On your fingers:
Simple ways of assessing simulation-observation fit

Warren Thorngate
Psychology Department, Carleton University
Ottawa, Ontario K1S 5B6 Canada
email = warren.thorngate@rogers.com

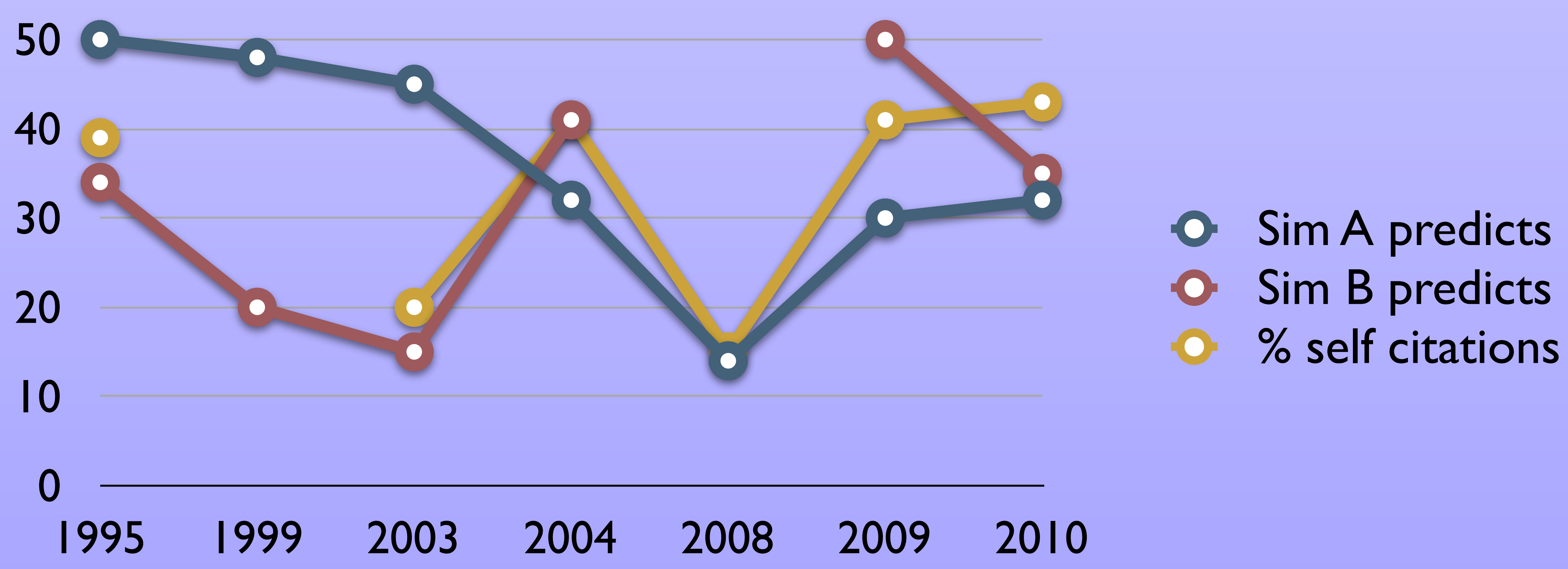
presented at the
Surrey Simulation Summer School
22 July 2011

Preface

- This is NOT a talk about yet another extension of the grossly over-rated & over-used General Linear Model
  - Prediction: History will show that the GLM did more to retard the development of the social sciences than any other research tool
- It is instead a short presentation of a few alternative ways to analyse data, invented for examining data from individuals studied over time.
- I have been working on it since 1975, and now have a book-full of techniques derived from its simple ideas.
- Maybe some of the techniques would be useful for examining the fit between outputs of agent-based models and very messy event-series data. Maybe not.
- I can show only the basics here, and point to some extensions. For more information, write me!

The Problem

- It is worthwhile to try validating simulation models by comparing their outputs to observations
- But how should we assess whether outputs of simulations A or B shown below
  - each give a good fit to observations?
  - show one simulation to be a better fit than the other?
- Wanted: useful indicators of prediction-observation fit
- Where to begin? Basic Principles



What my Stats Profs forgot to mention

- What is a statistic?
  - Best definition I know: *A number that represents some aspect of two or more other numbers*
  - If the numbers are measurements, then the resulting statistic is called "descriptive"
  - If the numbers are descriptive statistics, then both inferential and evidential statistics can be derived
- What make a statistic good?
  - mathematical properties (efficiency, sufficiency and all that)
  - conceptual properties (does it tell me something? Will the statistical tail wag the conceptual dog?)
- <u>Inferential statistics</u> are used to generalize from samples from populations
- <u>Evidential statistics</u> are used to generalize from predictions to observations
  - Because ABMers create simulations to make predictions, it seems reasonable to use evidential statistics.

