference

- Accommodate differences
- Not as data-intensive as separate individual analyses
- Evidence for clusters may be scientifically interesting

Fitting Aggregate Data

Fitting aggregate data: Response times

1. Fitting summary statistics of aggregate data
 - Typically easiest, with greatest downsides
 - Loses a great deal of information
2. Pretend everyone is the same
 - Common model, parameters, etc.
 - If assuming data are already conditionally independent, just lump them together

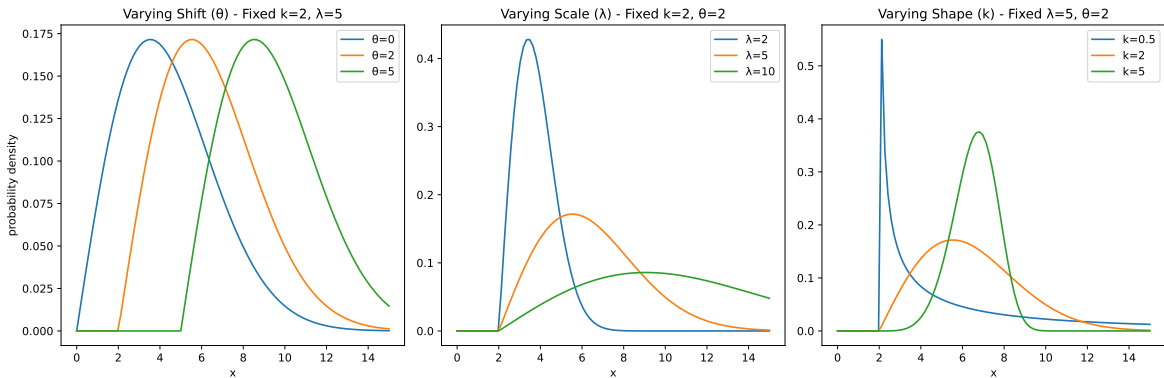
Example: Response times

Suppose we want to characterize people using a shifted Weibull distribution.

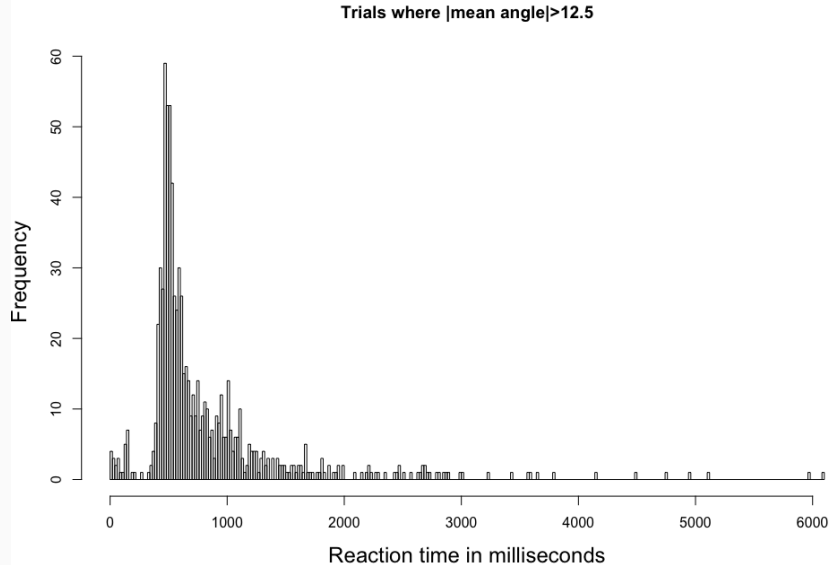
- Models an accumulator like the random walk
- Difference: “race” approach – first accumulator past the post
- Parameters: shift (θ), scale (λ), shape (k)

Example: Response times

Variations of shift, scale, shape of the shifted Weibull distribution (see tutorial 2):



Aggregate response times



Aggregate response times: Vincentizing

Pooling response times across participants can distort the shape of the underlying distributions.

