A Template for Describing Individual Differences in Repeated Measures Data, with Application to the Connection between Learning and Ability

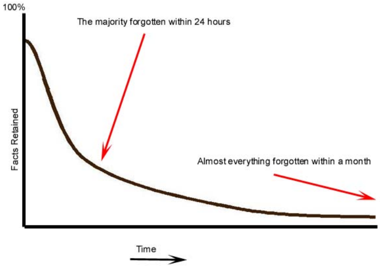
Bob Cudeck and Casey Codd
Ohio State University

June 18, 2010

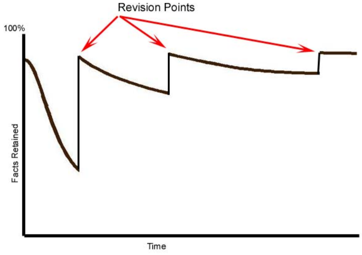
Learning, Forgetting, Reminiscence

From decades of experimental research, it is well known that learned material is forgotten very quickly unless it is rehearsed

Rapid and immediate forgetting is almost universal



The Curve of Forgetting

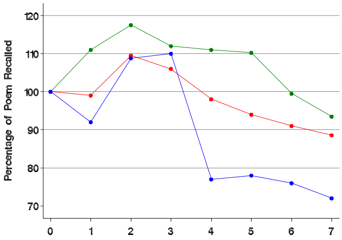


The Curve of Learning

Learning, Forgetting, Reminiscence

Ballard, 1913, a classic study

- 5,192 school children learned lines from "Rhyme of the Ancient Mariner"
- Immediate recall: Children wrote as many lines as they could remember
- Delayed recall: Re-tested one to seven days later
- Approx 50% of the sample recalled more after several days than they did at the initial test

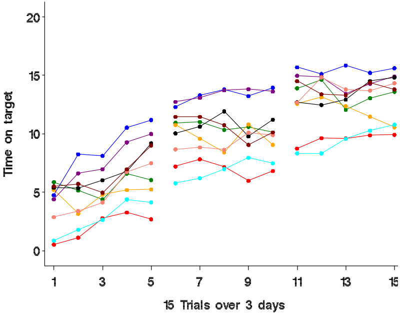


Percentage of lines of poem recalled,
1 - 7 days after initial learning
Scaled so pre-test is 100%

Reminiscence - unrehearsed improvement in memory after a delay

Learning, Forgetting, Reminiscence

- Ballard's demonstration of reminiscence effects sparked wide interest
- Pervasive effect
  - Found using nonsense syllables, 3-letter digits, abstract words, prose, pictures of objects and people
  - Occurs in younger and older learners, in depressed and nondepressed samples, high and low IQ, extroverts.



In this project: investigate reminiscence effects in the pursuit rotor task

Each day, 25 trials of 20 seconds, summarized into 5 trial blocks, 5 blocks per day

Some Subjects are Reminiscers, Some are Not

Reminiscence is ... Acquired skill? Inborn ability? Personality trait? A fluke?

- Can it be taught?
- Which characteristics are associated with it?

Most studies investigating reminiscence are based on comparison of means

- Mean comparisons are a problem because reminiscence is a characteristic of individuals who either improve without active practice or do not
- Averages conceal the strong individual differences in the effect

RCM in the Study of Individual Learning and Change

The random coefficient model RCM is ideal for individual differences in learning

Most important feature of the RCM:
It is a subject-specific model focused on individual patterns of change

- Gracefully handles diverse measurement designs, functional forms, between- and within-subject variability
- Major improvement over statistical models that are mean-focused where individual differences are averaged over

Aim 1 of this work: Use RCM to describe individual learning and study reminiscence in pursuit rotor data