What makes a useful evidential statistic?

- Many simulations predict complex, contingent, curvilinear relations
- Data for comparisons are often sparse, incomplete, and not statistically independent
- Measurements are often semi-orders, ordinal, ordered metric but not interval or ratio
- What NOT to do:
  - Average everything in hopes of eliminating "noise"
  - Conduct an omnivariate time series analysis or other useless variants of the General Linear Modal
  - Average over journals and test for least squares fitness
  - Try everything on SPSS menus
- What to do instead:
  - Conduct an interocular trauma test (as did Sengupta)
  - Stretch your fingers to prepare for Ordinal Pattern Analysis!

Denys Parsons' *Directory of Tunes & Musical Themes (1975)*

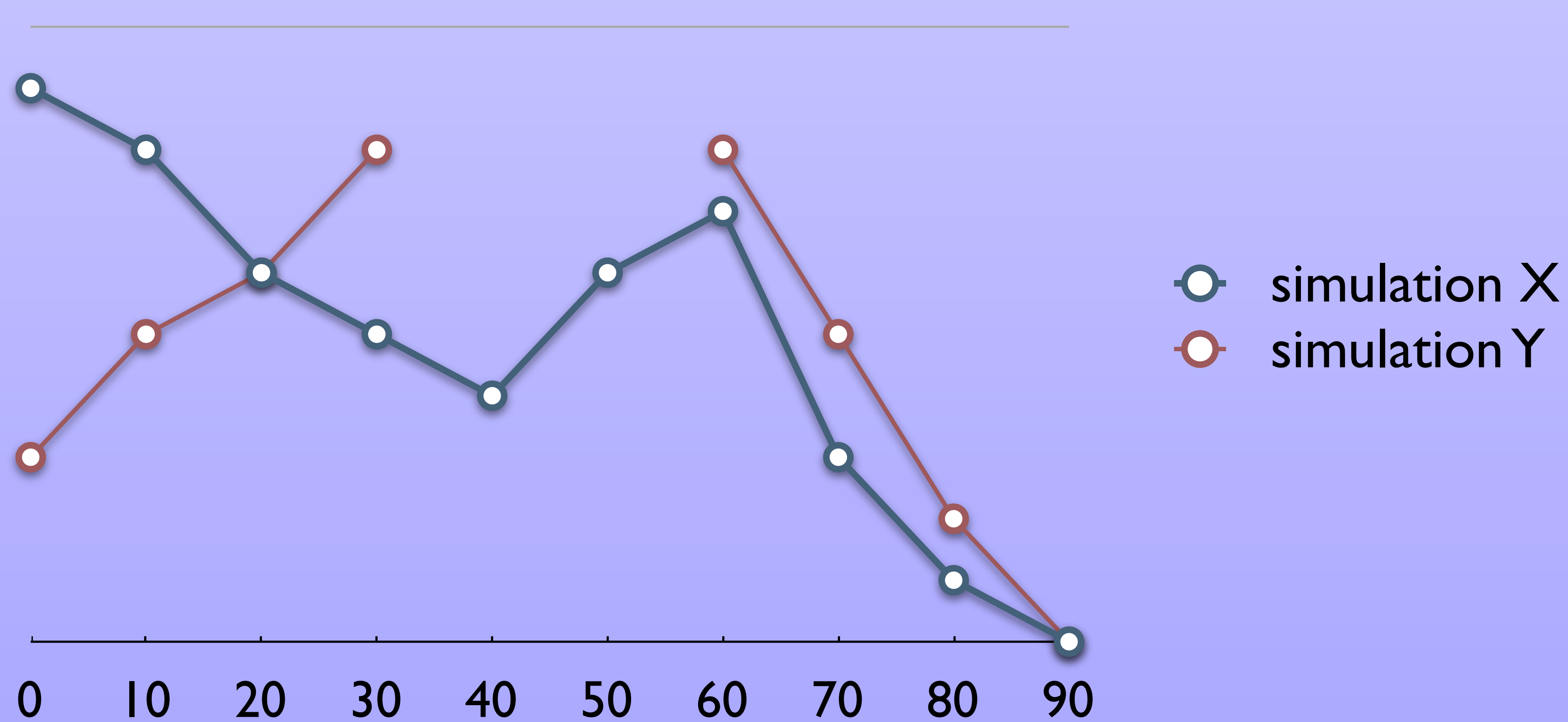
- 14,000 tunes coded by the ordinal relationship of their first 16 adjacent notes
  - H = higher than previous note, R = repeats previous, D = lower
  - Examples of entries in alphabetical order:
    - DDUUUUDDDUURRUDD = Star Spangled Banner
    - RHDHHHRHDDDHDDHH = God Save the Queen
    - URDUUUUUDHHRHHHH = O'Canada
    - URDRDHDRURUDRDUDH = Etude Numero 6 de Fernando Sor
  - To use: hum the first 17 notes, write down the UDR string, and look it up
  - Let's try it!
- Eureka! We want to the ordinal fit of our data (humming) with 14,000 theories
- So let's generalize:
  - examine ordinal pattern of non-adjacent pairs
  - examine ordinal pattern of differences in pairs
  - create useful indices of degree of fit. Examples:
    - p(hit) =#hits / (#hits - #misses)
    - Index of observed fit (IOF)
    - Index of predictive fit (IPF)

