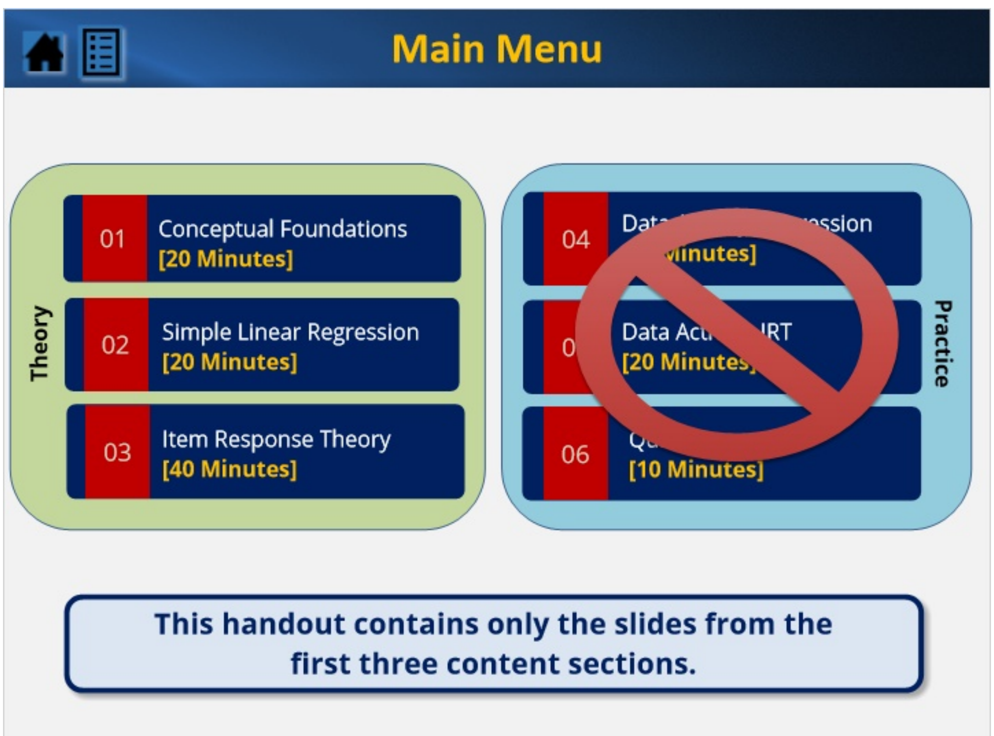
ITEMS Digital Module 06: Posterior Predictive Model Checking

This document contains all core content slides from sections 1-3 with the exception of slides that show video screens. In the digital module all slides can be accessed individually.

Module Organization

The module starts with an introductory section that leads to the main menu from which learners can select individual content and activity sections:



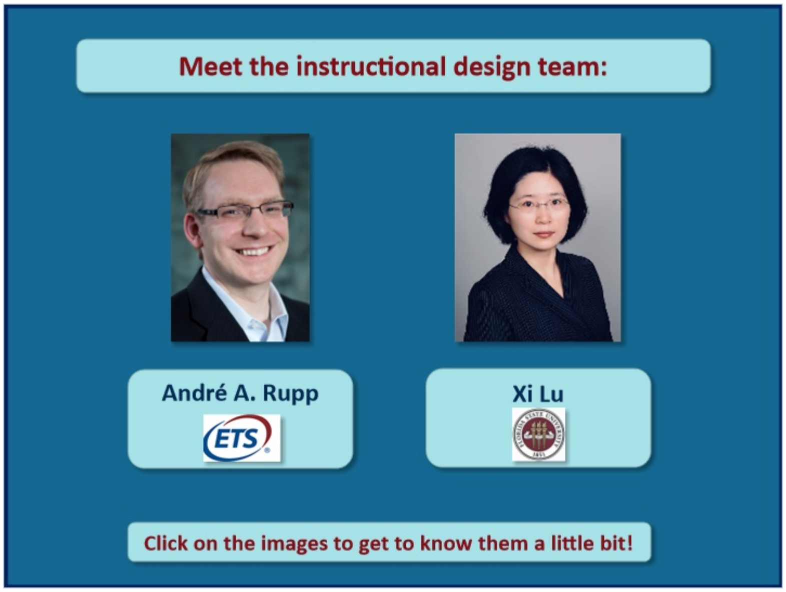
DM06 HANDOUT Version

1. Module Overview

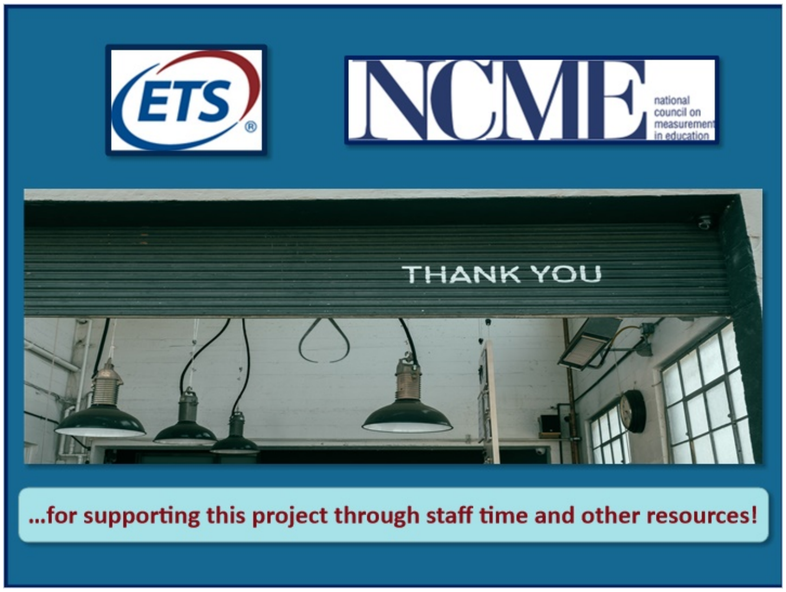
1.1 Module Cover

1.2 Content Team

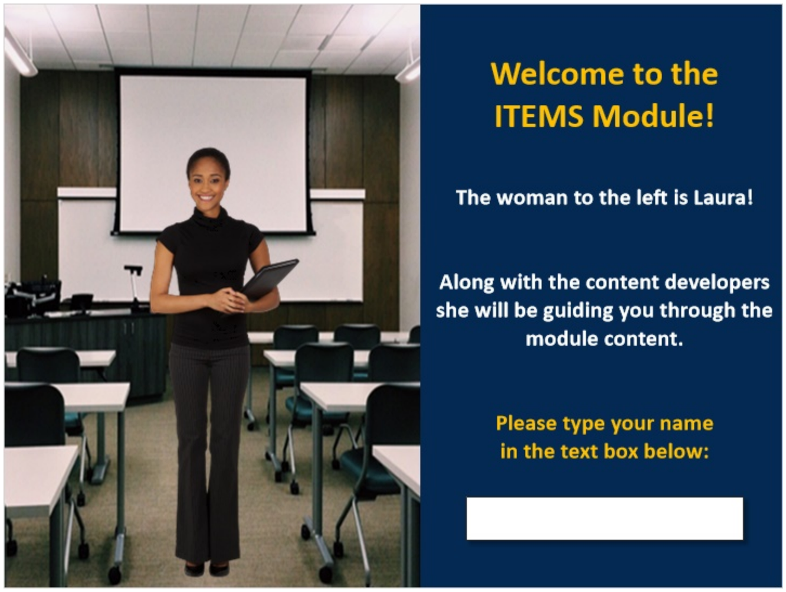
1.3 Design Team



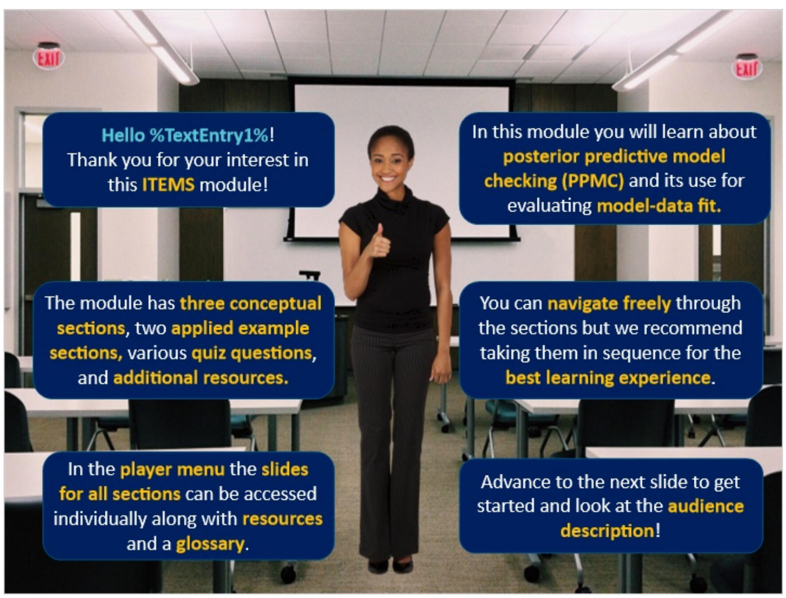
1.4 Support Team



1.5 Welcome



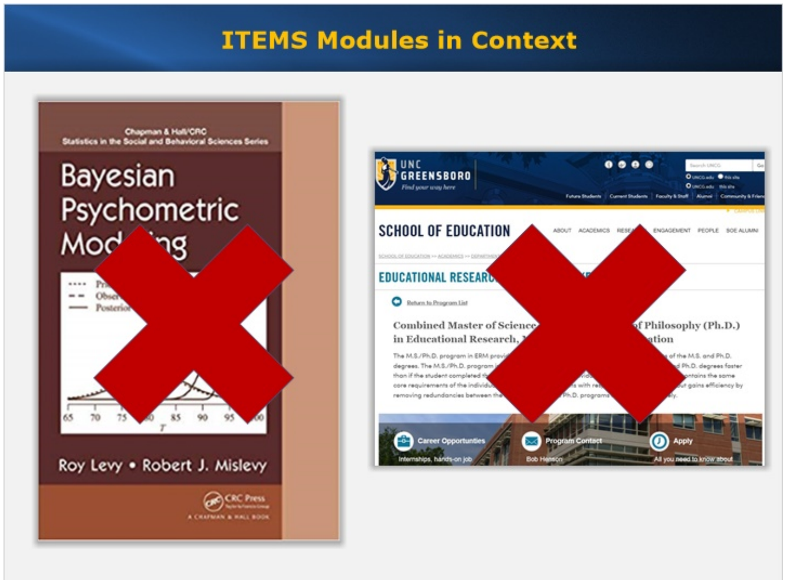
1.6 Overview



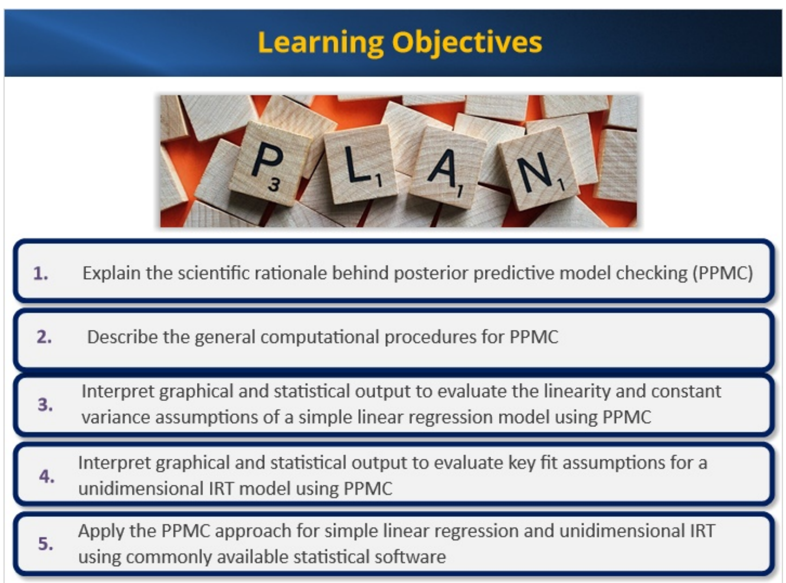
1.7 Target Audience

1.8 Expecations (I)

1.9 Expectations (II)



1.10 Learning Objectives



1.11 Prerequisites

Prerequisites

- **Working knowledge of foundational measurement concepts**:
  - ✓ Construct definitions / latent variables
  - ✓ Assessment formats
  - ✓ Item / task types
  - ✓ Scales and scale scores
  - ✓ Basic aspects of assessment development

- **Working knowledge of foundational statistical concepts**:
  - ✓ Descriptive statistics for distributions
  - ✓ Simple linear regression model
  - ✓ Statistical inference with p-values

1.12 Resources

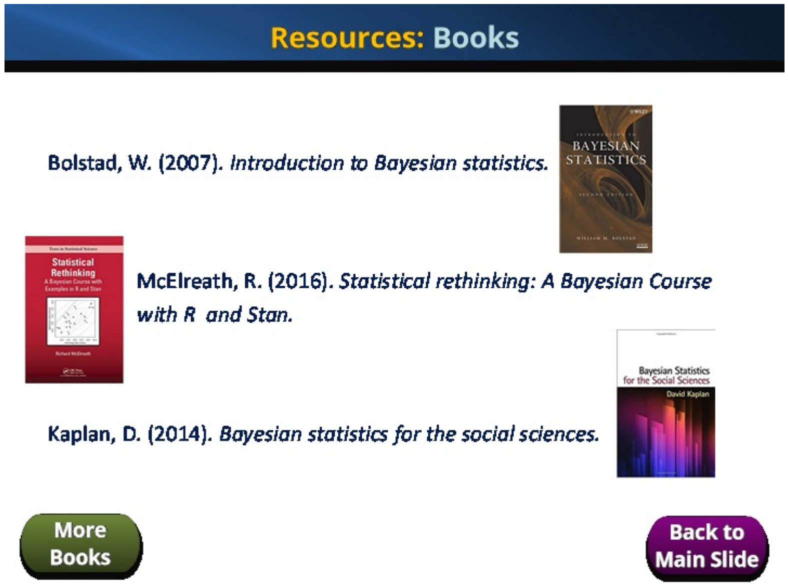
Resources

Selected free materials are in the 'Resource' section of the player interface.

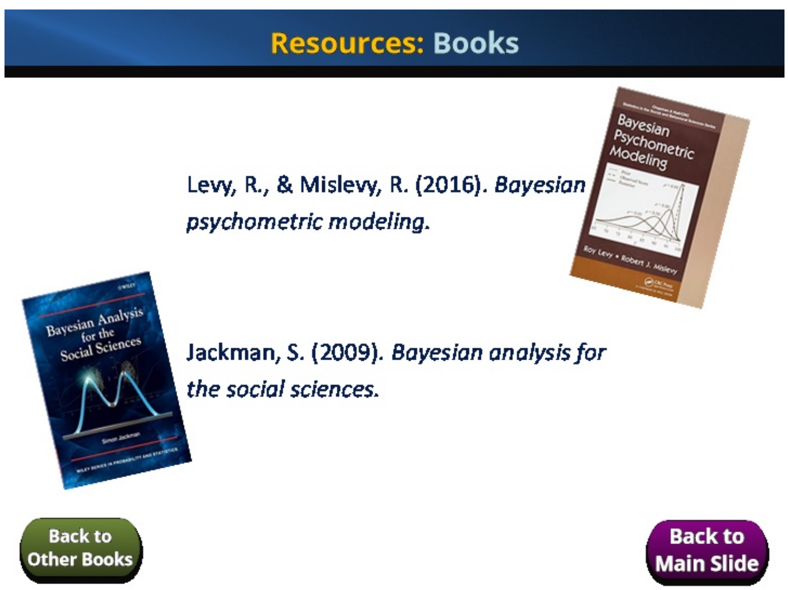
Links to purchase key reference books on Amazon are provided via the button below.

