

## Research Question

How can we **predict performance** on an exam question based on:

- an estimate of the **question difficulty** and
- an estimate of the **student's proficiency**?

## Background of ASSISTment System

- Computer system to **assist** student in learning material and **assess** a student's knowledge
- Generates an 'ASSISTment' score per student per question
- ASSISTment score = # hints requested + # questions attempted
- Historical data of student performance on questions in ASSISTment system allows us to estimate the **question difficulty**

## Data

- R Data frames
- 226,000 records spanning over **900 students**
- Rectangular shape

## Data Issues

- Removed blank questions
- Removed invalid student IDs
- Removed problem items
- Removed both questions and students with less than 10 attempts

# Computer Based Tutoring Systems: ASSISTments System

## Statistical Models

- Rasch Model

$$\log\left(\frac{p(X_{ij}=1)}{1-p(X_{ij}=1)}\right) = \theta_i - \beta_j$$

- Logistic Regression Model

$$\log\left(\frac{p(X_{ij}=1)}{1-p(X_{ij}=1)}\right) = H_{ij} + T_{ij} - \beta_j$$

- Modified Rasch Model

$$\log(A_{ij}) = \theta_i - \beta_j$$

$i$  = student

$j$  = main question

$$X_{ij} = \begin{cases} 1 & \text{if student } i \text{ gets question } j \text{ right} \\ 0 & \text{else} \end{cases}$$

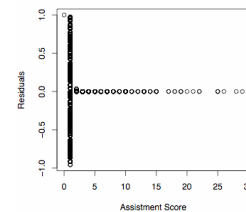
$A_{ij}$  = ASSISTment score for student  $i$  on question  $j$  (hints + attempts)

$H_{ij}$  = number of hints taken by student  $i$  on question  $j$

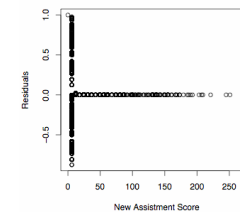
$T_{ij}$  = number of attempts made by student  $i$  on question  $j$

## Logistic Regression Residual Plots

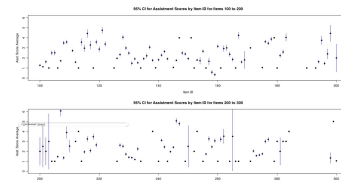
$$\theta_j = T_{ij} + H_{ij}$$



$$\theta_j = 6T_{ij} + 25H_{ij}$$



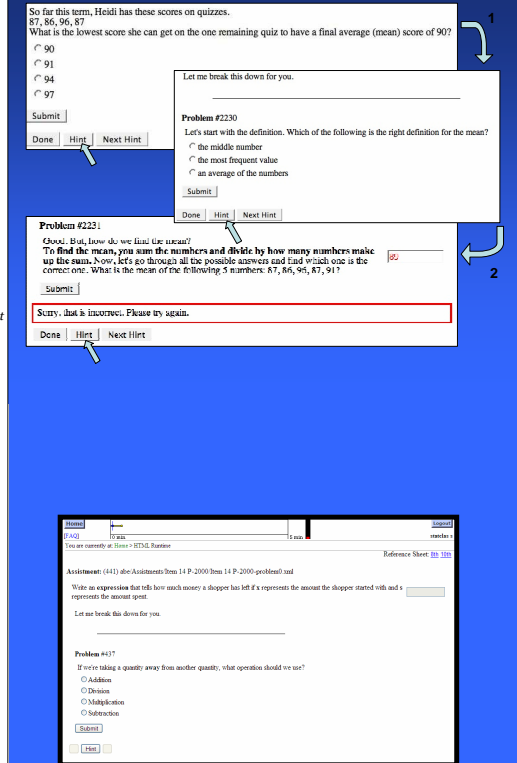
## Analysis



Plot of the 95% confidence intervals for the ASSISTment scores for questions 100-200 and 200-300.

## Logistic Regression Results

| Models                | Residual Deviance | Null Deviance | "R <sup>2</sup> " |
|-----------------------|-------------------|---------------|-------------------|
| $A_1 = T_i + H_i$     | 15604             | 20462         | 0.2374            |
| $A'_1 = 6T_i + 25H_i$ | 12805             | 20462         | 0.3742            |
| $T_i$                 | 17451             | 20462         | 0.1472            |
| Null; only betas      | 20462             | 20462         | 0                 |



## Acknowledgements

Ken Koedinger Ph.D

- Professor in the Human Computer Interaction Institute
- Head of Pittsburgh Science of Learning Center
- Develops computerized academic tutoring and evaluation systems

ASSISTments team including **Beth Ayers** from Carnegie Mellon University and **Mingyu Feng** from Worcester Polytechnic Institute

Professors **Brian Junker** and **Jim Delaney**



ASSISTments

WPI

Carnegie Mellon