Ordinal Pattern Analysis: The basics

- Task: to count how many of the ordinal properties of predictions match the ordinal properties of observations, and how many mismatch.
  - The greater the proportion of matches, the better the fit
- But please heed the creed of the idiographic researcher: it is always better to analyse before we aggregate than to aggregate before we analyse.
  - Do not average over individual cases then analyse the averages
  - Instead, analyse individual cases, then aggregate over analyses
- Examples to follow...

Simplest example? Happiness over the life span

- Simulation X shows that agents start high, decline until about 50, then increase until about 70, then decrease until death
- Simulation Y shows that agents start low, rise until about 40, then decrease after about 70 until death



Ordinal Predictions of Simulation X

- Agents start high, decline until about 50, then increase until about 70, then decline until death. Predicted ordinal pairs:
  - 0s>10s, 0s>20s, 0s>30s, 0s>40s
  - 10s>20s, 10s>30s, 10s>40s
  - 20s>30s, 20s>40s
  - 30s>40s
  - 60s>50s, 60s>40s
  - 50s>40s
  - 70s>80s, 70s>90s, 70s>100s
  - 80s>90s, 80s>100s
  - 90s>100s
- Scope of X (proportion of pairs that simulation addresses) = 19/(11*10/2) = 19/55 = 0.34