Reference Books

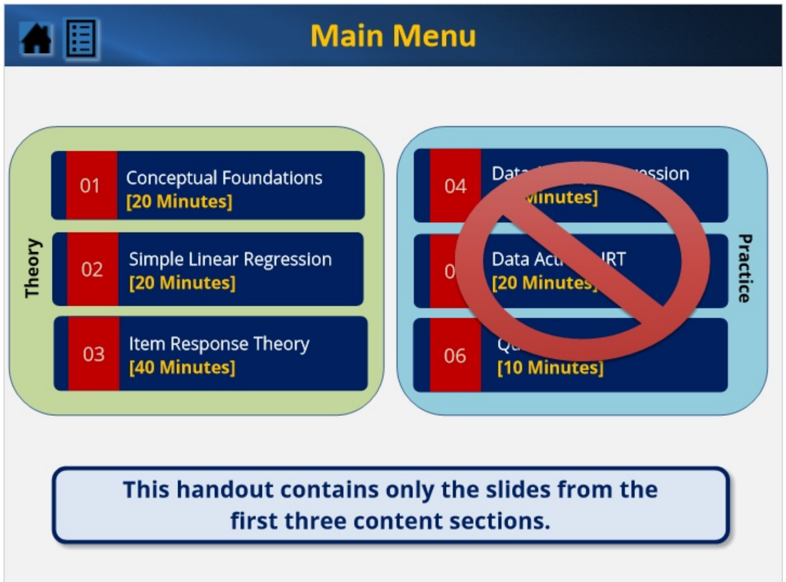
Resources 1 (Slide Layer)



Resources 2 (Slide Layer)



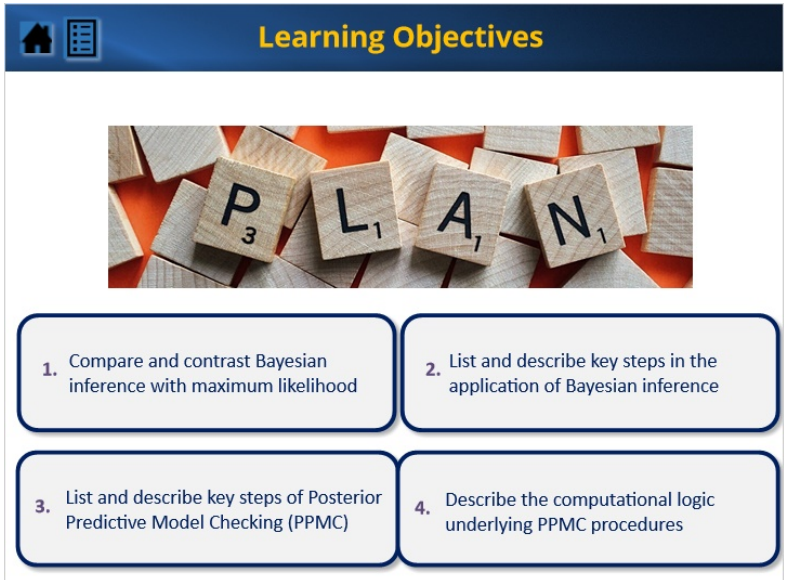
1.13 Main Menu



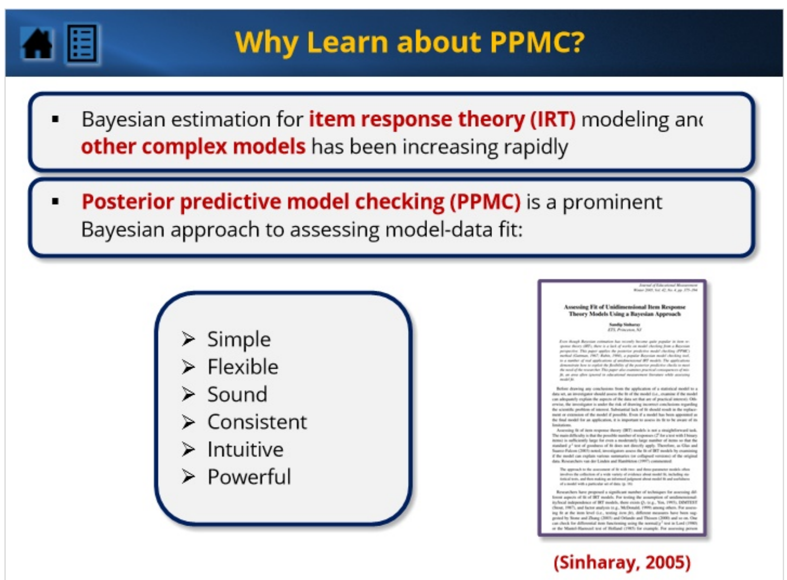
2. Conceptual Foundations

2.1 Cover: Foundations

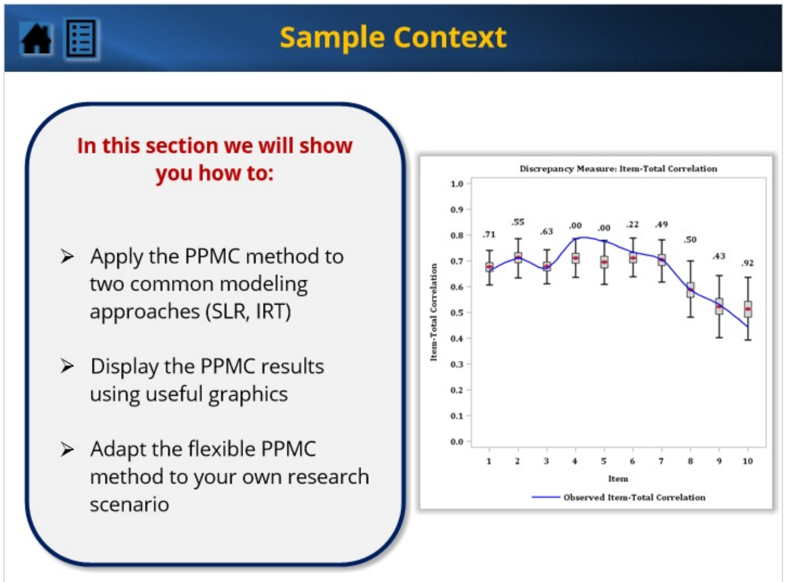
2.2 Objectives: Foundations



2.3 Motivation



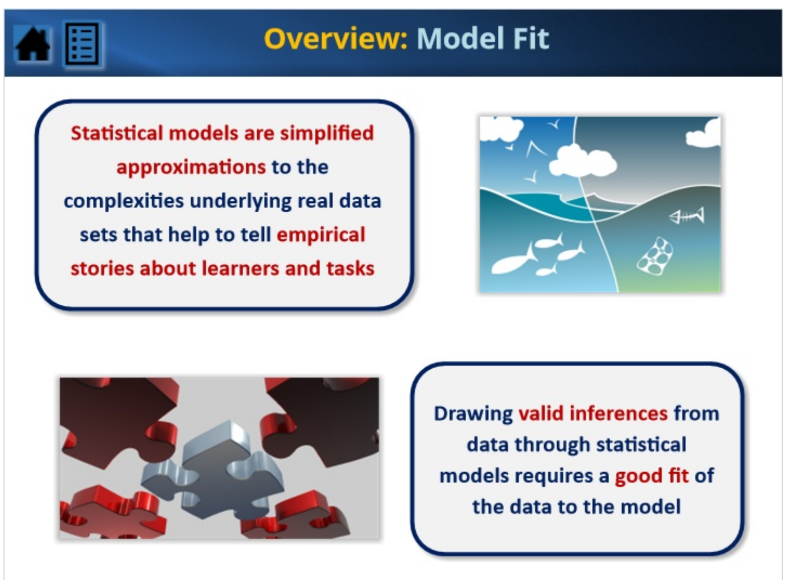
2.4 Sample Context



2.5 Topic Selection

2.6 Bookmark: Theoretical Framework

2.7 Overview: Model Fit



2.8 Overview: Model Fit

Overview: Model Fit

We may wish to ask questions such as:

- ➢ What aspects of the observed data are not captured by my model?
- ➢ Are the aspects not captured by the model critical for my inferences?
- ➢ Are there ways that I can refine my model in order to improve fit?
- ➢ Do I have the statistics available that help assess all critical aspects?
- ➢ How much do my prior beliefs influence my interpretations?

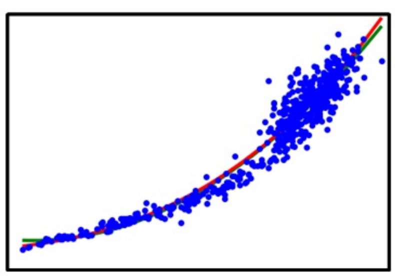
Answering these questions is the foundation of Posterior Predictive Model Checking (PPMC), the Bayesian approach to model-data fit.

2.9 Consequences of Misfit (I)

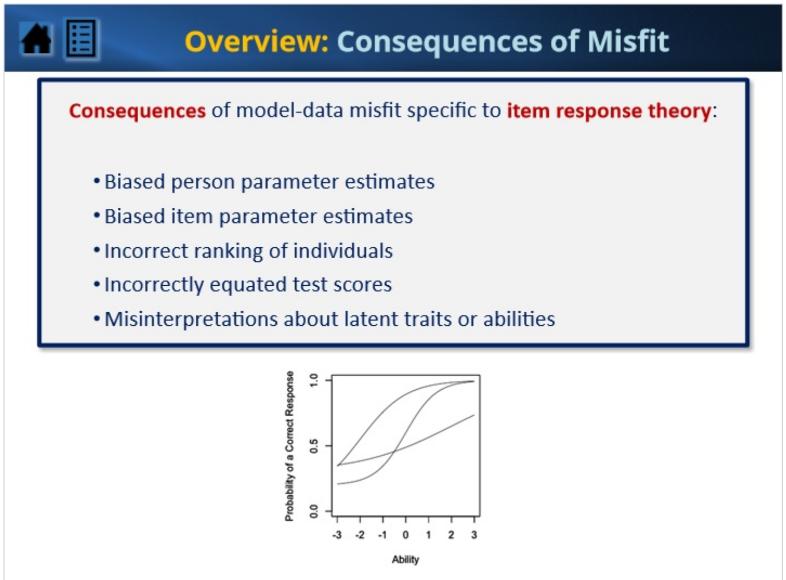
Overview: Consequences of Misfit

Consequences of model-data misfit specific to **linear regression**:

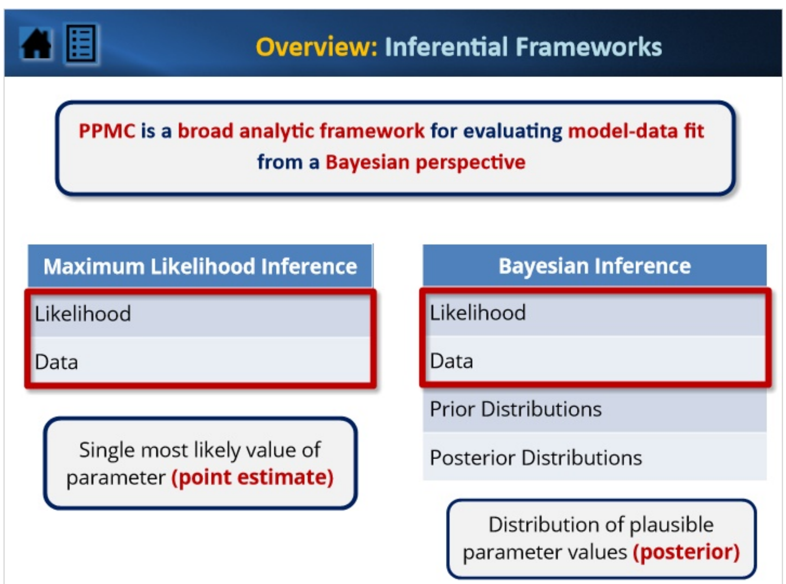
- Biased regression parameters
- Incorrect and imprecise predictions for the outcome variable
- Misinterpretations about substantive relationships



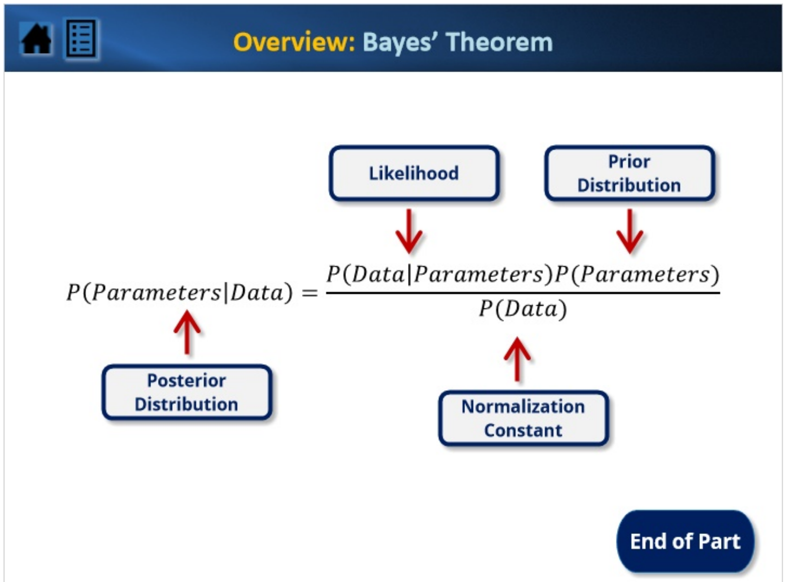
2.10 Consequences of Misfit (II)