Vincentizing preserves more distributional information:

1. Compute the same quantiles for each participant (e.g., .1, .3, .5, .7, .9).
2. Average corresponding quantiles across participants.
3. Fit the model to these averaged quantiles.

This gives a single group-level distribution while retaining information about the shapes of the individual distributions.

Aggregate response times: Vincentizing

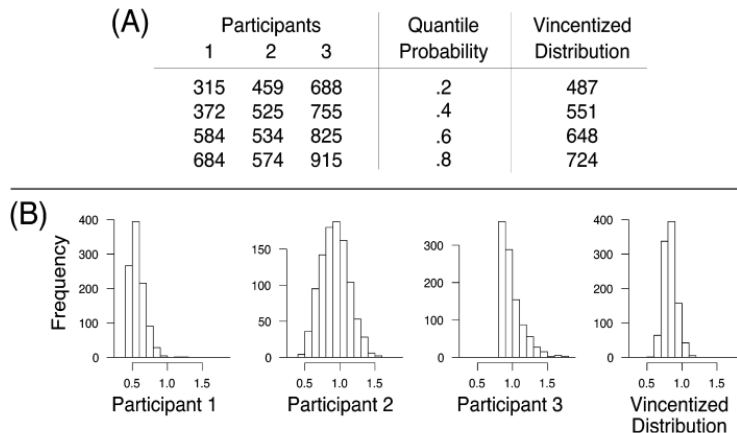


Figure 1. Examples of Vincentizing. Panel A shows how quantiles are averaged. Panel B shows that Vincentized distribution is an “average” of the individual distributions.

Aggregate response times: Vincentizing

- Better preserves distributional shape than simply pooling all RTs.
- But Vincentizing is **not guaranteed to give consistent parameter estimates** for every RT distribution.
- For the shifted Weibull example, it performs well when individual samples are small.

Time for a short quiz on WooClap!

Go to wooclap.com

Enter the event code in the top banner

Event code

OONNLCJ



Fitting Individual Participants

Fitting individual participants

Instead of aggregating, fit the model separately to each participant:

$$\hat{\theta}_i = \operatorname{argmax}_{\theta_i} P(\mathbf{y}_i | \theta_i)$$

- Obtain a separate set of parameter estimates for each person.
- Preserve variability between participants.
- Use the resulting parameter estimates for population-level inference.

Why not always fit individuals?

Advantages:

- Reveals participant-level variability.
- Allows us to relate model parameters to individual differences.

Limitations:

- Individual participants may contribute little data.
- Parameter estimates may therefore be noisy or unstable.
- Separate fits do not exploit similarities between participants.

This motivates methods that model **structure across participants**.

Modeling Differences between Participants

Discrete differences between participants

Sometimes participants may fall into distinct groups or use different strategies.

Two related approaches:

- **Mixture models**: model data as arising from several latent generating processes.
- **K-means clustering**: group similar participants based on their observed behavior.

Mixture components may correspond to different participant groups, but can also represent different processes within a participant.

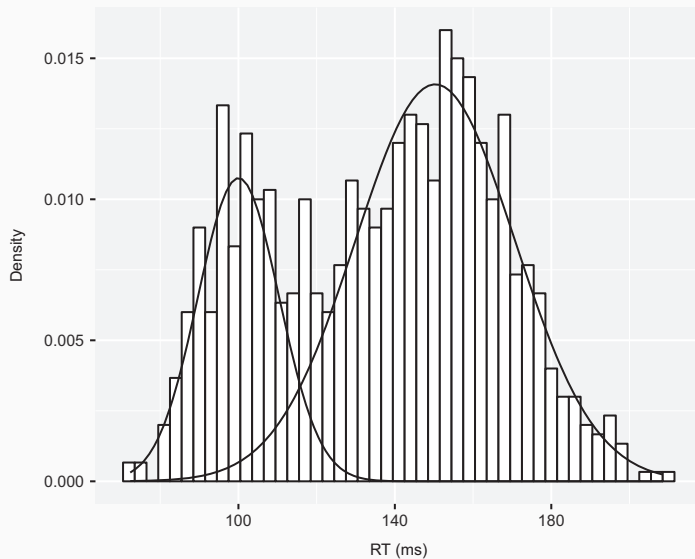
These methods are useful when individual differences are **discrete rather than continuous**.