RCM in the Study of Individual Learning and Change

Considerable enthusiasm for the RCM because of the subject-specific focus

- However in fact, subject-specific capabilities of RCM get short shrift
  Def'n short shrift: "brief and unsympathetic treatment"
- Strange practice:
  Praise model as ideal for studying individual change
  But in practical research ignore the individuals and analyze the means
- Why do we avoid studying individuals, even with a SS model?
  - Harder than analyzing means
  - Has more subjective decisions than does the analysis of means
  - But (according to all textbooks) significant return on statistical investment

Aim 2 of this work: Suggest a preliminary EDA-type procedure in conjunction with the RCM to study individual differences in repeated measures data

Thurstone (1919) "The Learning Curve Equation"

Presented a disciplined, preliminary EDA-type procedure

1. Consistent focus on individual learning functions
   Accumulated from individual analyses, generalized to broad themes
2. Examined many alternative functions, more than 40, before picking one
   Used a model that worked well with the learning data that he studied
3. Imposed strict criteria for inclusion and exclusion of participants
   83 students initially, but 32 (39%) excluded
   20 for missing sessions, 3 for variable responding, 9 unusual learning curve

Goal of the study was to describe a typical subject

- Identified representative learners who learned in a characteristic way

> *"We have eliminated 12 out of 63 complete records. Generalizing from this fact we may conclude that [function for this problem works] in about four cases out of five. This justifies our reference to it as the typical but not as the universal form of learning curve"*

Sample and Data

Thomas Bouchard's incomparable study of twins reared apart

Adult twins, plus some family members and close associates

- Separated early in life, reared in adoptive families, reunited in adulthood
- N = 176 adults, 63 men and 113 women, 18 to 77 years old

Data for this work selected from a larger archive

- Learning task: Standard pursuit rotor apparatus, 3 30 minute sessions, 5 trials over 3 days
- Covariates: 37 cognitive ability tests, used to explain individual difference in learning on the pursuit rotor task

Twin project primarily for behavior genetics; however sample has been used to investigate substantive questions in measurement and statistics

Selecting the Response Function

Goal is to represent individual learning on the pursuit rotor

- Initially, use EDA procedures to plot individual data and candidate functions in trellis displays
- Designed to work toward a full RCM
- In both phases, the function should provide a good fit for the majority of individuals

For EDA, use nonlinear least squares for "individual-specific regressions" (Davidian & Giltinan)

- Different functions work better for some individuals than others and there are lots of learning styles
- Goal is to decide on a function that is most appropriate for the largest number of individuals

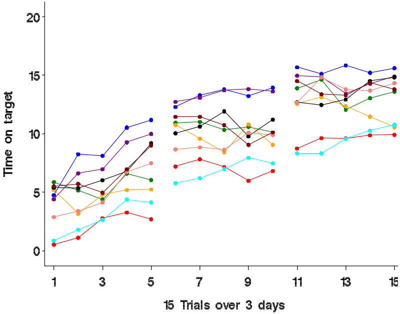
Modern statistical software makes it easy to fit the functions and do the graphs

Selecting the Response Function

Model for person i, occasion j, based on trial x_{ij} and parameters $\boldsymbol{\theta}_i$

$$\boxed{y_{ij} = f(\boldsymbol{\theta}_i, x_{ij}) + e_{ij}}$$

First four candidates ignore the structure of trials within days for simplicity



1 Linear

$$f_j = \beta_0 + \beta_1 x_j$$

2 Exponential

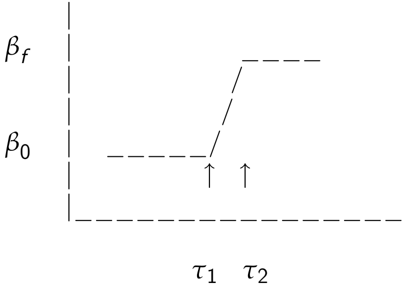
$$f_j = \beta_f (1 - \exp(-\eta x_j))$$

3 Logistic

$$f_j = \frac{\beta_0 \beta_f}{\beta_0 + (\beta_f - \beta_0)\exp(-\eta x_j)}$$

4 Three continuous linear segments with unknown knots

$$f_j = \begin{cases} \beta_0 & x_j \leq \tau_1 \\ \beta_0 + \dfrac{\beta_f - \beta_0}{\tau_2 - \tau_1}(x_j - \tau_1) & \tau_1 < x_j \leq \tau_2 \\ \beta_f & \tau_2 < x_j \end{cases}$$