2.11 Overview: Inferential Frameworks

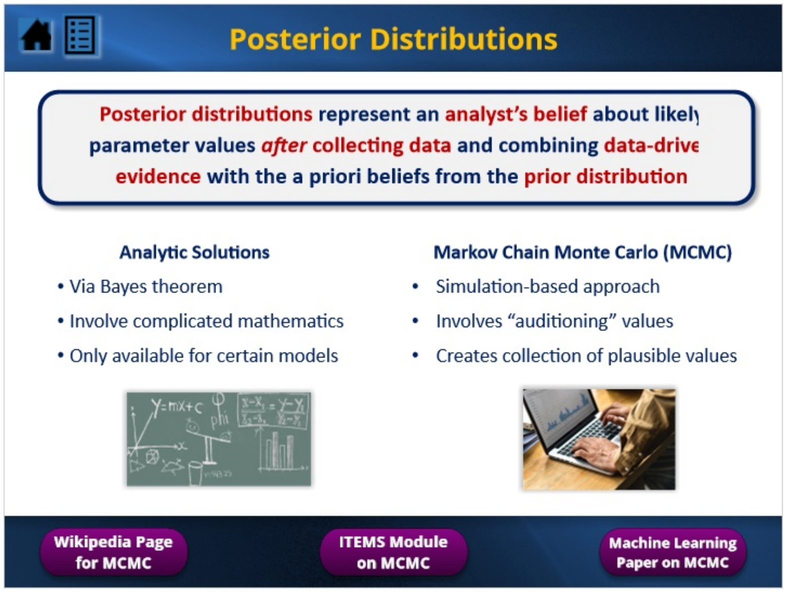


2.12 Overview: Bayes Theorem

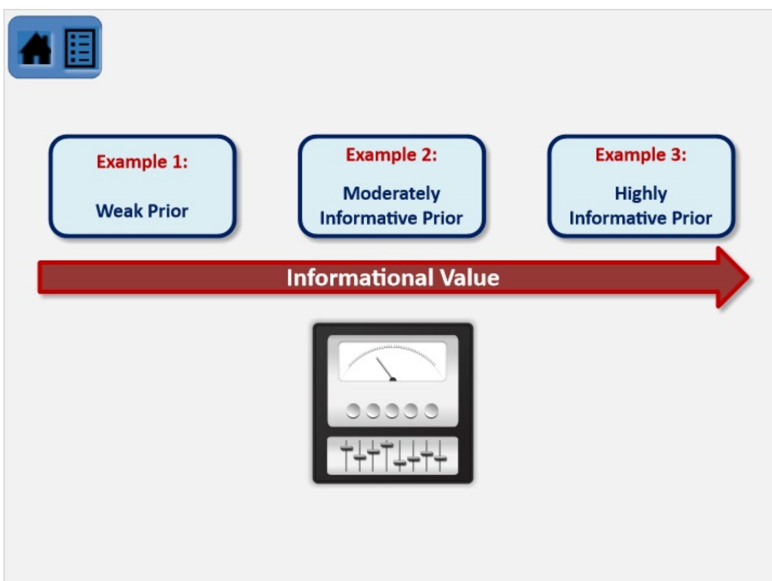


2.13 Bookmark: Posterior

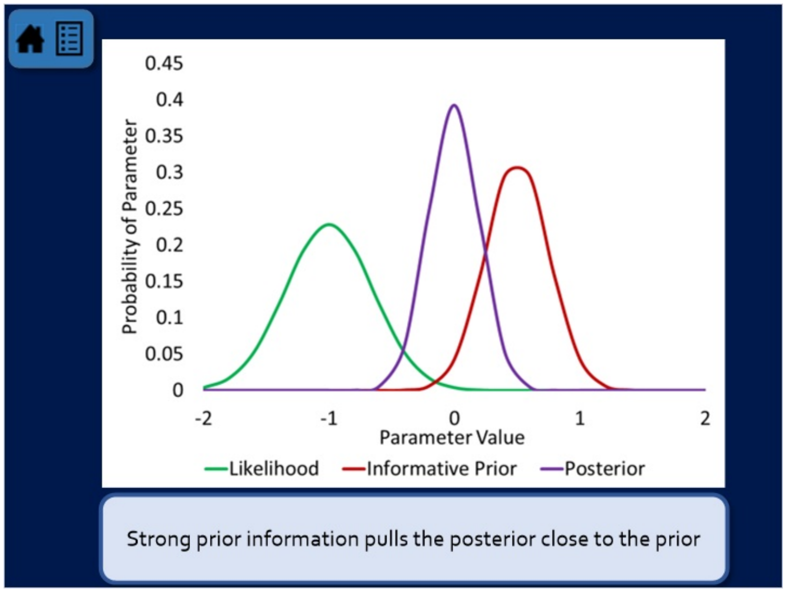
2.14 Posterior Distributions



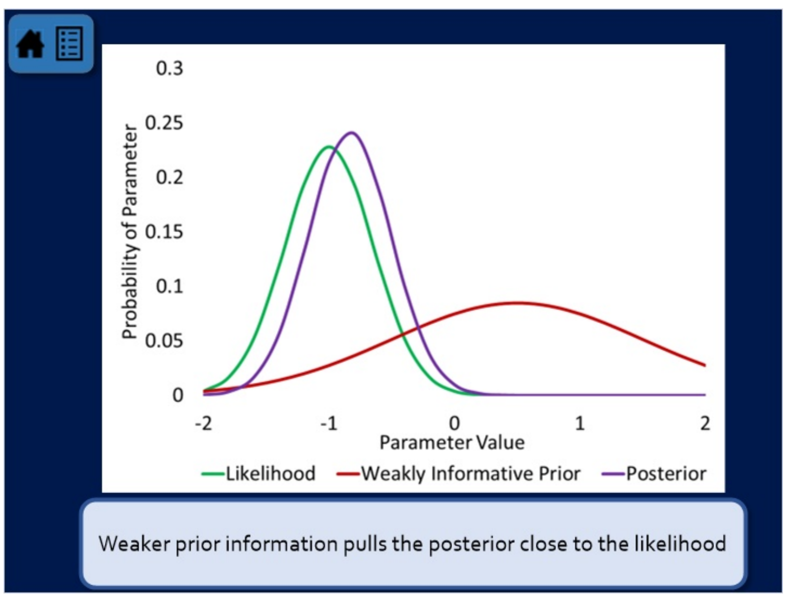
2.15 Example Selection



2.16 Posterior Example (I)

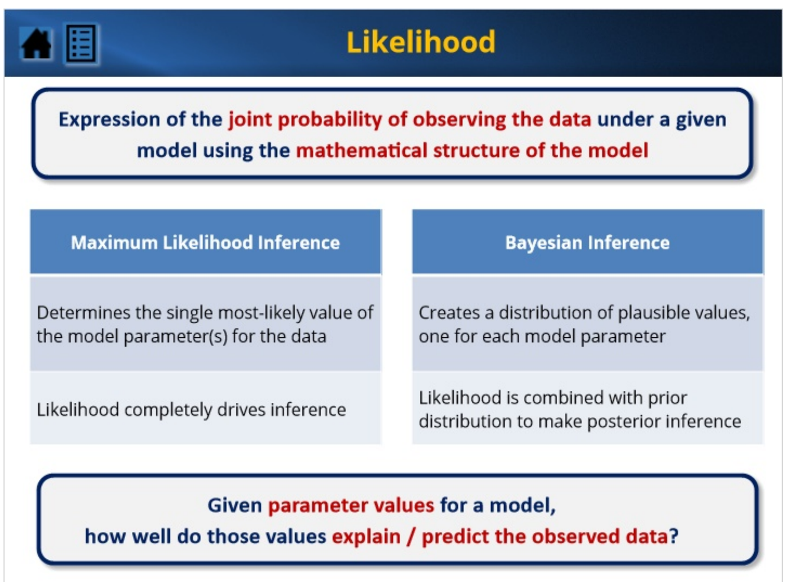


2.17 Posterior Example (II)

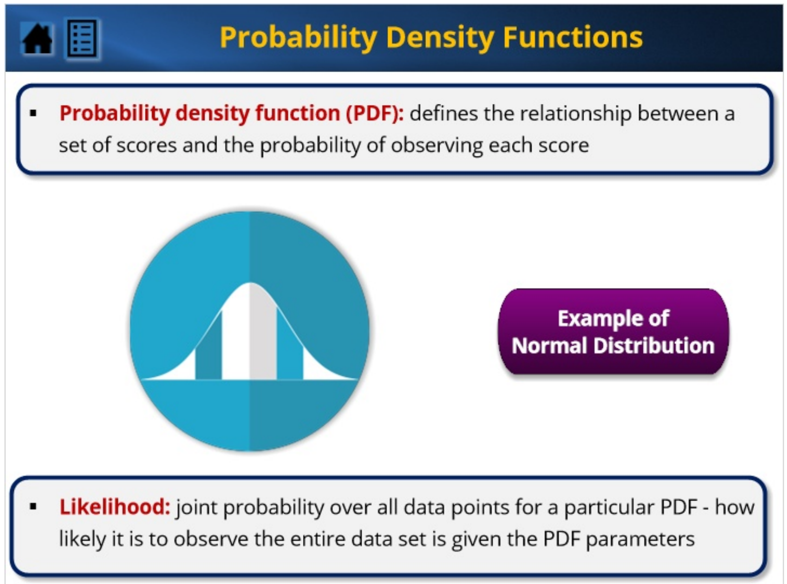


2.18 Bookmark: Likelihood

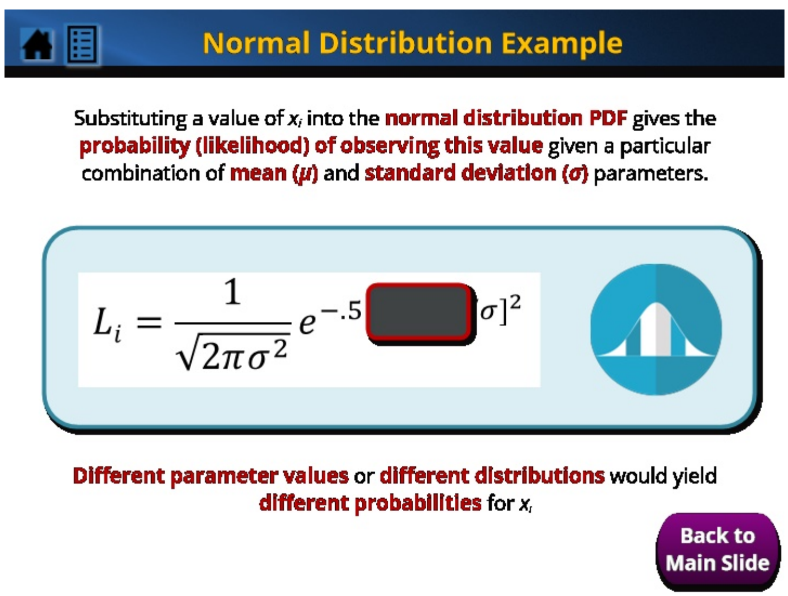
2.19 Likelihood (I)



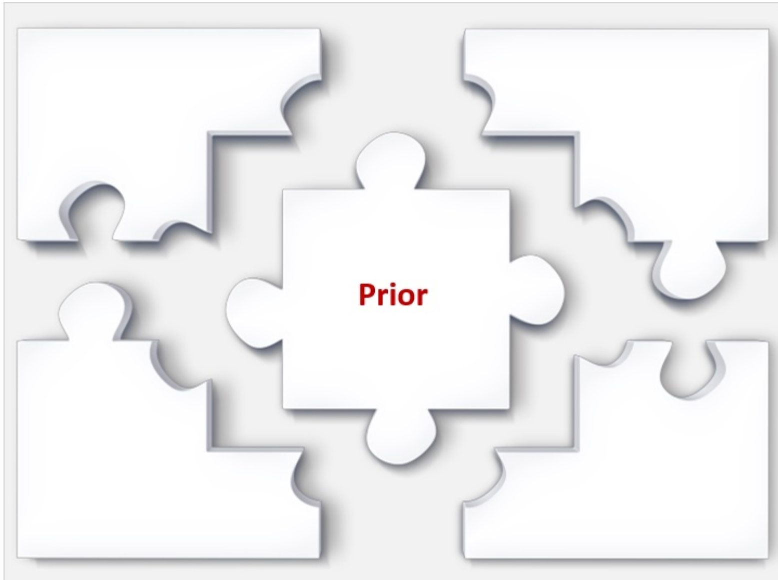
2.20 Likelihood (II)



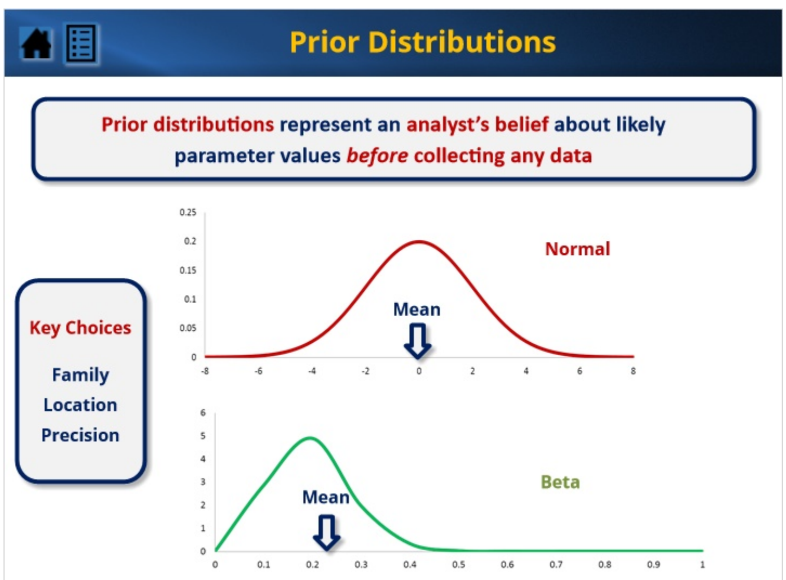
Example (Slide Layer)



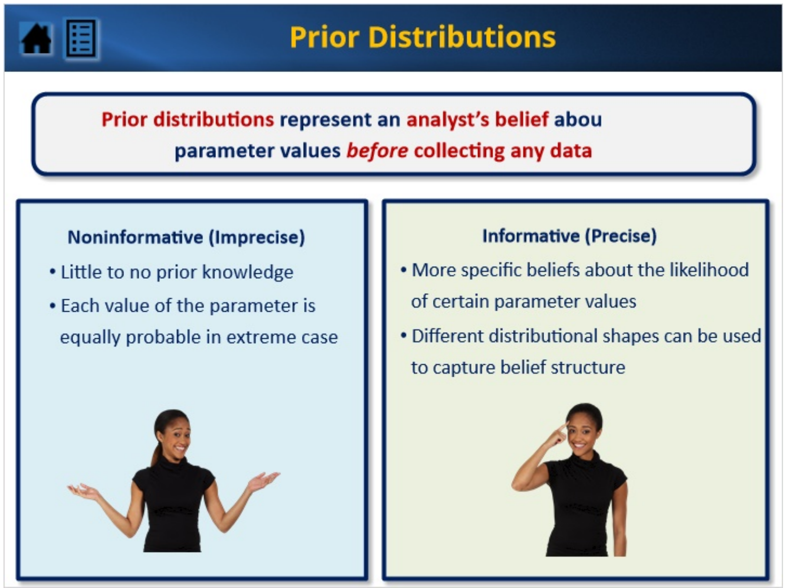
2.21 Bookmark: Prior



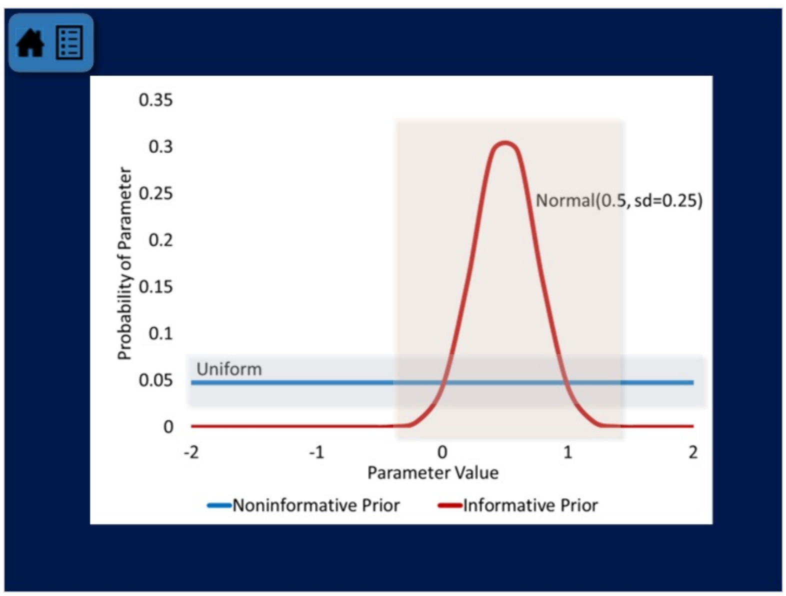
2.22 Prior Distributions (I)



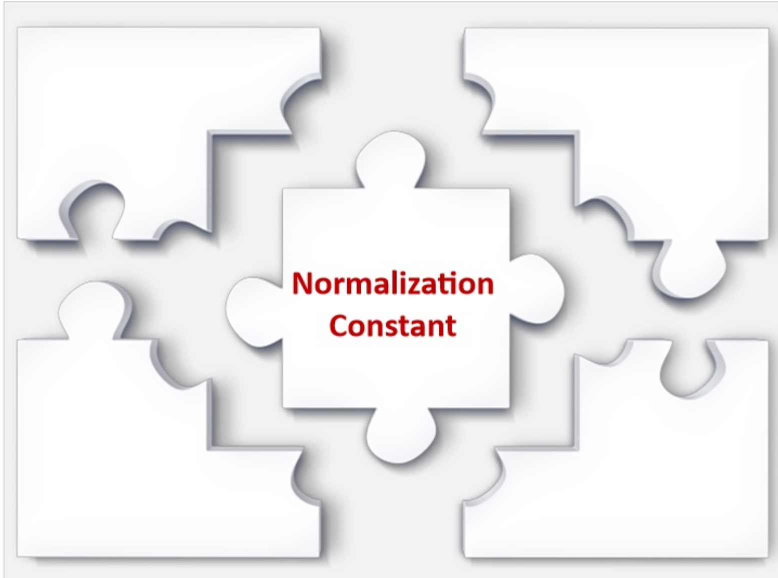
2.23 Prior Distributions (II)



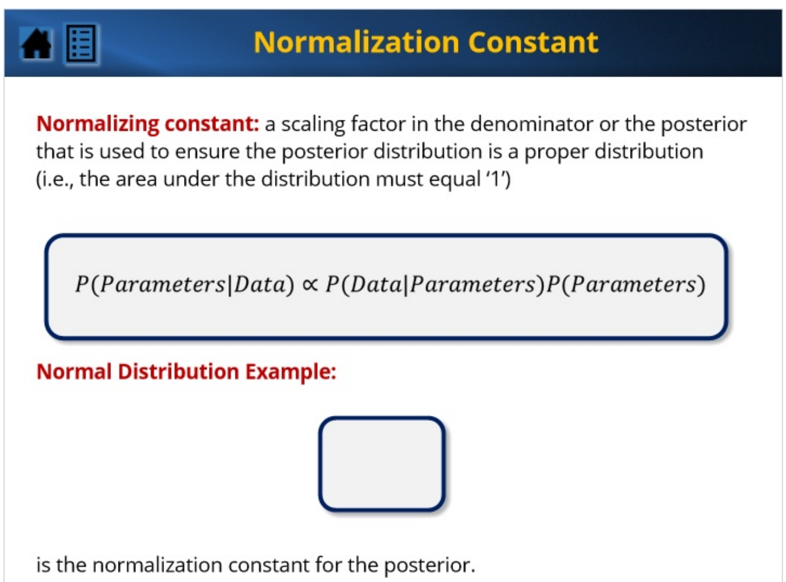
2.24 Prior Distributions (III)