Ordinal Predictions of Simulation Y

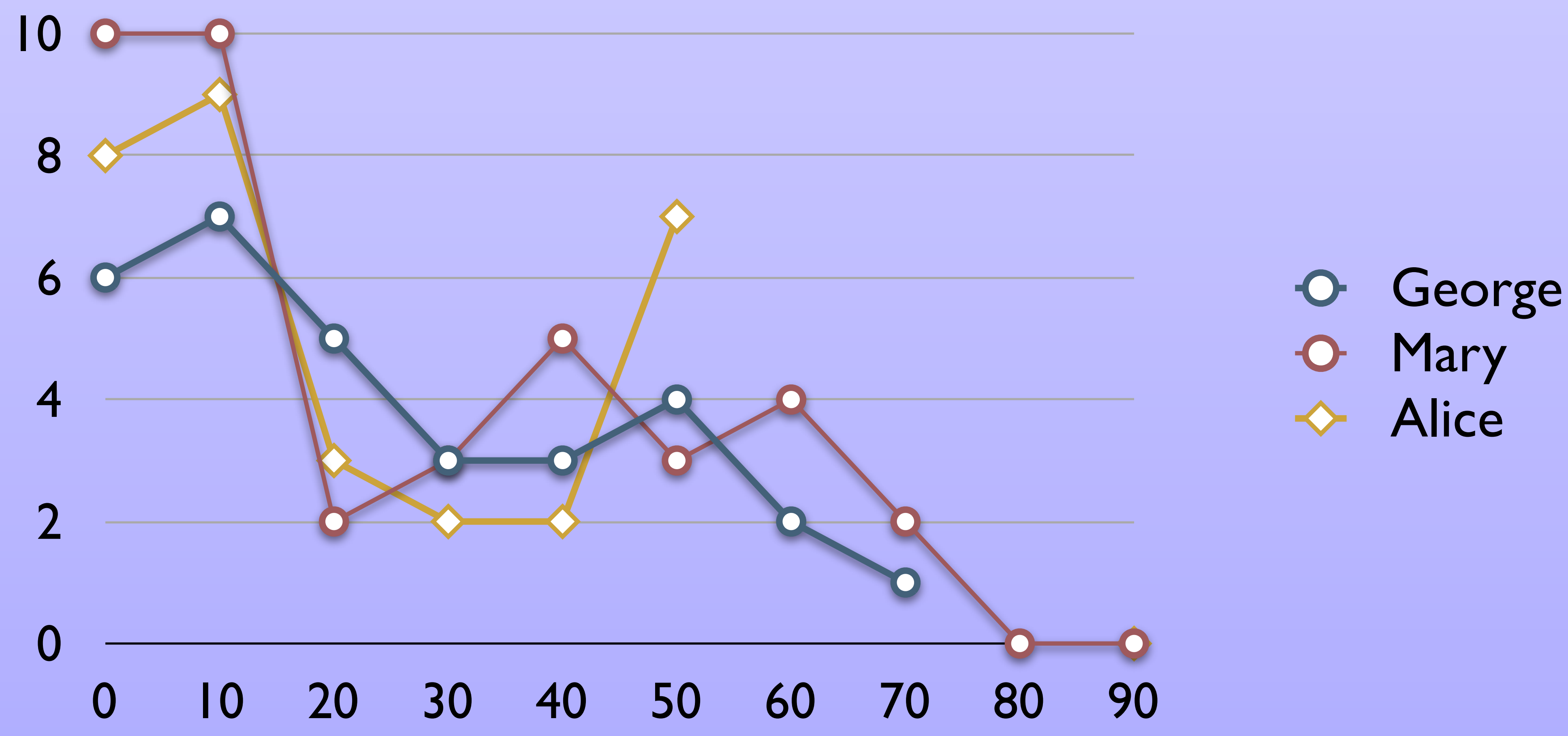
- Agents start low, rise until about 40, then decrease after about 70 until death. Predicted ordinal pairs:
  - 30s>20s, 30s>10s, 30s>0s
  - 20s>10s, 20s>0s
  - 10s>0s
  - 60s>70s, 60s>80s, 60s>90s, 60s>100s
  - 70s>80s, 70s>90s, 70s>100s
  - 80s>90s, 80s>100s
  - 90s>100s
- Scope of Y (proportion of pairs that simulation addresses) = 16/(11*10/2) = 16/55 = 0.29

Common and Opposing predictions

- Common
  - 70s>80s, 70s>90s, 70s>100s
  - 80s>90s, 80s>100s
  - 90s>100s
- Opposing
  - Simulation X
    - 0s>10s, 0s>20s, 0s>30s
    - 10s>20s, 10s>30s
    - 20s>30s
  - Simulation Y
    - 30s>20s, 30s>10s, 30s>0s
    - 20s>10s, 20s>0s
    - 10s>0s

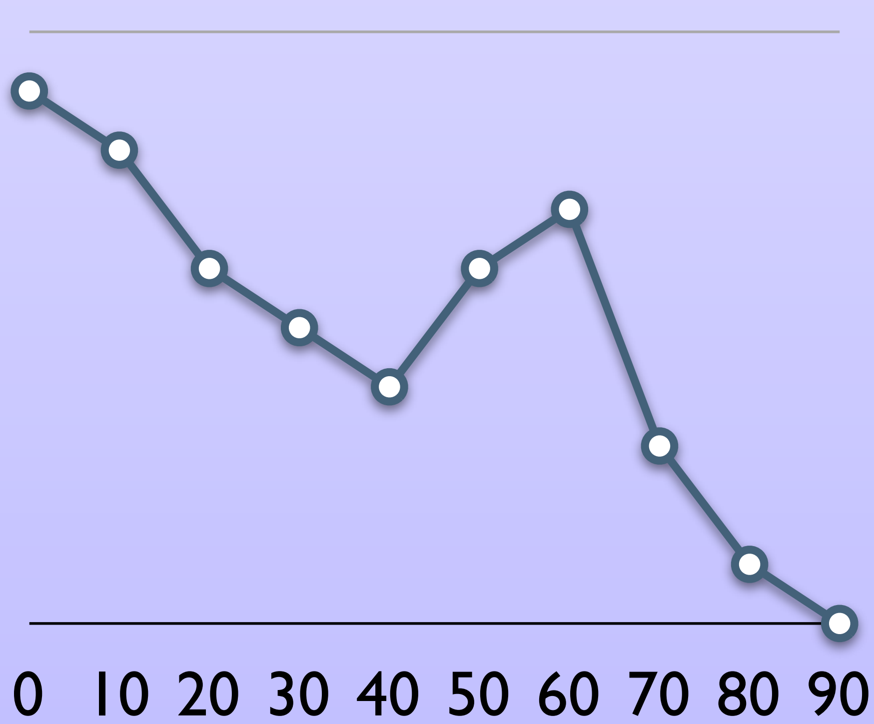
Now some data

- Ask old people to plot their happiness over their lifespan and most can do it. Part of a critical incident technique. Results below.