Selecting the Response Function

5 Three discontinuous linear segments, one for each day

$$f_j = \begin{cases} \alpha_0 + \alpha_1 x_j & 1 \le x_j \le 5 \\ \beta_0 + \beta_1 x_j & 6 \le x_j \le 10 \\ \gamma_0 + \gamma_1 x_j & 11 \le x_j \le 15 \end{cases}$$

Define ρ_1 and ρ_2 to be the difference in performance between days 1 and 2, and between days 2 and 3. ρ_1 and ρ_2 are reminiscence effects under the model.

$$\begin{aligned} \rho_1 &= \beta_0 + \beta_1 x_1^* - (\alpha_0 + \alpha_1 x_1^*),\ x_1^* = 5.5 \\ \rho_2 &= \gamma_0 + \gamma_1 x_2^* - (\beta_0 + \beta_1 x_2^*),\ x_2^* = 10.5 \end{aligned}$$

Re-write the function to include ρ_1 and ρ_2.

$$f_j = \begin{cases} \alpha_0 + \alpha_1 x_j & 1 \le x_j \le 5 \\ \alpha_0 + \alpha_1 x_1^* + \rho_1 + \beta_1 (x_j - x_1^*) & 6 \le x_j \le 10 \\ \alpha_0 + \alpha_1 x_1^* + \rho_1 + 5\beta_1 + \rho_2 + \gamma_1 (x_j - x_2^*) & 11 \le x_j \le 15 \end{cases}$$

Selecting the Response Function

Individual-specific Regressions

$$y_{ij} = f(\boldsymbol{\theta}_i, x_{ij}) + e_{ij}$$

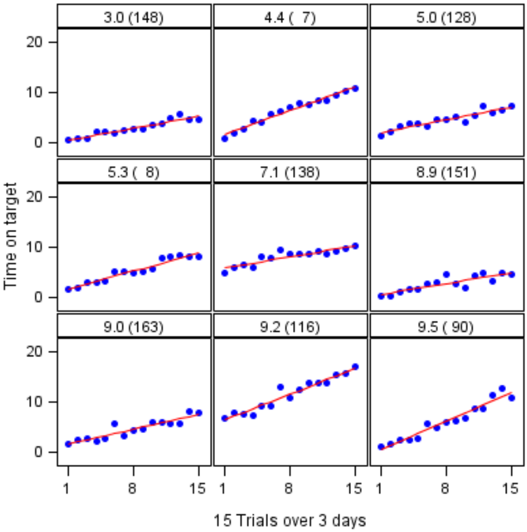
Fit functions to every individual by least squares and record whether the procedure converges

- If usable solution, then calculate individual mean square residual,

$$MSR_i = n_i^{-1} \sum_j (y_{ij} - f_{ij})^2$$

- Record overall measure of fit $mdn(MSR_i)$
- Judging success by number of individuals reasonably well fit often preferable to an overall goodness of fit test