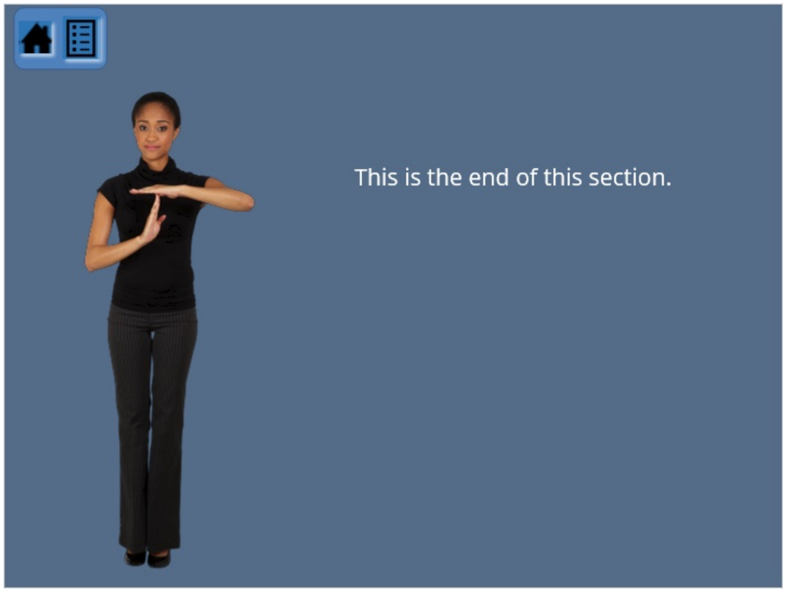
2.25 Bookmark: Normalization



2.26 Normalization Constant

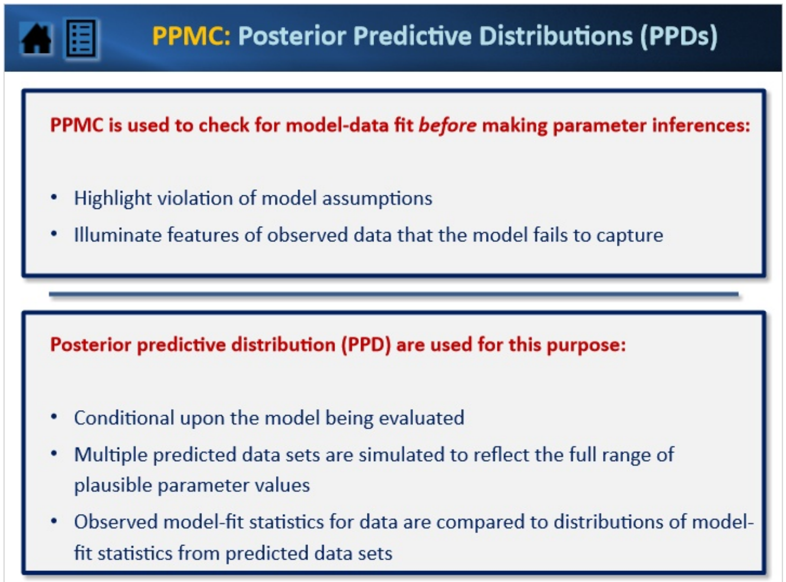


2.27 Bookend: Theoretical Framework

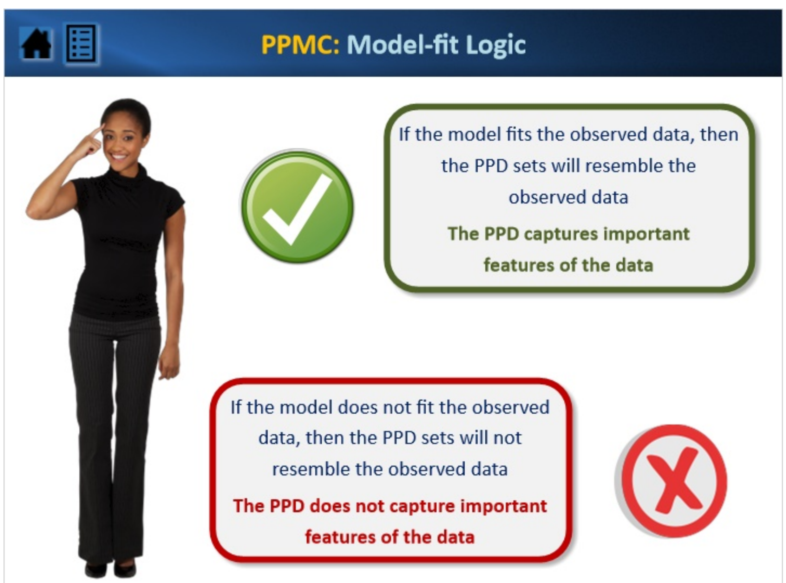


2.28 Bookmark: Computational Approach

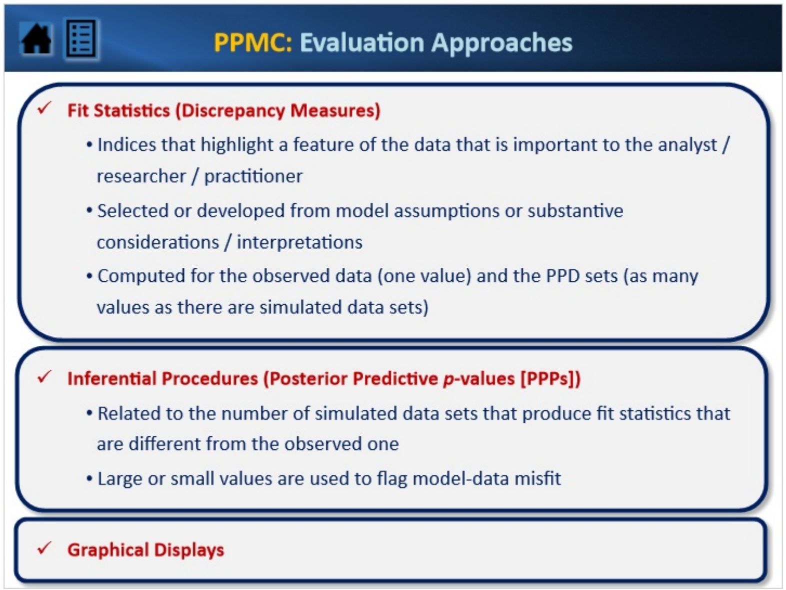
2.29 Posterior Predictive Distributions



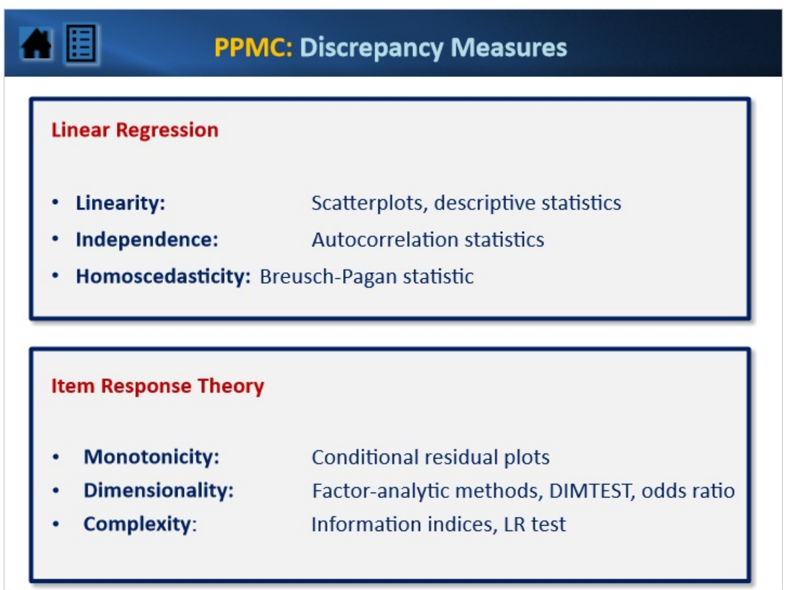
2.30 Model-fit Logic



2.31 Evaluation Procedures



2.32 Discrepancy Measures



2.33 Computational Steps

PPMC: Computational Steps

1. **Sample randomly a single value for each model parameter** from the associated posterior distribution
2. **Simulate one posterior predictive data** set using the random parameter draw(s) and the statistical model structure
3. **Compute the discrepancy measure** for the predicted and the observed data sets
4. **Compare the two values of the discrepancy measure** and record whether the predicted value is larger

5. **Repeat steps 1-4** a large number of times (e.g., 1,000 times or 10,000 times)

6. **Tabulate how often** the value of the predicted discrepancy measure was larger than the corresponding value for the observed data (posterior predictive *p*-value)
7. **Interpret the resulting percentage** to make a judgment about model fit to suggest model modification or replacement

2.34 Data Examples

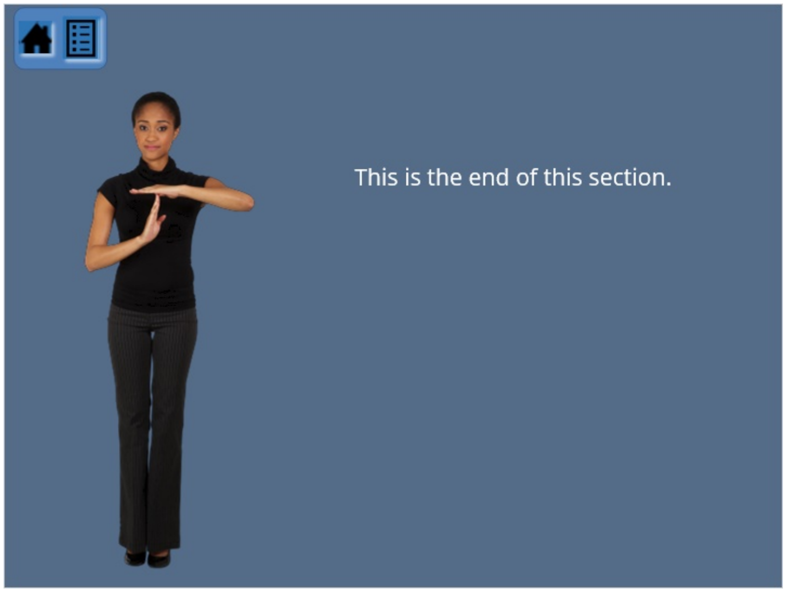
Data Examples

There are **two extended examples** in this digital module that you can access from the **main menu** or the **final slide in this section**:

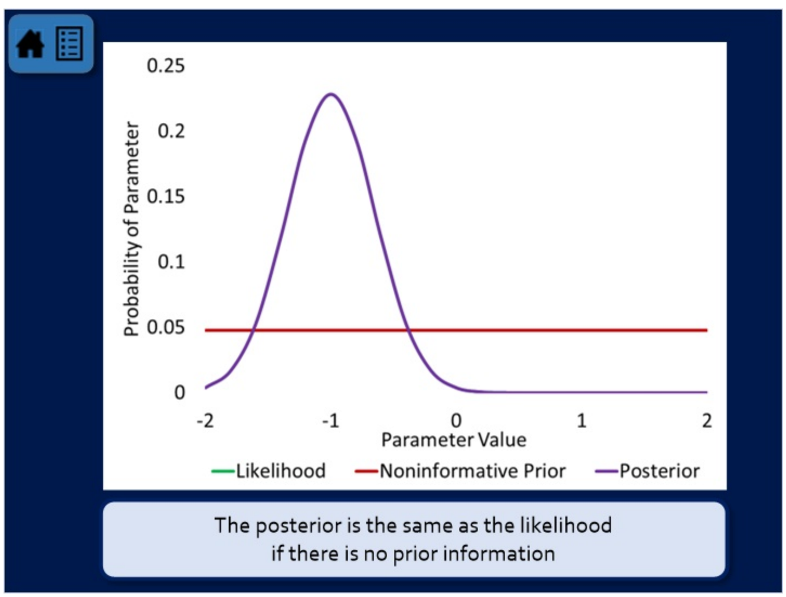
- **Simple linear regression**
- **Item response theory**

The general steps for **Bayesian inference and PPMC** are the same but the **statistical model** and associated **discrepancy measures** vary!

2.35 Bookend: Computational Approach



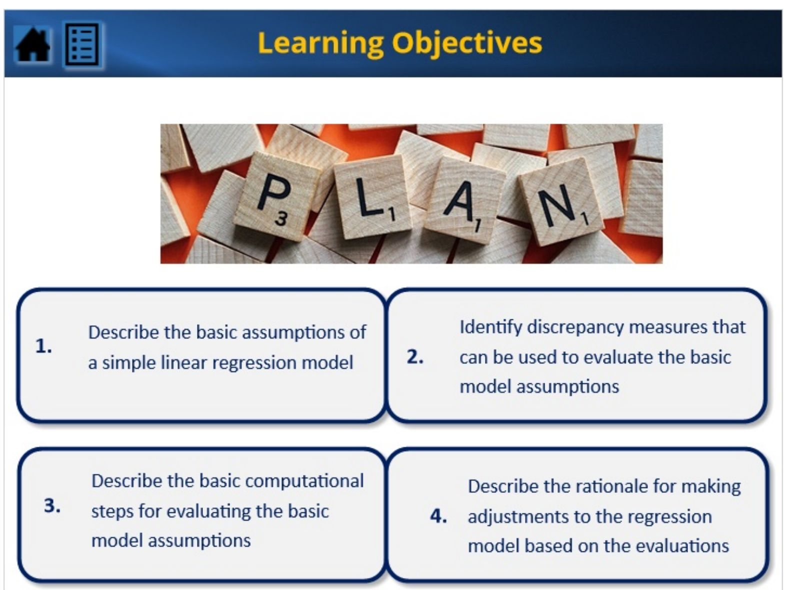
2.36 Posterior Example (III)



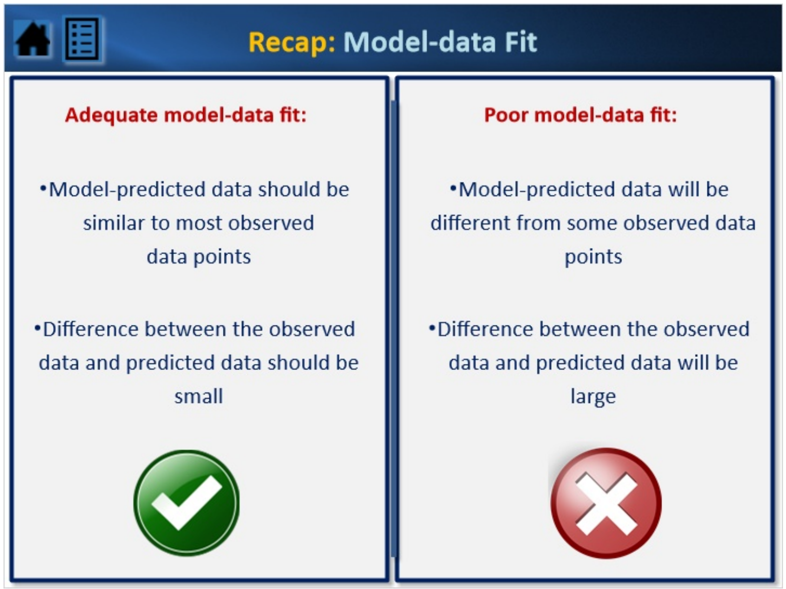
3. Simple Linear Regression

3.1 Cover: Regression

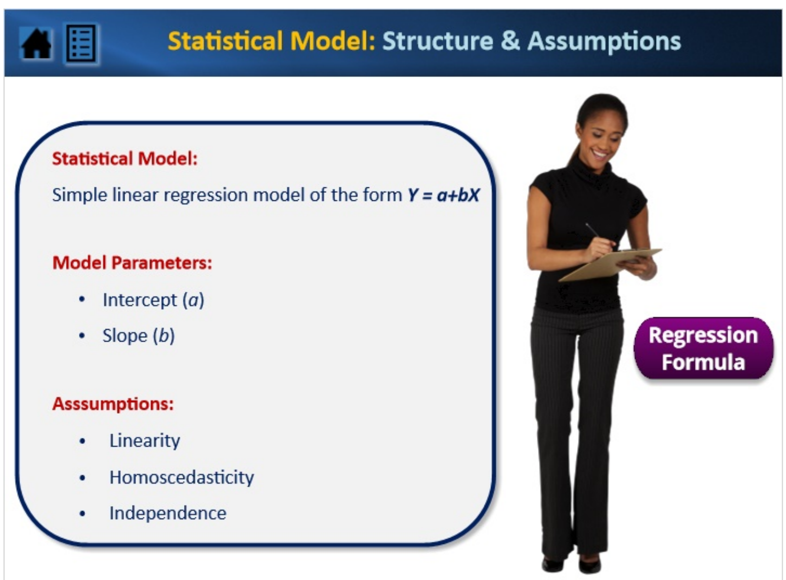
3.2 Learning Objectives: Regression



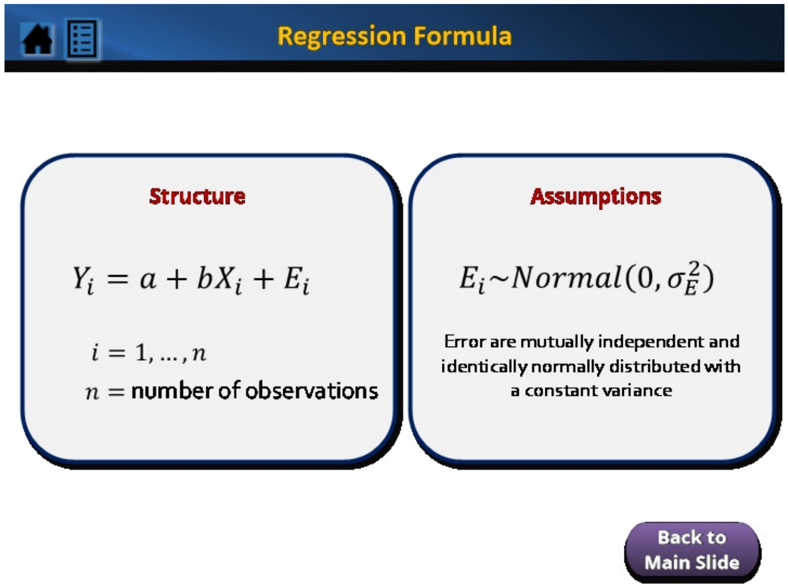
3.3 Model-data Fit (III)