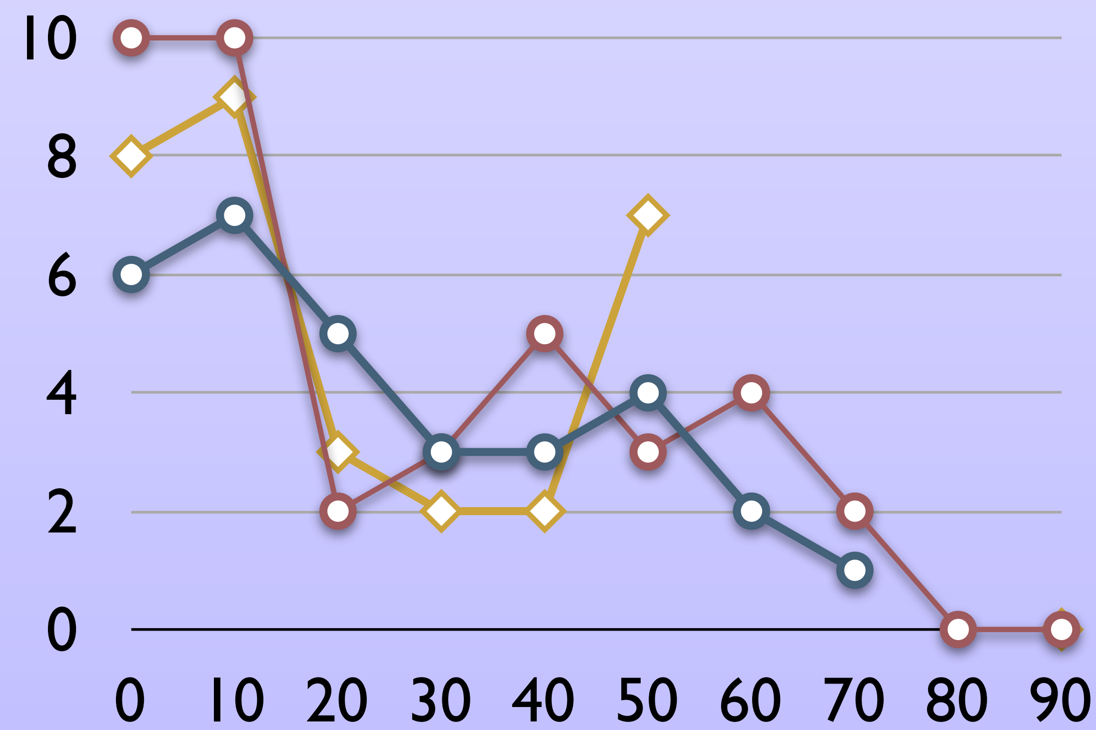


Testing Simulation X

Predictions



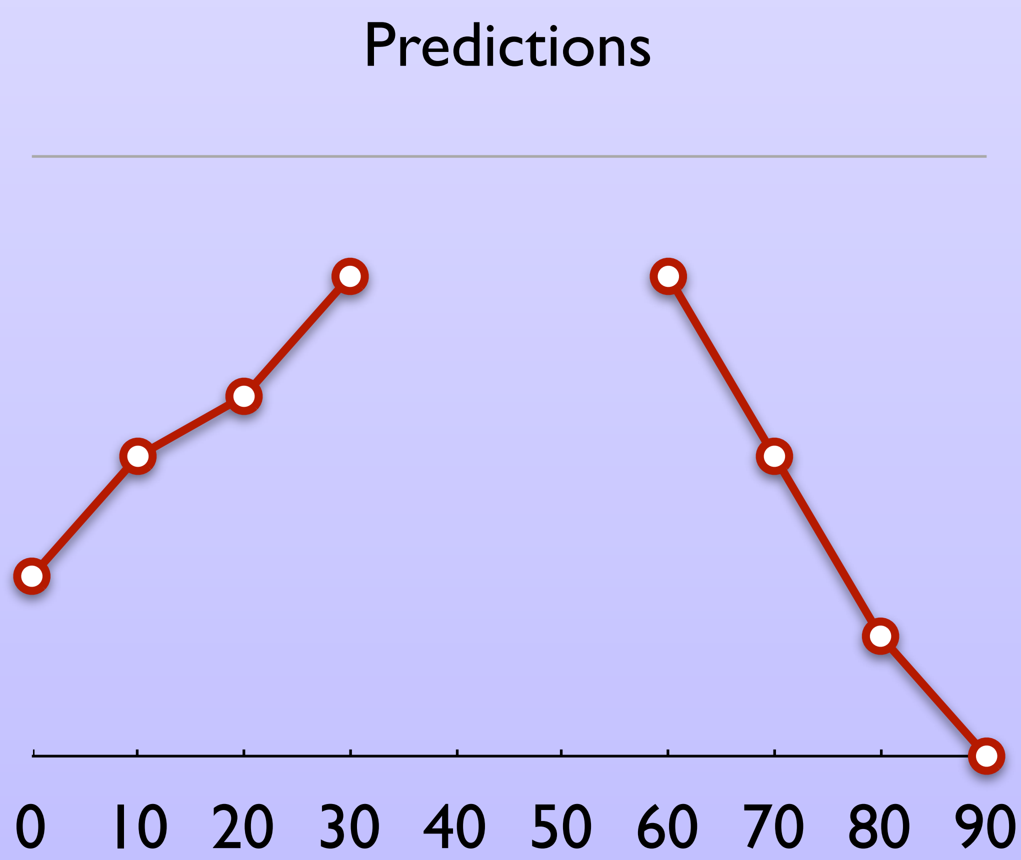
Observations

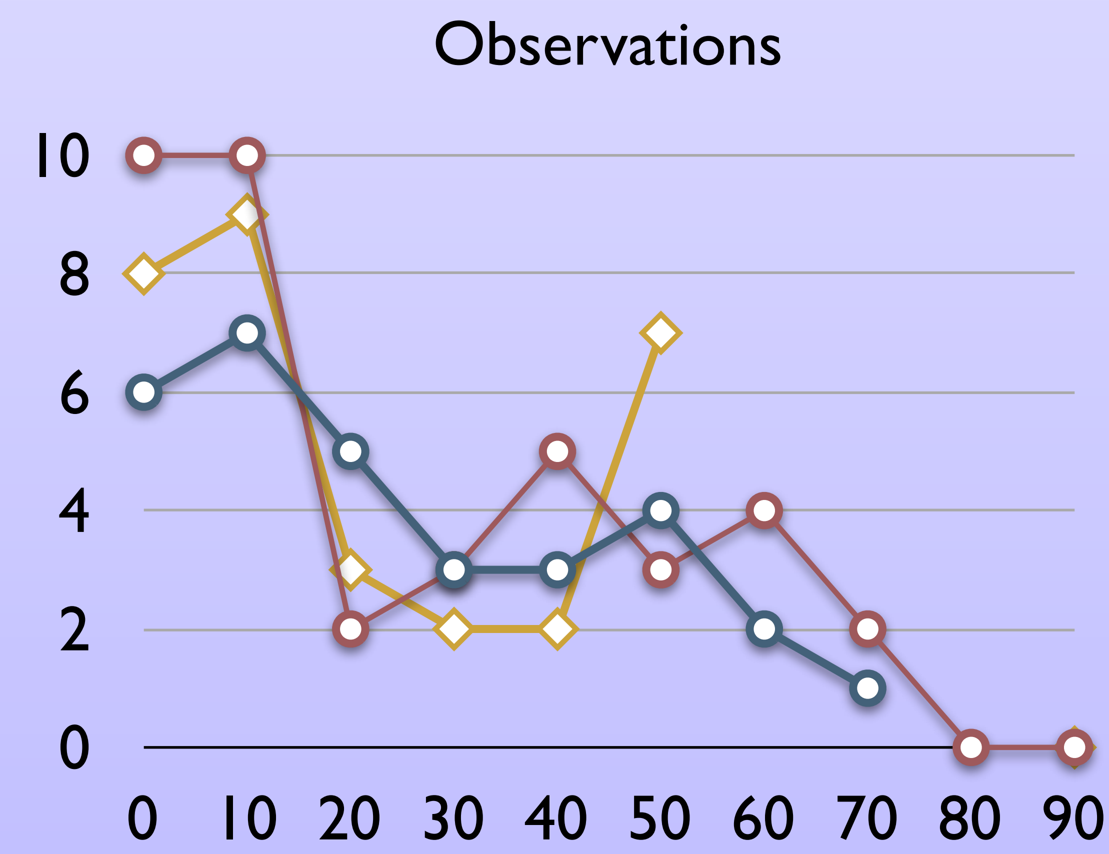


- 0s>10s, 0s>20s, 0s>30s, 0s>40s, 10s>20s, 10s>30s, 10s>40s, 20s>30s, 20s>40s, 30s>40s
- 60s>50s, 60s>40s, 50s>40s
- 70s>80s, 70s>90s, 70s>100s, 80s>90s, 80s>100s, 90s>100s