Linear

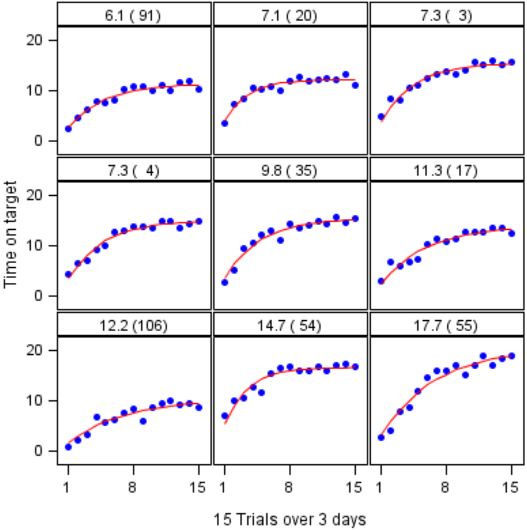


$$f_j = \beta_0 + \beta_1 x_j$$

Mdn(msr): 35.0

Solutions 176 / 176

Exponential

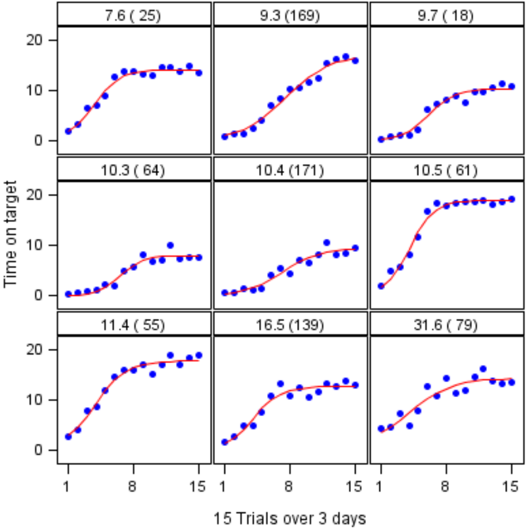


$$f_j = \beta_f(1 - \exp(-\eta x_j))$$

Mdn(msr): 26.9

Solutions 162 / 176

Logistic

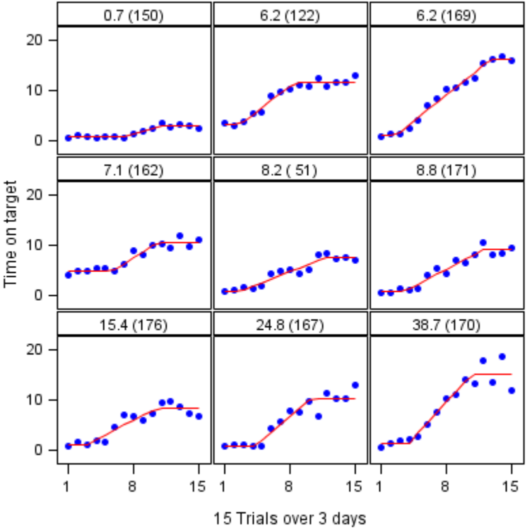


$$f_j = \frac{\beta_0 \beta_f}{\beta_0 + (\beta_f - \beta_0)\exp(-\eta x_j)}$$

Mdn(msr): 17.1

Solutions 170 / 176

Three-phase Spline, Continuous, Unknown Knots

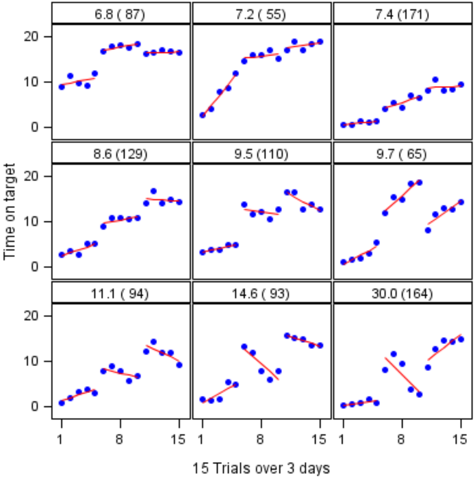


$$f_j = \begin{cases} \beta_0 \\ \beta_0 + \dfrac{\beta_f - \beta_0}{\tau_2 - \tau_1}(x_j - \tau_1) \\ \beta_f \end{cases}$$

Mdn(msr): 15.4

Solutions 128 / 176

Three-phase Spline, Discontinuous, Reminiscence Effects



$$f_j = \begin{cases} \alpha_0 + \alpha_1 x_j \\ \alpha_0 + \alpha_1 x_1^* + \rho_1 + \beta_1 (x_j - x_1^*) \\ \alpha_0 + \alpha_1 x_1^* + \rho_1 + 5\beta_1 \\ \qquad + \rho_2 + \gamma_1 (x_j - x_2^*) \end{cases}$$

Mdn(msr): 7.3

Solutions 176 / 176

Which Function?