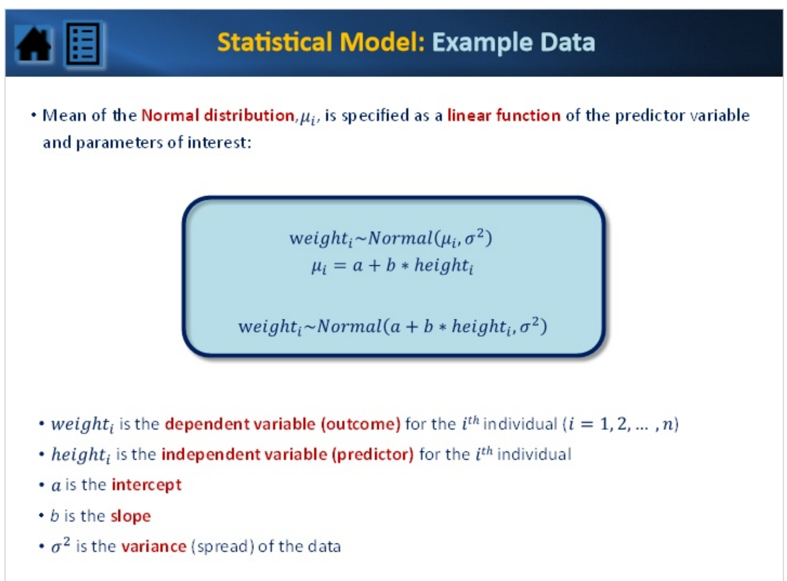
3.4 Example: Overview



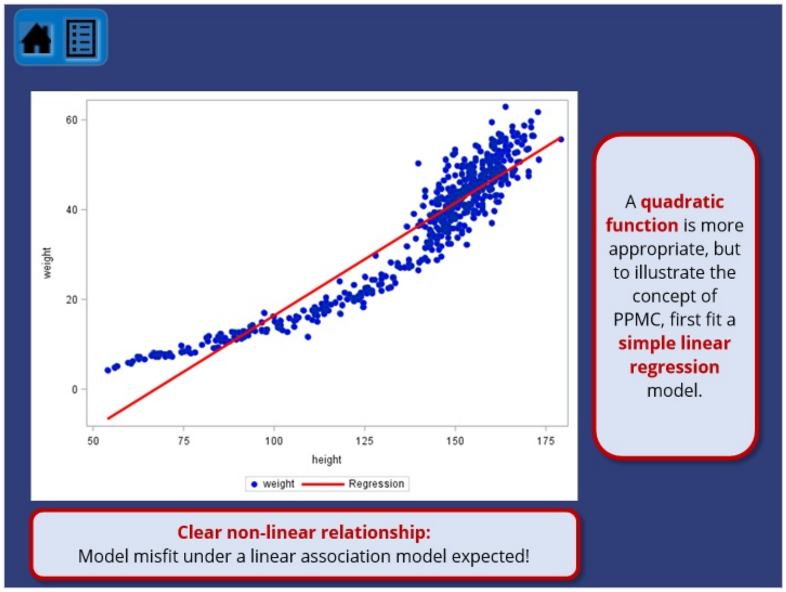
Regression Formula (Slide Layer)



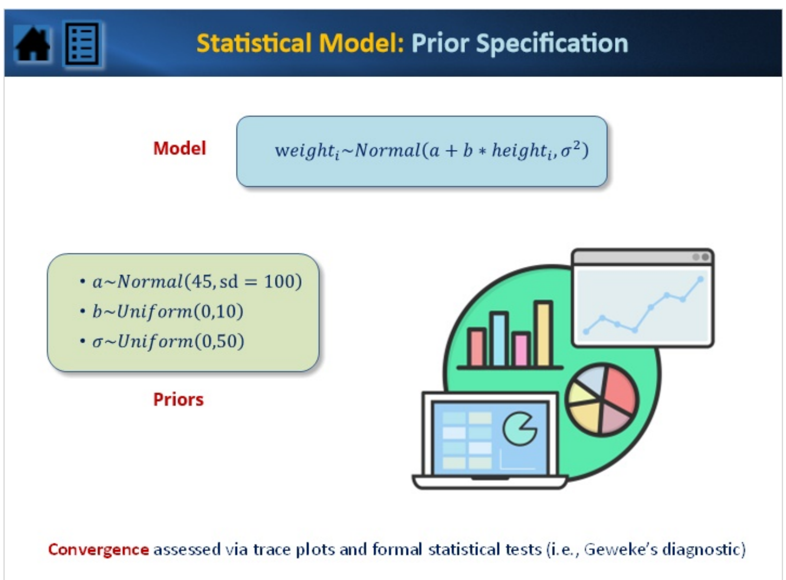
3.5 Regression Example (II)



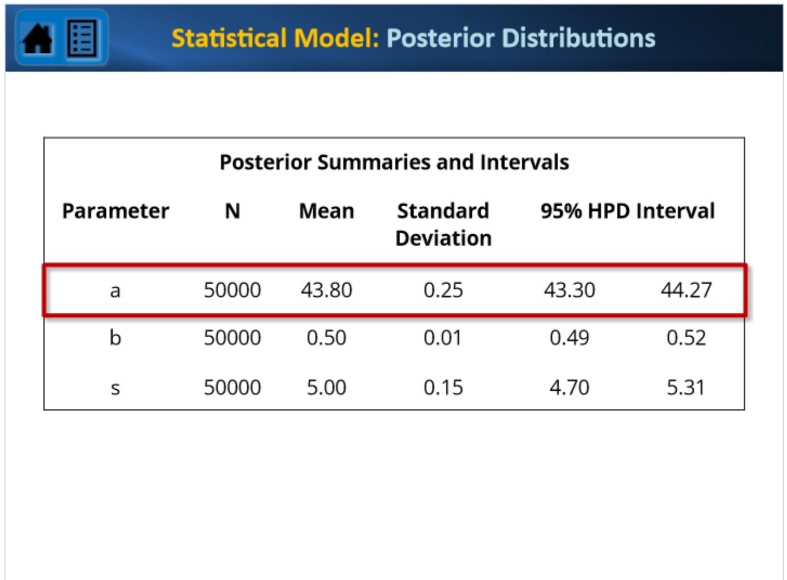
3.6 Regression Example (I)



3.7 Regression Example (III)



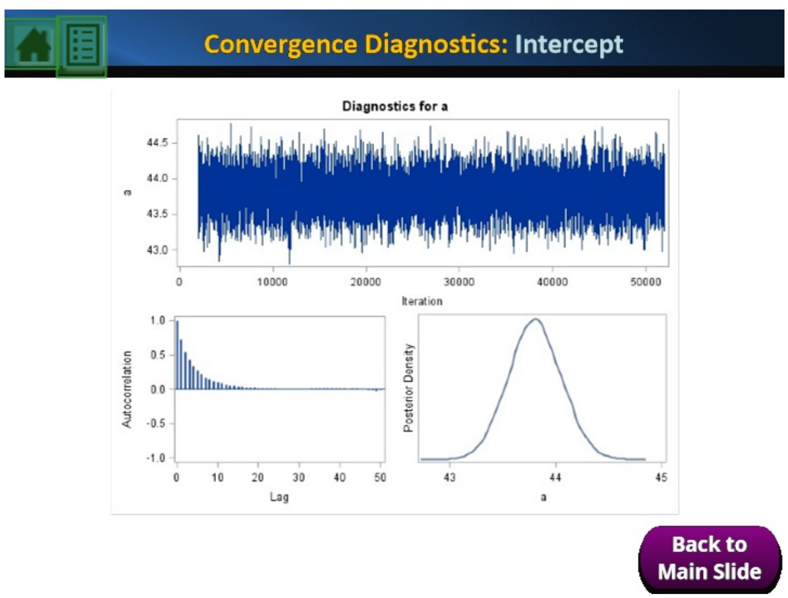
3.8 Regression Example (IV)



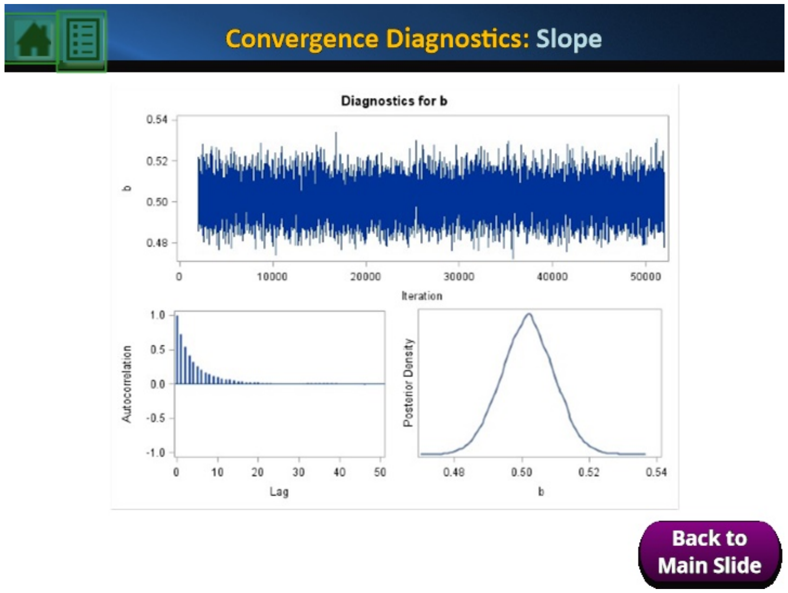
Posterior Summaries and Intervals

Parameter	N	Mean	Standard Deviation	95% HPD Interval	
a	50000	43.80	0.25	43.30	44.27
b	50000	0.50	0.01	0.49	0.52
s	50000	5.00	0.15	4.70	5.31

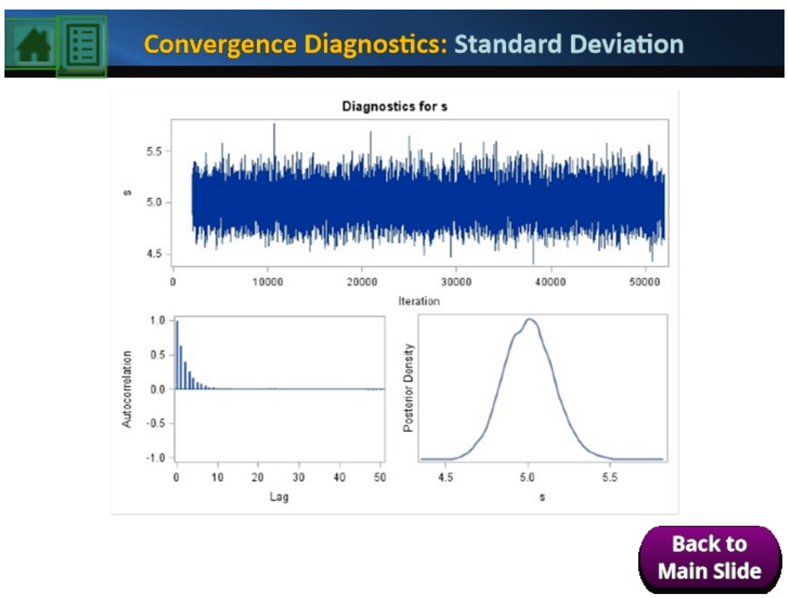
Diag a (Slide Layer)



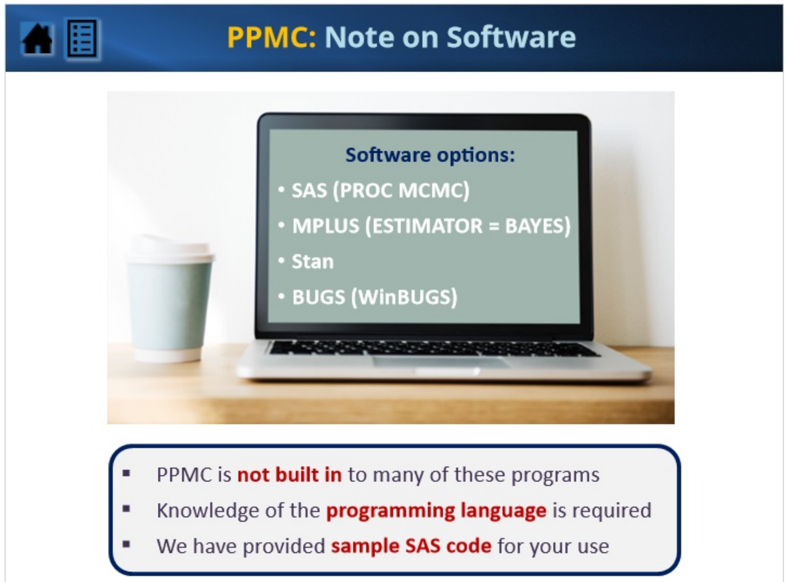
Diag b (Slide Layer)



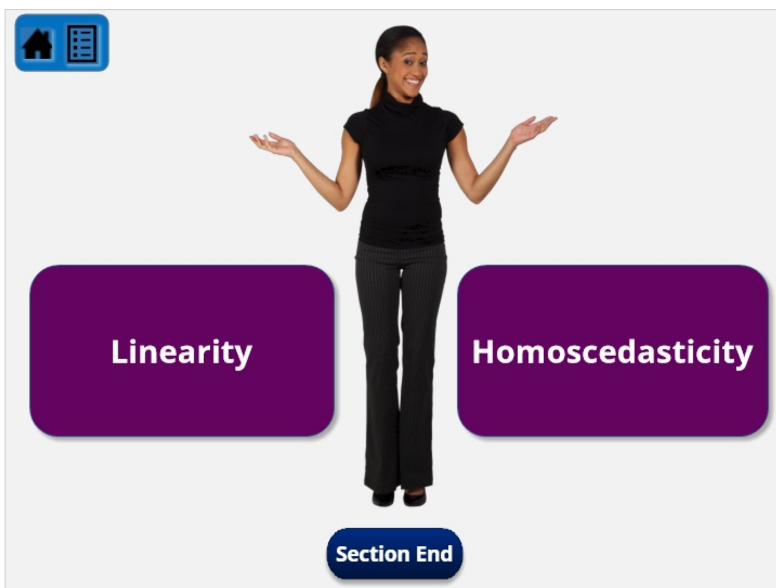
Diag s (Slide Layer)