- Let's start using our fingers! Then calculate Index of Observed Fit (IOF) = (hits-misses)/(hits+misses)
  - Hits for **George** = 9; Misses = 4; IOF = (9-4)/(9+3) = 5/13 = +0.38
  - Hits for **Mary** = 9; Misses = 6; IOF = (9-6)/(9+6) = 3/15 = +0.20
  - Hits for **Alice** = 8; Misses = 3; IOF = (8-3)/(8+3) = 5/11 = +0.45
  - Hits for all three = 26; misses =13; IOF = (26-13)/(26+13) = +0.33

Testing Simulation Y





- 30s>20s, 30s>10s, 30s>0s, 20s>10s, 20s>0s, 10s>0s
- 60s>70s, 60s>80s, 60s>90s, 60s>100s, 70s>80s, 70s>90s, 70s>100s, 80s>90s, 80s>100s, 90s>100s

- Let's continue using our fingers! Then calculate Index of Observed Fit (IOF) = (hits-misses)/(hits+misses)
  - Hits for **George** = 2; Misses = 5; IOF = (2-5)/(2+5) = -2/7 = -0.28
  - Hits for **Mary** = 6; Misses = 6; IOF = (6-6)/(6+6) = 0/12 = +0.00
  - Hits for **Alice** = 1; Misses = 5; IOF = (1-5)/(1+5) = -4/6 = -0.67
  - Hits for all three = 9; misses =16; IOF = (9-16)/(9+16) = -7/25 = -0.28

Opposing and common predictions of X and Y

- Opposing:
    - Simulation X
        - 0s>10s, 0s>20s, 0s>30s, 10s>20s, 10s>30s, 20s>30s
    - Simulation Y
        - 30s>20s, 30s>10s, 30s>0s, 20s>10s, 20s>0s, 10s>0s
    - Hits for X (Misses for Y) : George = 5; Mary = 4; Alice =5; IOF(X) = +0.56
    - Hits for Y (misses for X): George = 1; Mary = 2; Alice = 1; IOF(X) = -0.56

- Common
    - Simulations X and Y
    - 70s>80s, 70s>90s, 70s>100s, 80s>90s, 80s>100s, 90s>100s
    - Hits for X (Misses for Y) : George = ?; Mary = 2; Alice =?; IOF(X) = +1.00
    - Hits for Y (misses for X): George = ?; Mary = 2; Alice = ?; IOF(X) = +1.00

In Summary...