Function (5) seems best among contenders

- From graphs it is appropriate for large majority
- Converged solution for all and had best $mdn(MSR_i)$
- Includes parameters for reminiscence effects

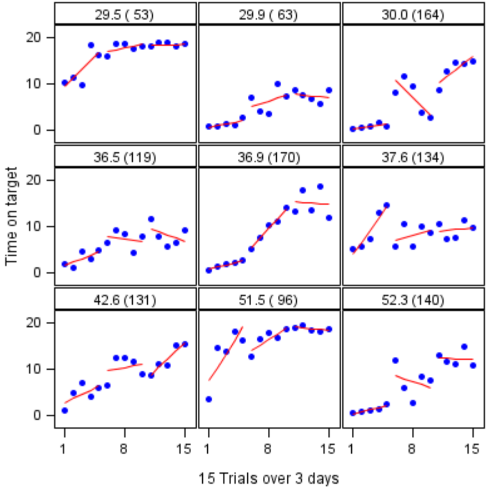
NB: Function (5) is not universally superior
Each function works best for some

As always, tremendous individual differences in learning patterns

Function	# parms	N*	MSR
(1) Linear	2	176	35.0
(2) Exponential	2	162	26.9
(3) Logistic	3	170	17.1
(4) 3 segments, contin.	4	128	15.4
(5) 3 segments, discontin	6	176	7.3

Poorest Fits to (5)

Poor fits often diagnostic about whether a function is really appropriate
Anomalies? Outliers? Latent classes? Different functions?



#164: Strong reminiscence effects, deterioration in active practice

#53: 10 second improvement after trial 3, then near-perfect performance

#170: Beautiful linear improvement, then inconsistent performance

Overall, relatively few poor fits, and no major surprises

RCM for the Rotor Pursuit Data

Full RCM model has features unavailable with individual-specific regressions

- But still want to maintain subject-specific perspective
- Function (5) is nonlinear in form but linear in the random effects, so the RCM setup is straightforward
- With $\boldsymbol{\theta}_i = (\alpha_{i0}, \alpha_{i1}, \rho_{i1}, \beta_{i1}, \rho_{i2}, \gamma_{i1})$, model for an individual is

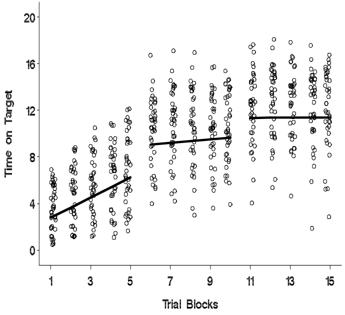
$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\theta}_i + \mathbf{e}_i$$

$$\{\mathbf{X}\}_{j\cdot} = \begin{cases} (1,\ x_{ij},\ 0,\ 0,\ 0,\ 0) & 1 \leq x_{ij} \leq 5 \\ (1,\ x_1^*,\ 1,\ (x_{ij} - x_1^*),\ 0,\ 0) & 6 \leq x_{ij} \leq 10 \\ (1,\ x_1^*,\ 1,\ 5,\ 1,\ (x_{ij} - x_2^*)) & 11 \leq x_{ij} \leq 15 \end{cases}$$

Standard assumptions: $\boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}, \Phi_{\theta\theta})$ and $\mathbf{e}_i \sim N(\mathbf{0}, \Psi_e)$, Ψ_e is diagonal.