3.9 PPMC: Computational Steps (II)

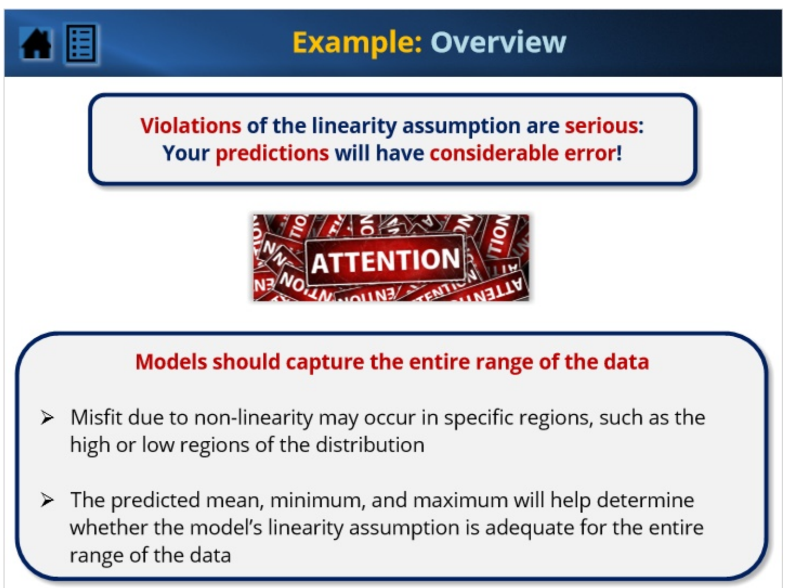


3.10 Topic Selection

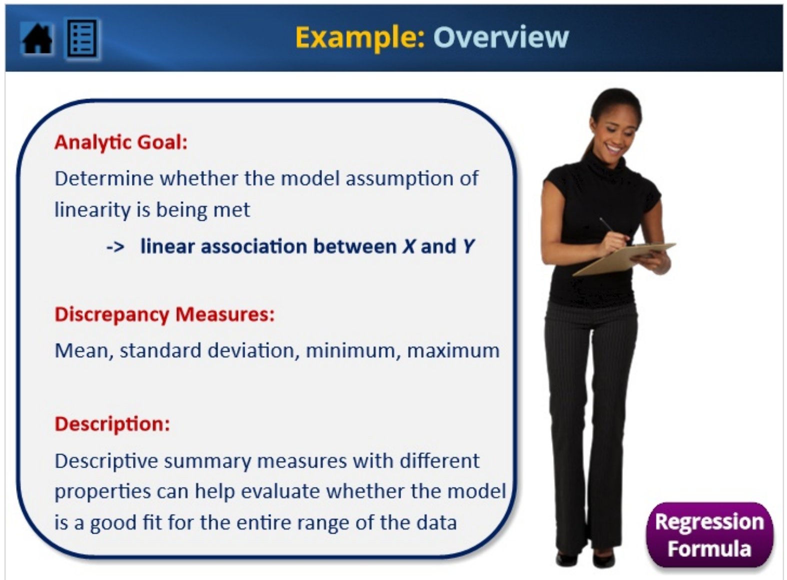


3.11 Bookmark: Linearity

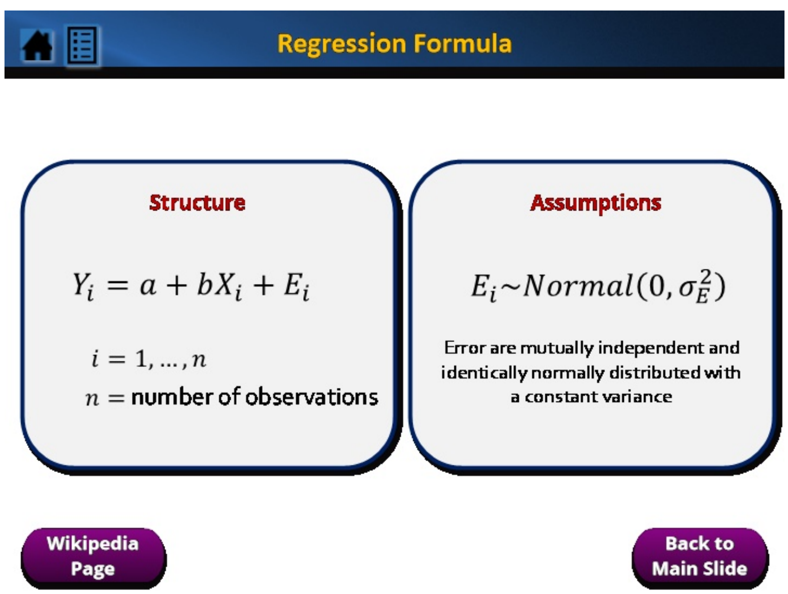
3.12 Untitled Slide



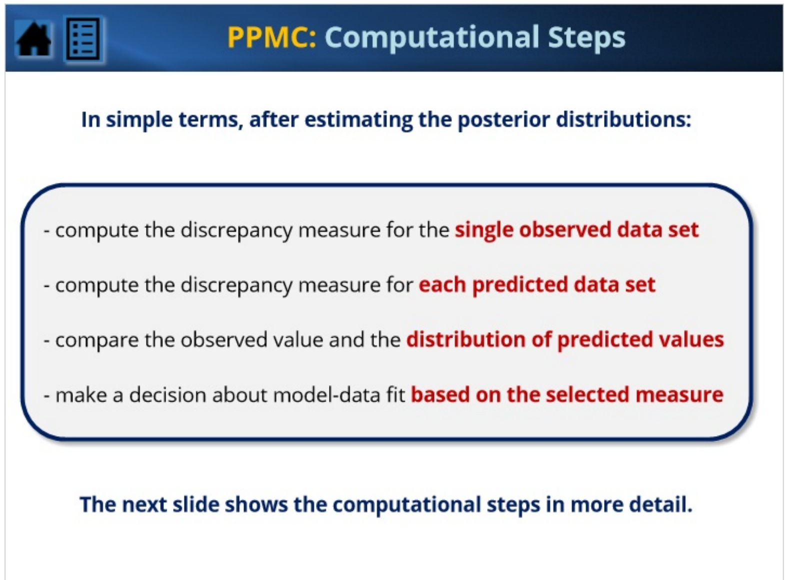
3.13 Example: Overview



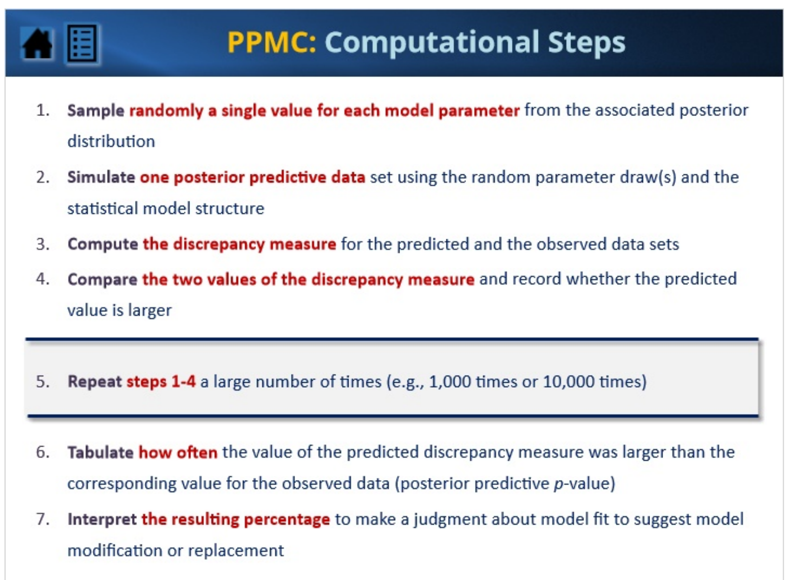
Regression Formula (Slide Layer)



3.14 PPMC: Computational Steps (I)



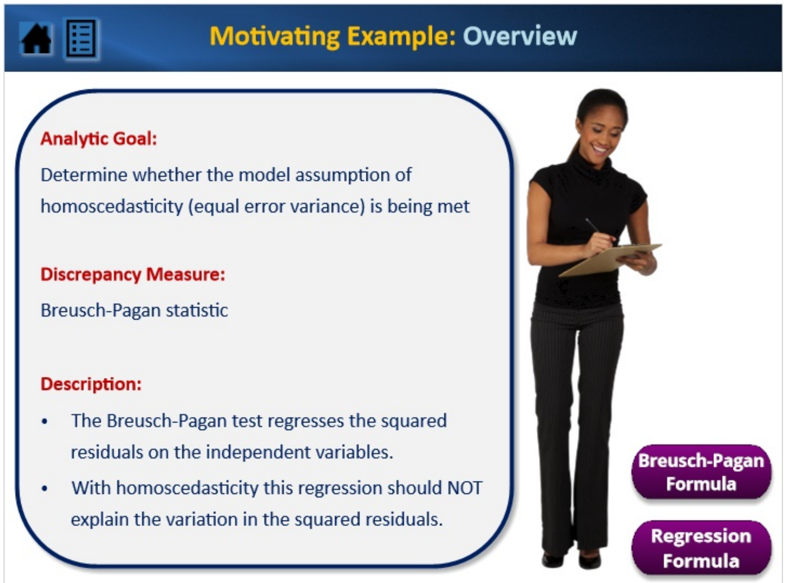
3.15 PPMC: Computational Steps (II)



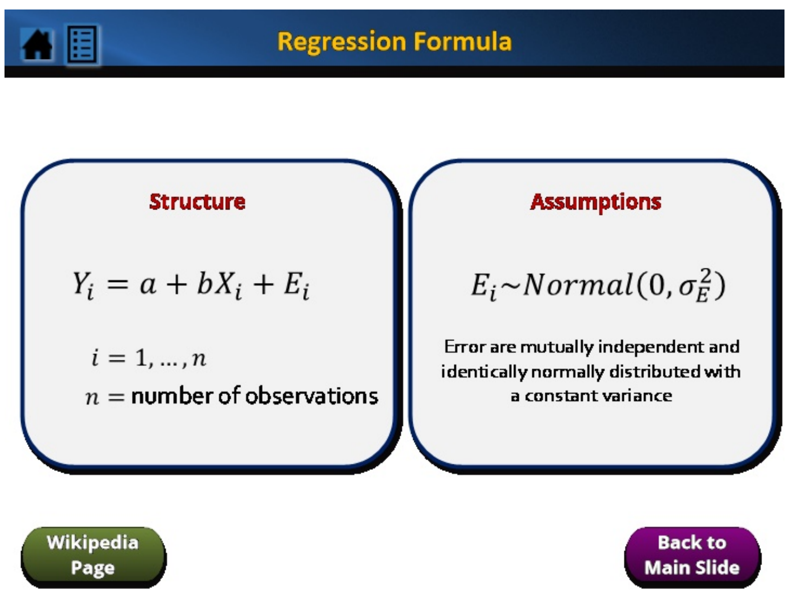
3.16 Bookend: Linearity

3.17 Bookmark: Homoscedasticity

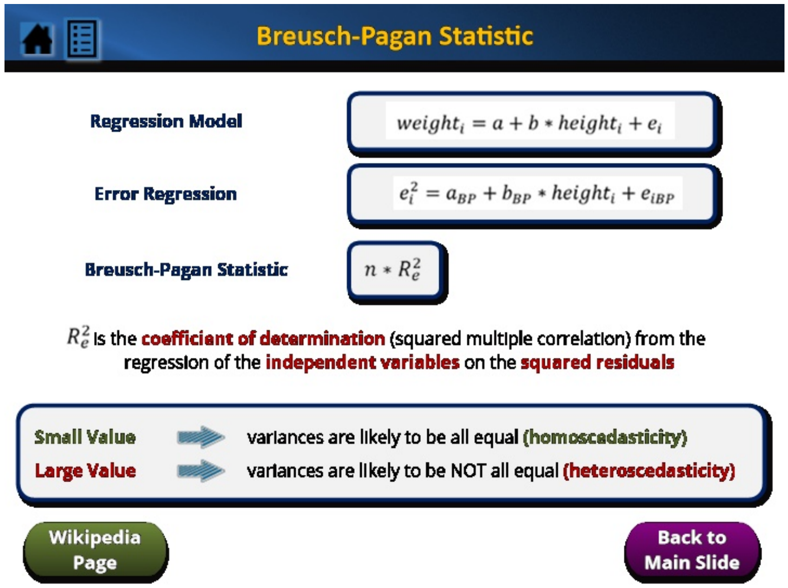
3.18 Example: Overview



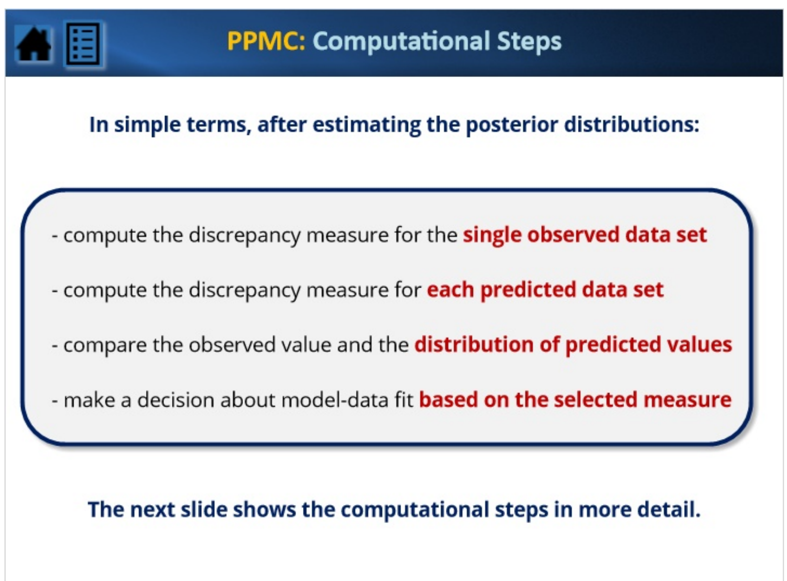
Regression Formula (Slide Layer)



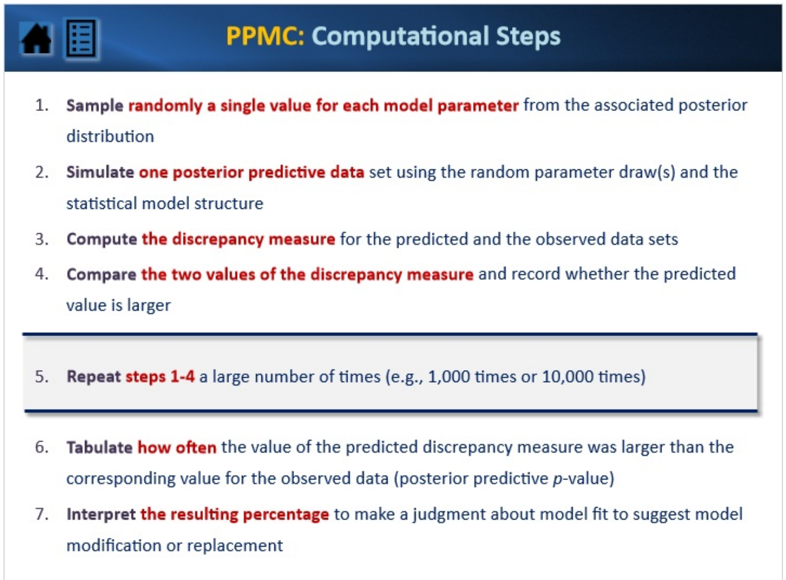
Breusch-Pagan Formula (Slide Layer)



3.19 PPMC: Computational Steps (I)



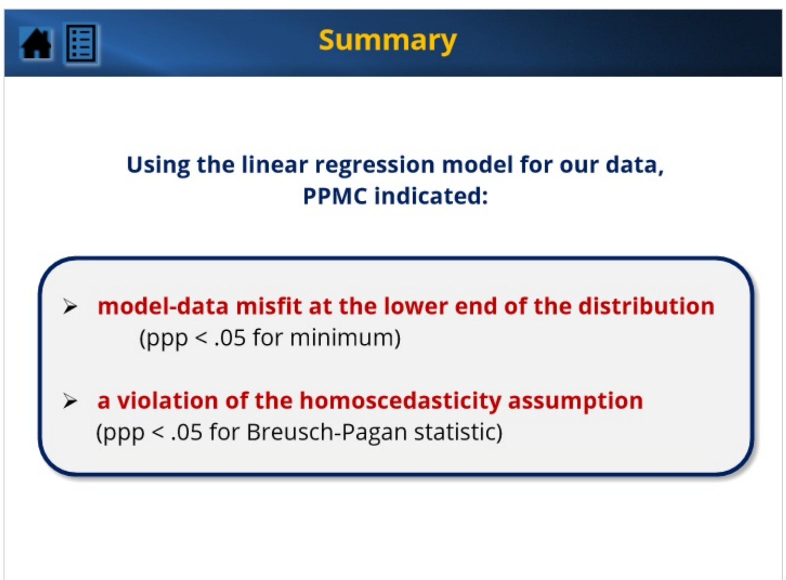
3.20 PPMC: Computational Steps (II)



3.21 Bookend: Homoscedasticity

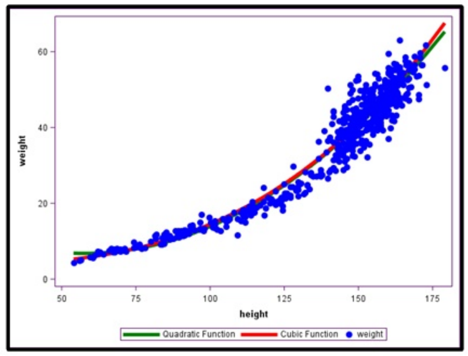
3.22 Bookmark: Summary

3.23 Summary



3.24 Other Models: Structure

Other Models: Structure



$$weight = a + b_1 * height + b_2 * height^2$$

$$weight = a + b_1 * height + b_2 * height^2 + b_3 * height^3$$

3.25 Other Models: PPMC Values

Other Models: PPMC Values

Discrepancy Statistic	Linear	Quadratic	Cubic
Mean	.540	.479	.480
Minimum	<.001	.249	.456
Maximum	.678	.900	.860
Standard Deviation	.770	.480	.540
Breusch-Pagan	.040	.049	.023

Quadratic and cubic models fit the lower range better but still have problems of heteroscedasticity!