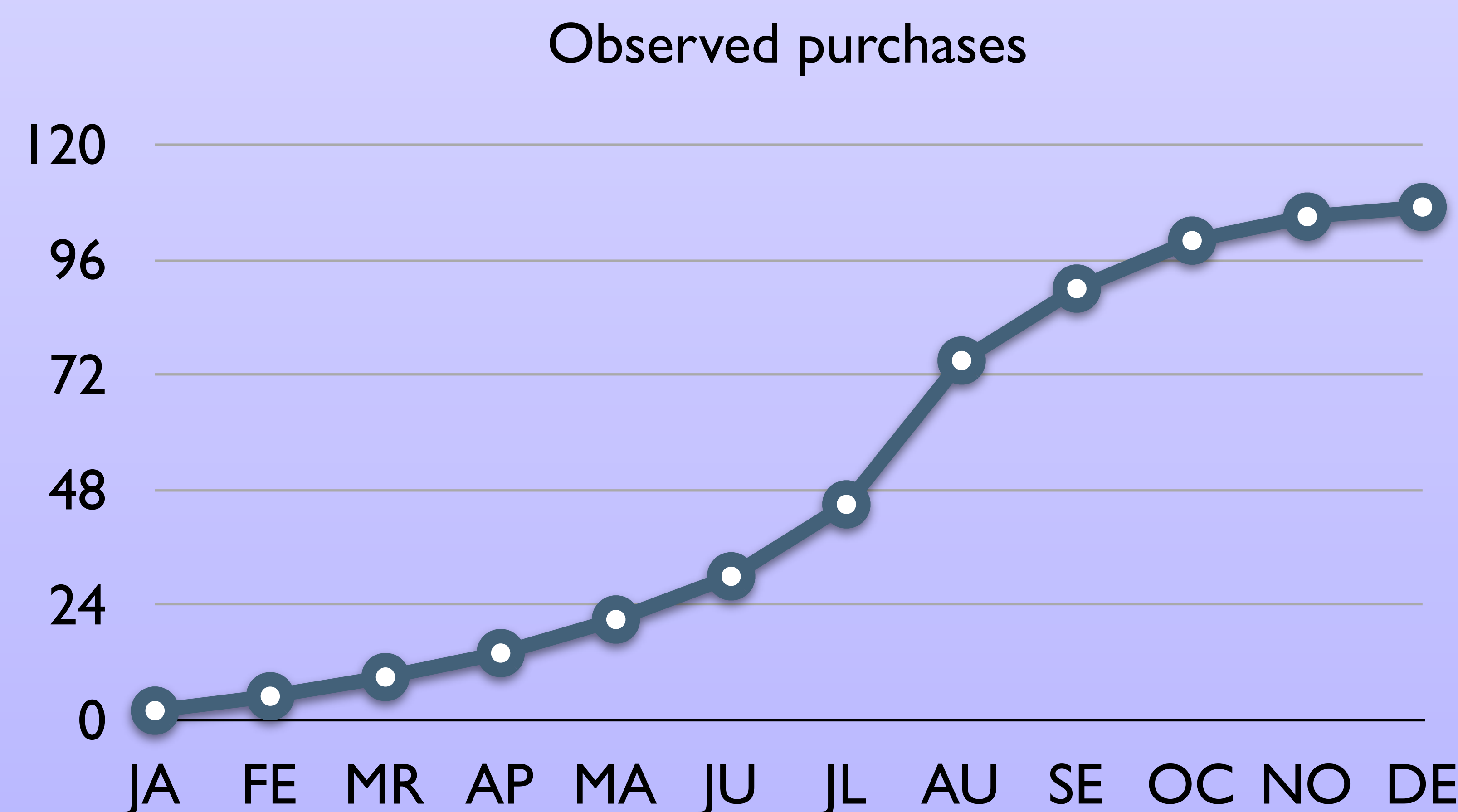
	Sim X	Sim Y
Scope	0.34	0.29
IOF George	0.38	-0.28
IOF Mary	0.20	0.00
IOF Alice	0.45	-0.67
IOF opposing	0.56	-0.56
IOF common	1.00	1.00

Ordinal patterns of differences

- When our simulations predict differences in differences, and when our data show properties of an interval scale, we can also do OPA on the differences
- Examples:
    - purchases of a new product will accelerate to a turning point, then decelerate to an asymptote
    - acts of revenge will peak soon after a public injustice, then decay exponentially
    - attitudes among men will polarize faster than attitudes among women (sex difference in variance)

Product Sales Example:

"purchases of a new product will accelerate to a turning point, then decelerate to an asymptote"



Turning point = AU, so it must be excluded from predictions (degrees of freedom and all that)

but we can predict:

JL-JU > JU-MA, JL-JU > MA-AP, JL-JU > AP-MR, JL-JU > MR-FE, etc. etc.

The silly question of "statistical significance"

- Evidential statistics were not invented to answer questions about generalizing from sample to population.
- But most people brainwashed by the Neyman-Pearson approach to inference do not feel comfortable without a "p value" to report -- including most editors.
- What to do?
  - When comparing 2+ simulations, pick an IOF of interest, and ask how many individual cases are best predicted by simulation X, simulation Y, simulation Z etc.
  - When trying to see if the IOF of a single simulation is doing better than chance, do a resampling, randomization test (a la Fisher, Diaconis & Efron)

A small sample of other topics

- Clustering people according to their fit with various simulations
- Nominal Pattern Analysis
- Derivation of prototypes via modal orderings
  - Use of these prototypes in generating new simulations
- Rules for handing ties
- Combining results across studies/experiments

But OPA is not supported by SPSS...

- And I hope it never is!
- But if you want a computer programme to help you calculate, be patient until September and I will send you a new version of one
  - I am working on one now, using R
  - If you want a copy, just ask:
    - warren_thorngate@carleton.ca
    - warren.thorngate@rogers.com

Fini

Kheili mamnoon!

Questions? Comments? Coffee?