Estimates of the RCM

- $\hat{\alpha}_1 = 0.86$, $\hat{\beta}_1 = 0.15$: Improvement under practice day 1 and day 2
- $\hat{\rho}_1 = 2.32$, $\hat{\rho}_2 = 1.61$: Reminiscence effects after day1 and day2
- $\hat{\gamma}_1 = 0.01$, $V(\gamma_{i1}) = 0.176$: Day 3 n.s., but individual slopes vary
- Plot of typical values, a subject with coefficients at estimates



MLE of function parameters

Parameters of Function (5)		Coefficient Variances ($\Phi_{\theta\theta}$)	
$\hat{\alpha}_0$	2.79(.15)	α_{i0}	4.64(.47)
$\hat{\alpha}_1$	0.86(.04)	α_{i1}	0.351(.04)
$\hat{\beta}_1$	0.15(.04)	β_{i1}	0.167(.02)
$\hat{\gamma}_1$	0.01(.04)	γ_{i1}	0.176(.03)
$\hat{\rho}_1$	2.32(.19)	ρ_{i1}	6.38(.76)
$\hat{\rho}_2$	1.61(.19)	ρ_{i2}	6.15(.78)

Latent Variable Covariates

Learning the pursuit rotor is not a unitary process

- Lots of complicated individual differences in learning style
- Is learning this laboratory task an independent, distinct skill?
- Are any specific abilities associated with it?
- Can any variable account for IDs on reminiscence? (a 100 year old question)

Now include in 5 latent variables to try to explain learning on this task

- Memory, Spatial, Object Rotation: Seemingly relevant to skill acquisition
- Verbal and Quantitative: Generally important in ability research
- Age as a single MV: Because pursuit rotor is a performance task

	Latent Variables					
	Memory	Spatial	Verbal	Quantitative	Rotation	Age
MVs	4	3	3	3	4	1

Sketch of Expanded Model

Factor analysis model for ability tests

$$\mathbf{z}_i = \Lambda \mathbf{f}_i + \mathbf{d}_i \qquad \mathbf{f}_i \text{ factors} \qquad \mathbf{d}_i \text{ residuals}$$

RCM for repeated measures on the learning task

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\theta}_i + \mathbf{e}_i \qquad \boldsymbol{\theta}_i \text{ coefficients} \qquad \mathbf{e}_i \text{ residuals}$$

Typically with the RCM, covariates make up Level 2 regression

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \Gamma \mathbf{f}_i + \mathbf{u}_i$$

In exploratory studies, often best to simply examine the covariance matrix

$$\Phi_{f\theta} = cov(\mathbf{f}_i, \boldsymbol{\theta}_i')$$

Connecting Learning and Ability

- Some moderate correlations between intercept and slope of day 1
- No connection to slopes on day 2 and 3
- No evidence that reminiscence is correlated with traditional abilities.

Correlations between ability LVs and learning the pursuit rotor

	Regression Coefficients					
Ability Factors	α_{i0}	α_{i1}	β_{i1}	γ_{i1}	ρ_{i1}	ρ_{i2}
Memory	.26	.01	.13	-.22	.15	.14
Spatial	.30	.30	.06	-.03	.23	.05
Verbal	-.01	.04	-.07	-.03	.06	.07
Quant	.01	.13	.06	.06	.13	-.02
Rotation	.40	.25	.11	.05	.16	.00
age	-.36	-.34	-.11	-.02	-.23	.05

Summary

Aim 1: Substantive - Study connections between ability and learning pursuit rotor

- Topics like this are a major activity in differential psychology, and representative of similar research enterprises
  - General ability and performance in school
  - General health and specific disease progression

For the ability/learning connection, only modest payoff on this aim

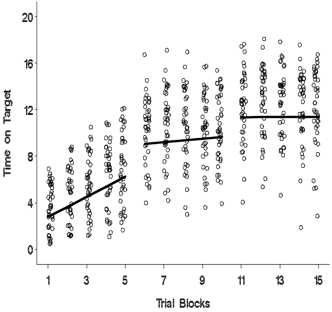
- Medium correlation between initial learning and spatial, image rotation and age
- If laboratory learning is not a component of general ability, what is it?
- If reminiscence is not an ability, what is it? It has significant IDs.
- Maybe other covariates - personality characteristics, motivational factors - would be more predictive

At the least, making reminiscence part of a function is a better way to measure it than just subtracting scores between adjacent periods.