3.26 Other Models: Conclusions

Other Models: Conclusions

Homoscedasticity

- The **ppp values** for the Breush-Pagan statistic are **below .05** for the linear, quadratic, and cubic models
- The issue of **heteroscedasticity has not been solved** by use of the quadratic or cubic model

Linearity

- The **ppp values** for the minimum are **.249** and **.456** for the quadratic and cubic models
- These models are a **better fit to data at the lower end** of the distribution than the simple linear model

3.27 Summary

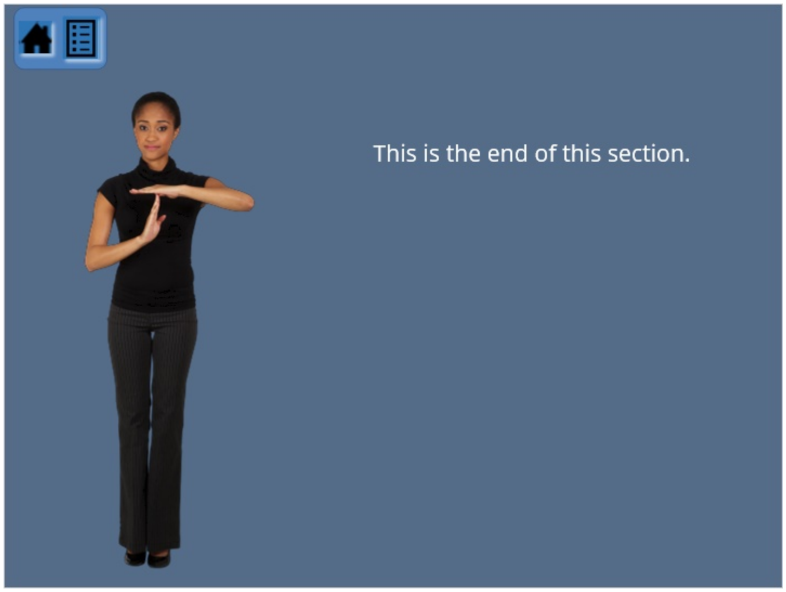
Final Recommendation

Choosing the appropriate statistic matters!

- The mean as a discrepancy statistic would not highlight the data's heteroscedasticity
- The Breusch-Pagan test as discrepancy statistic does not indicate violations of the independence assumption

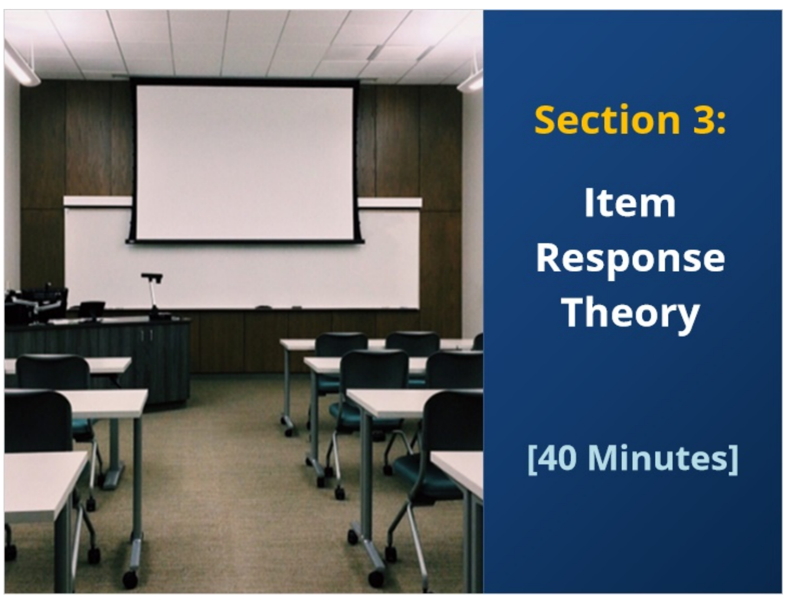
Regardless of the discrepancy statistic or model, the PPMC process is identical.

3.28 Bookend: Regression



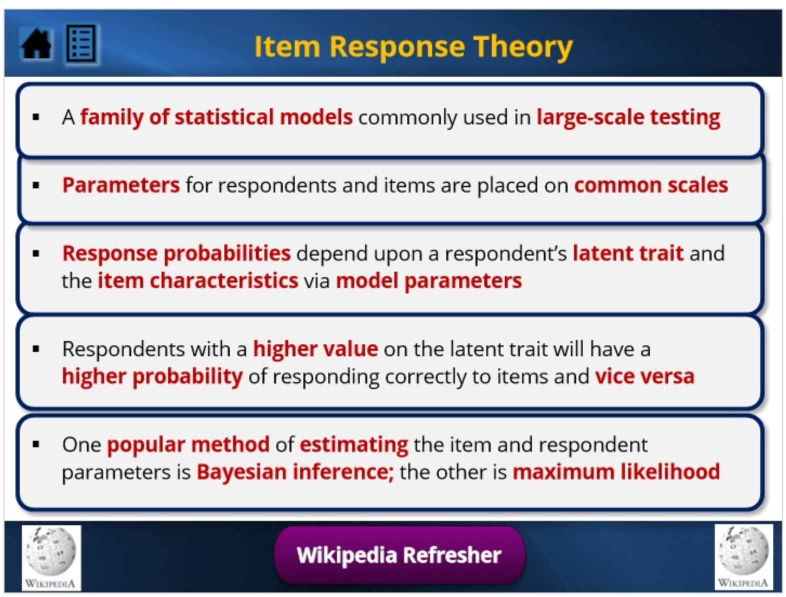
4. Item Response Theory

4.1 Cover: IRT

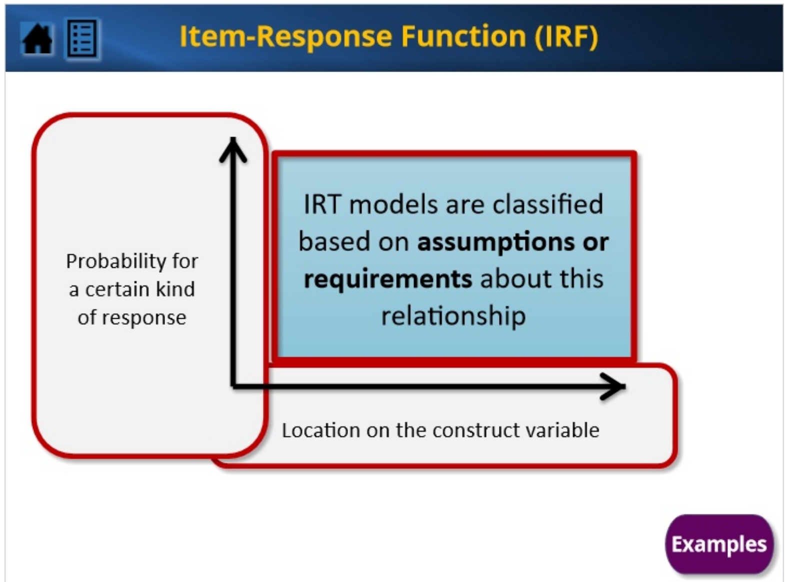


4.2 Objectives: IRT

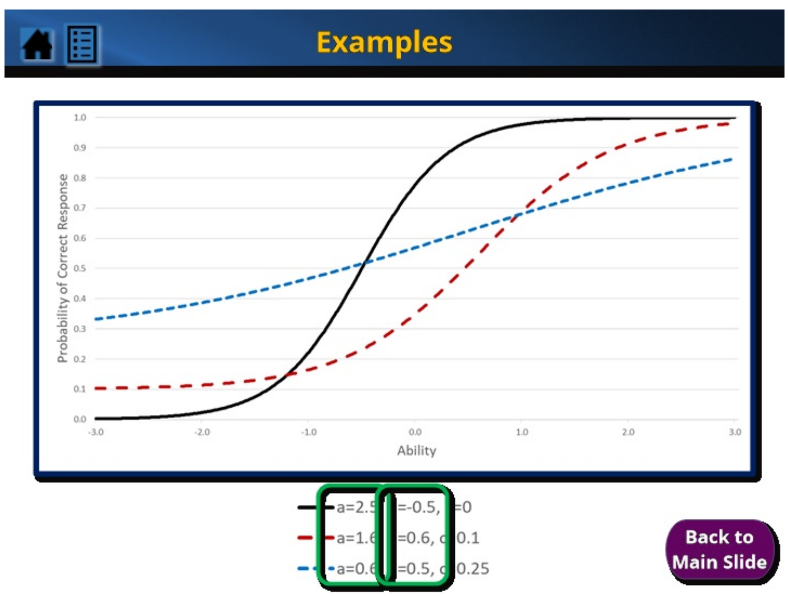
4.3 IRT Models (I)



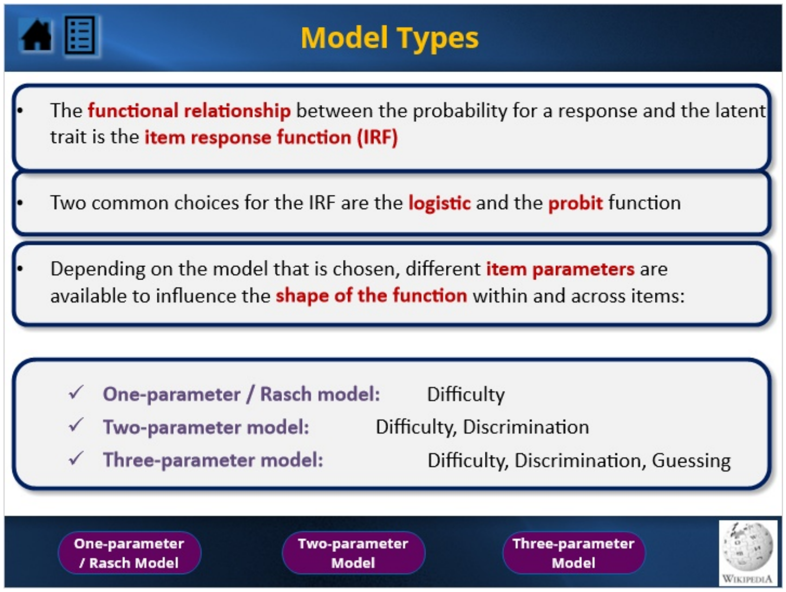
4.4 IRT Models (II)



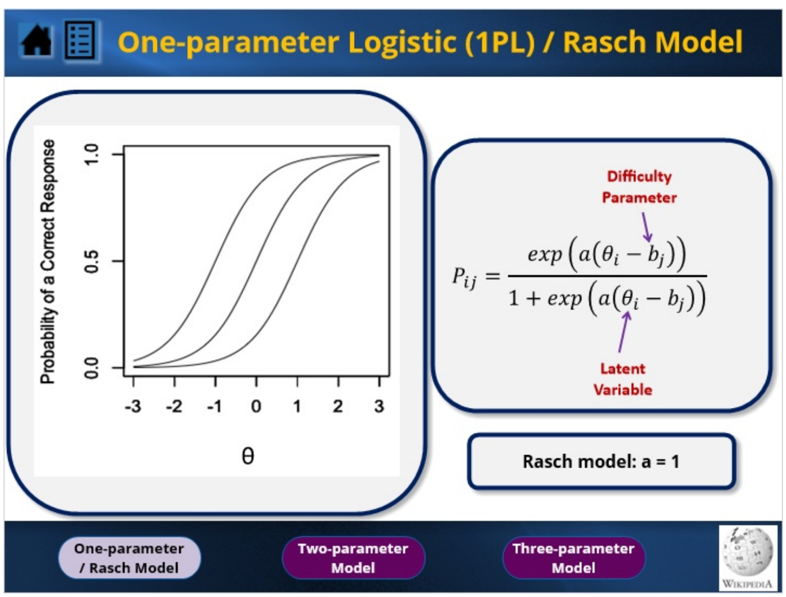
Examples (Slide Layer)