For two components:

$$P(x) = \pi P(x|\theta_1) + (1 - \pi)P(x|\theta_2)$$

- Each observation is generated by one component.
- Component membership is not directly observed.
- Parameters and membership probabilities can be estimated iteratively, e.g., with Expectation–Maximization (EM).

Example: Saccadic RT distribution with two Gaussian components



K-means clustering

K-means assumes observations belong to one of *K* clusters.

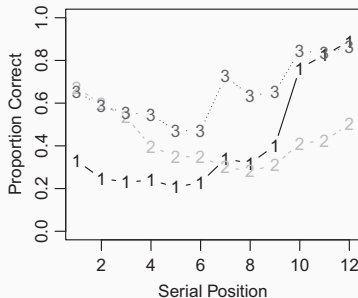
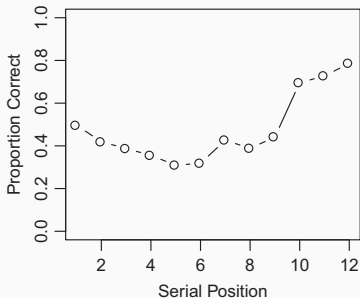
1. Initialize *K* cluster centroids.
2. Assign each observation to the nearest centroid.
3. Recompute the centroids.
4. Repeat until assignments stop changing.

The objective is to minimize within-cluster variation:

$$\sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2$$

Example: recall strategies

Average serial-position curve and clusters identified by *K*-means:



Averaging can hide qualitatively different patterns:

- predominantly recency;
- predominantly primacy;
- both primacy and recency.

Continuous individual differences

Not all differences form discrete groups.

Instead, model parameters can be treated as measures of **continuous psychological differences**.

- Fit the cognitive model to each participant.
- Use fitted parameters as measures of underlying cognitive processes.
- Relate these parameters to other measures of ability or behavior.

This approach has been called **cognitive psychometrics**.

Example: diffusion model and cognitive ability

Schmiedek et al. (2007) fitted a diffusion model to choice-RT tasks.

Key parameters:

- drift rate v : rate of information accumulation;
- boundary separation a : amount of evidence required;
- non-decision time T_{er} : perceptual/motor processes.

They then examined how these parameter estimates related to individual differences in cognitive ability.

Main result: drift rate showed the strongest relationship with working-memory performance.

Summary

- **Aggregation can be misleading:** averages may obscure or even reverse patterns present in individual data.
- When fitting models to multiple participants, we can:
 - aggregate the data;
 - fit each participant separately;
 - model groups or clusters of participants.
- **Vincentizing** can preserve distributional shape better than pooling observations, but it is not appropriate for every RT model.
- Fitting individuals preserves participant-level variability, but parameter estimates can be noisy when data are sparse.
- Individual differences can be:
 - **discrete:** mixture models, K -means clustering;
 - **continuous:** relate fitted model parameters to differences in ability or behavior.

References

- Farrell, Simon and Stephan Lewandowsky. 2018. *Computational Modeling of Cognition and Behavior*. Cambridge University Press, Cambridge.
- Rouder, Jeffrey N. and Paul L. Speckman. 2004. An evaluation of the vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review* 11(3):419–427.
- Schmiedek, Florian, Klaus Oberauer, Oliver Wilhelm, Heinz-Martin Süss, and Werner W. Wittmann. 2007. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General* 136(3):414–429.