Summary

Aim 2: Methodological - Recommend an EDA-type procedure together with the RCM to study individual differences in repeated measures data

Subject-specific capabilities of the RCM and rich individual differences tradition in the social sciences have not been investigated much.



The function of the typical values shows a beautiful learning pattern over days

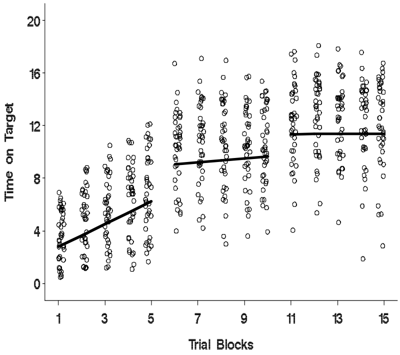
D_1: Practice and improvement
D_1: Effortless (magic) boost from reminiscence

D_2: More practice, more improvement
D_2: Another boost from reminiscence

D_3: No additional benefit from practice

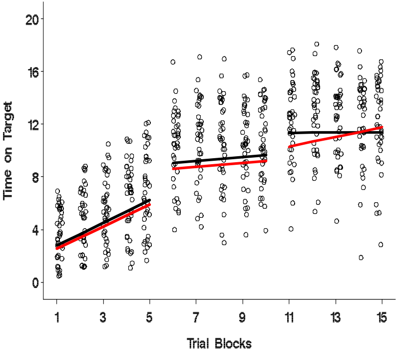
Summary

But in the sample of N = 176, how many individuals have a function like those based on the typical values?



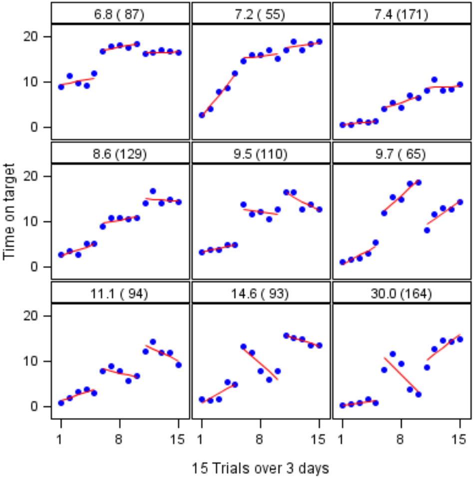
Summary

But in the sample of N = 176, how many individuals have a function like those based on the typical values? Perhaps a few. And the curve of the typical values is just one "person's" curve.



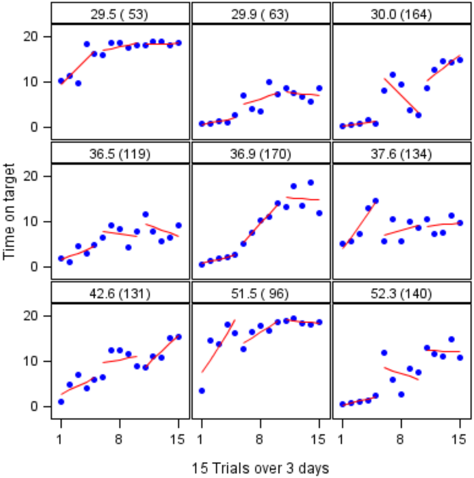
Summary

Always there are many other interesting patterns such, as these . . .



Summary

. . . and these



Summary

- If research with a subject-specific models is to pay off, we should do business differently than we did with statistics that work on the mean profile
- At the least we should be taking advantage of the new kind of information available in models like the RCM
- Thurstone's study on learning equations is a kind of template

◊ Approach the problem at the individual level

† For example, want a function that works well with majority of individuals

† Not the same as saying the function has the best overall AIC or LnL

◊ Try several response functions - linear misses a lot of valuable science

◊ Consider criteria for inclusion and exclusion to homogenize the sample

◊ Trellis displays can be hugely helpful