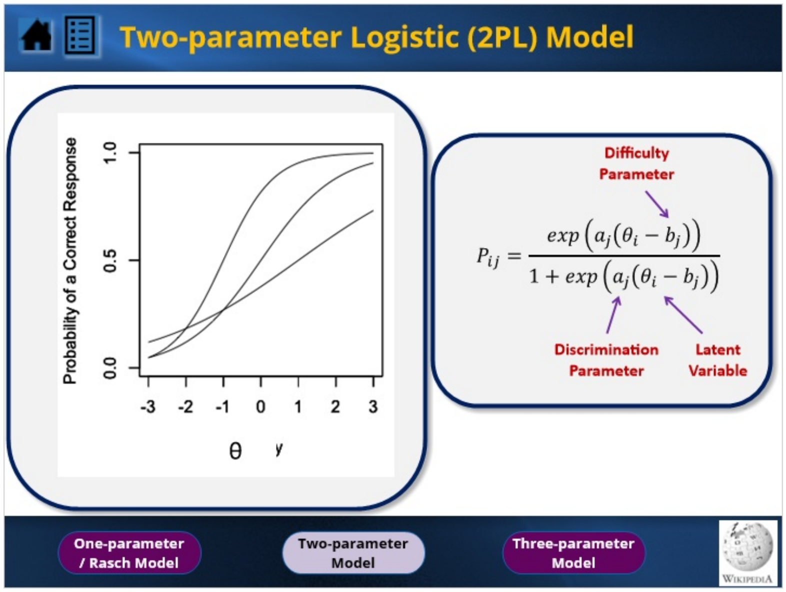
4.5 Parametric Item Response Theory



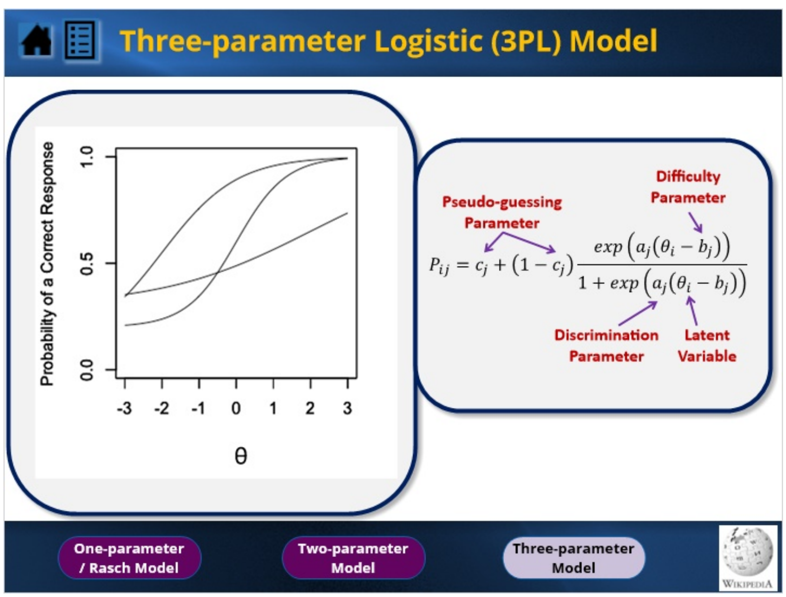
4.6 One-parameter Model



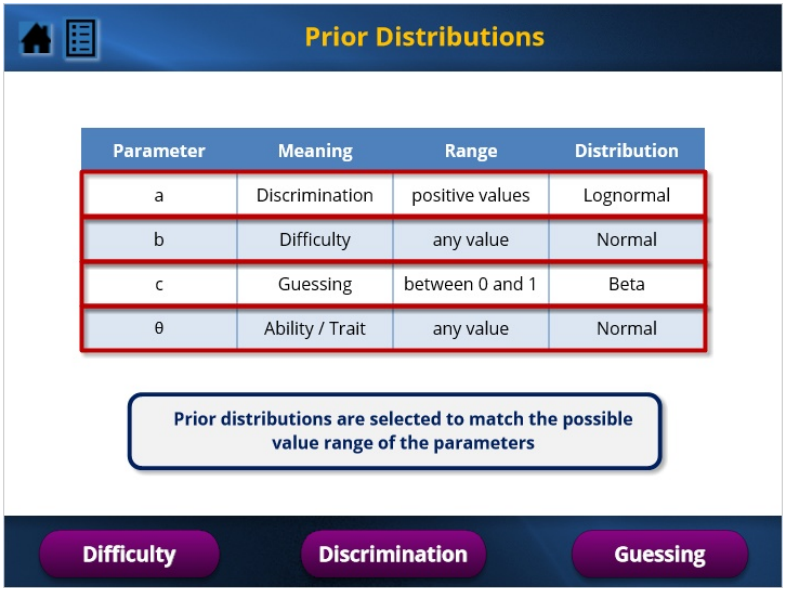
4.7 Two-parameter Model



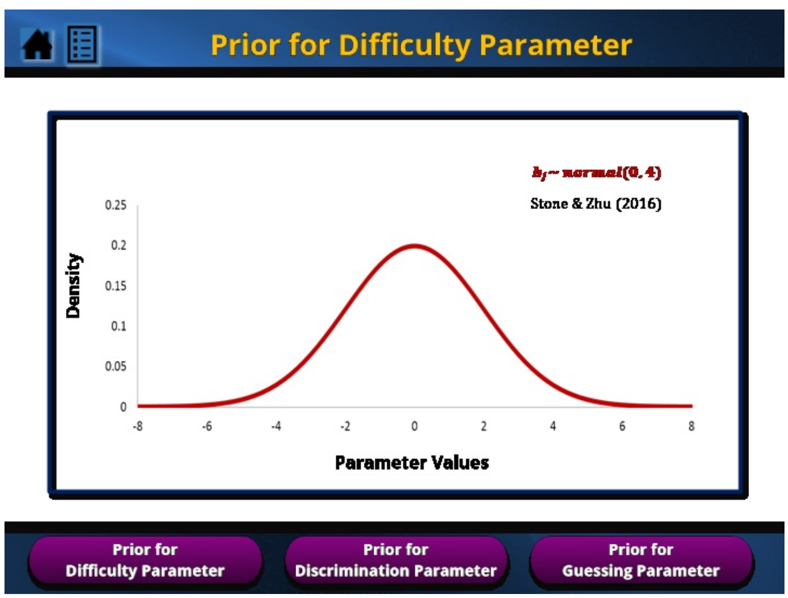
4.8 Three-parameter Model



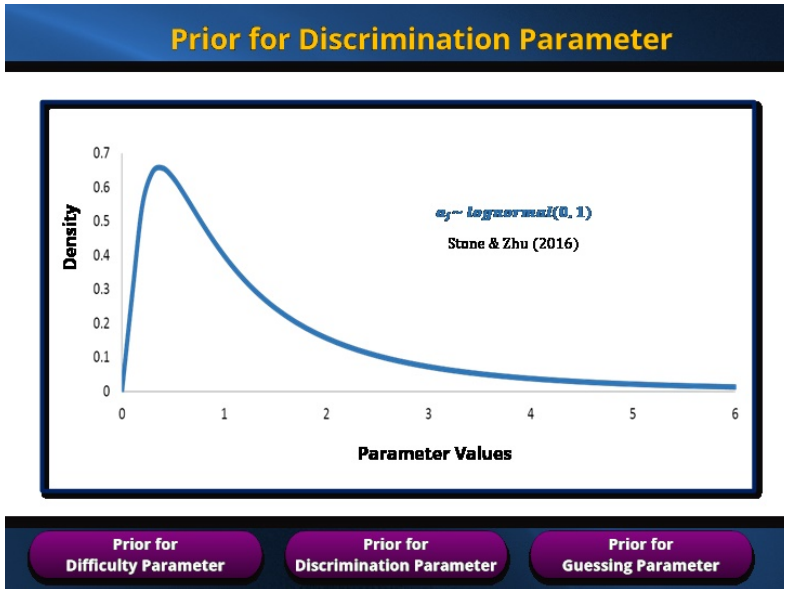
4.9 Prior Distributions (III)



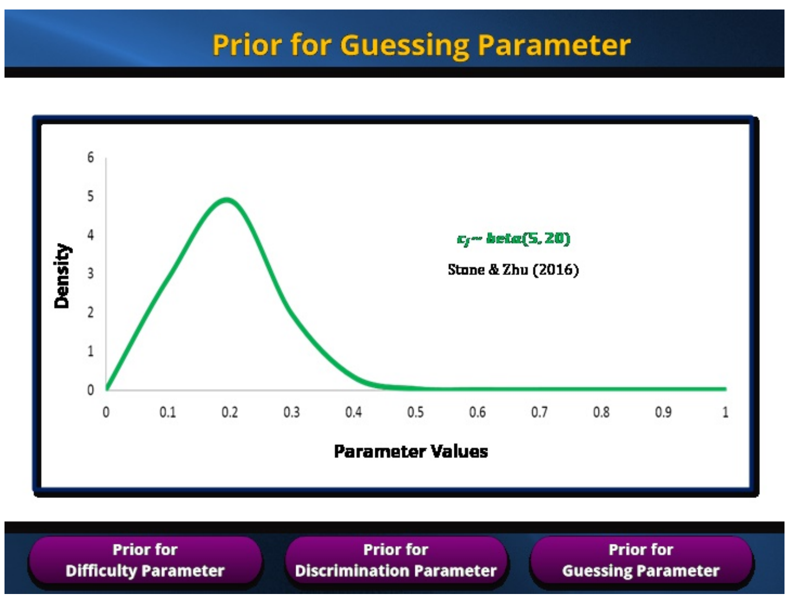
Prior Difficulty (Slide Layer)



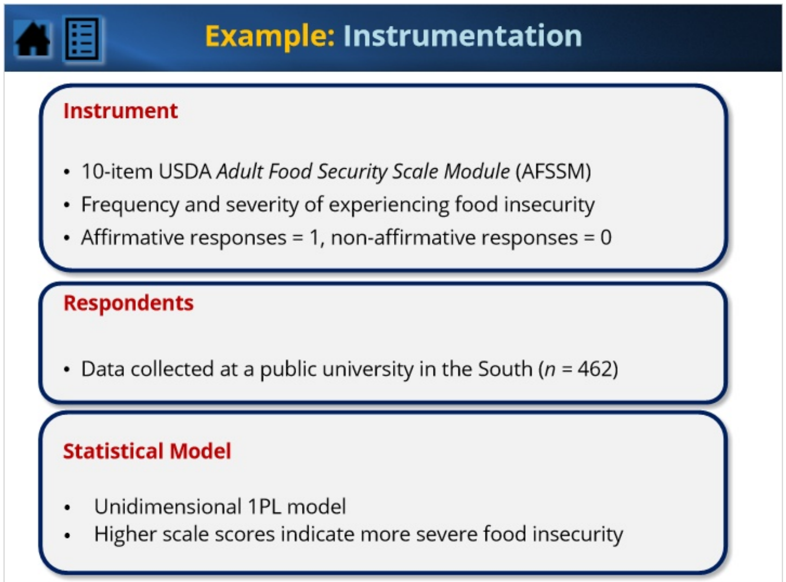
Prior Discrimination (Slide Layer)



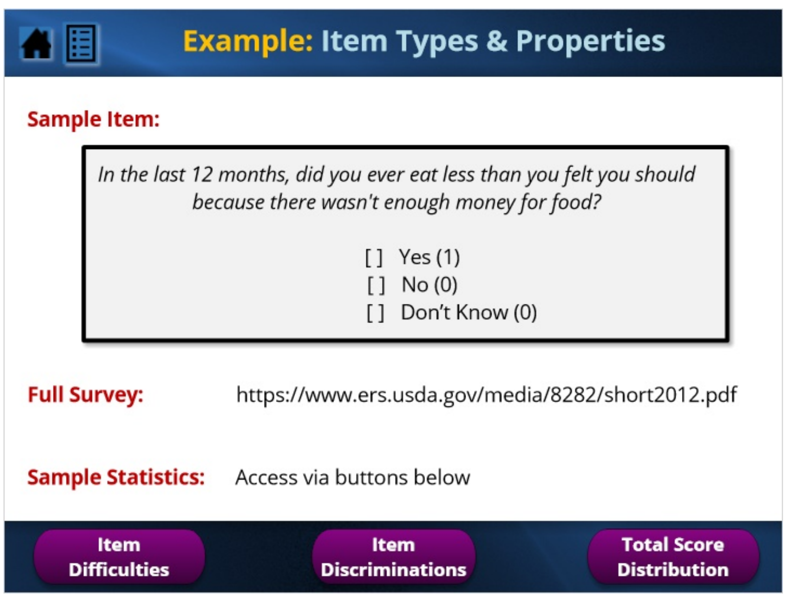
Prior Guessing (Slide Layer)



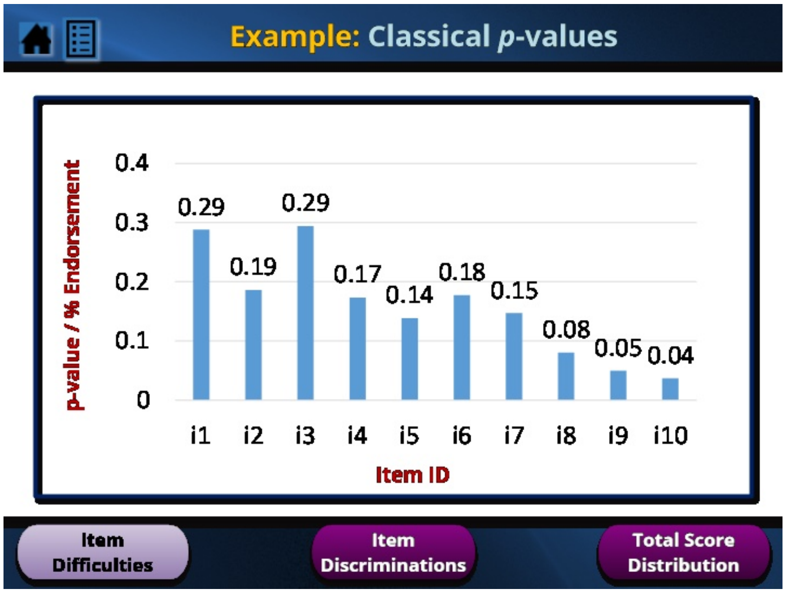
4.10 Example: Instrumentation



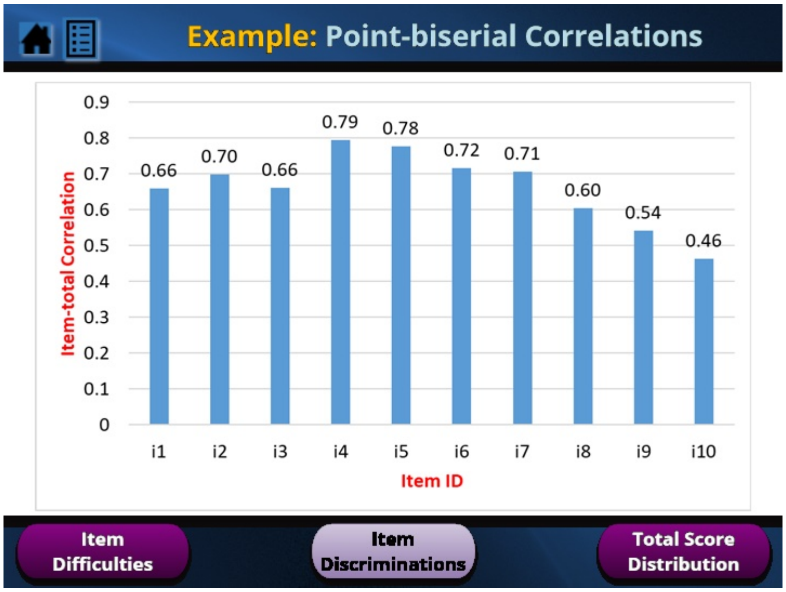
4.11 Example: Sample Items



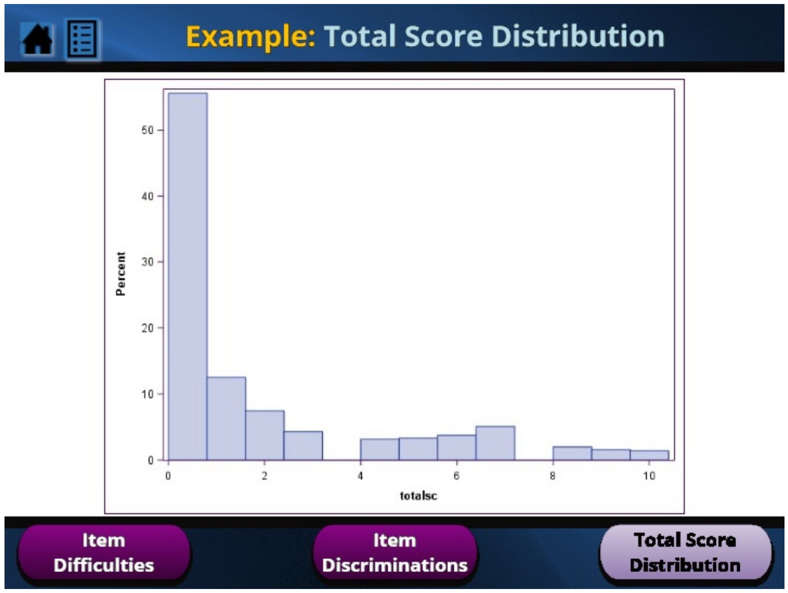
Difficulty (Slide Layer)



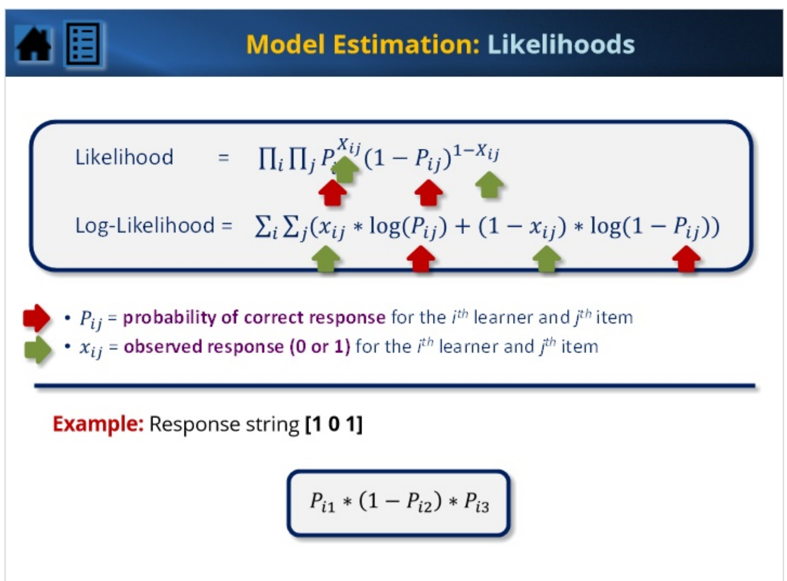
Discrimination (Slide Layer)



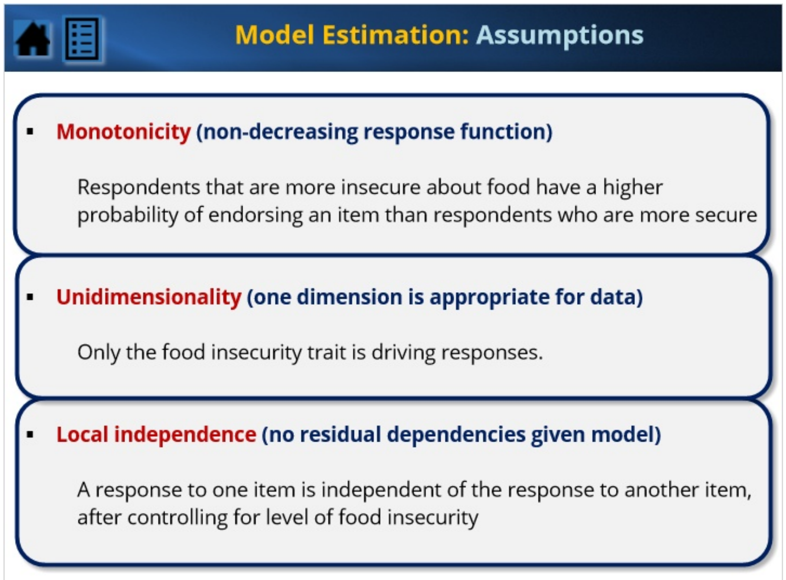
Total score distribution (Slide Layer)



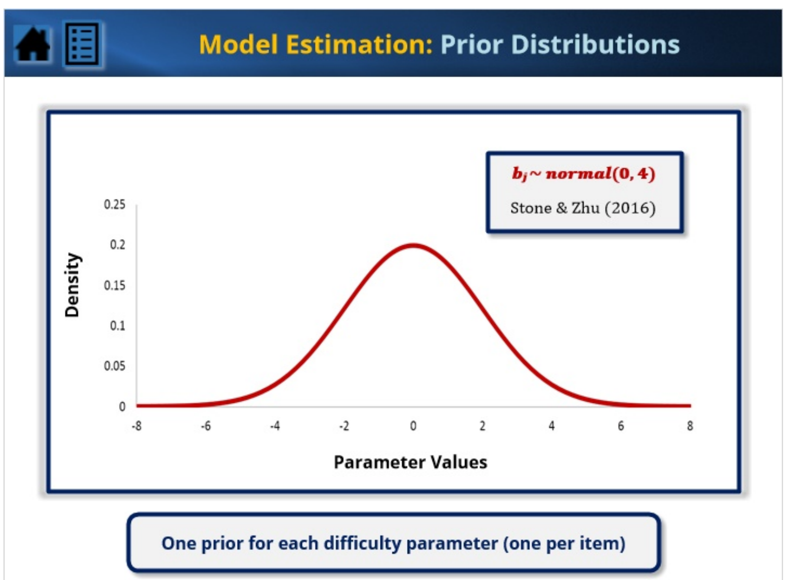
4.12 Model Estimation: Structure & Assumptions



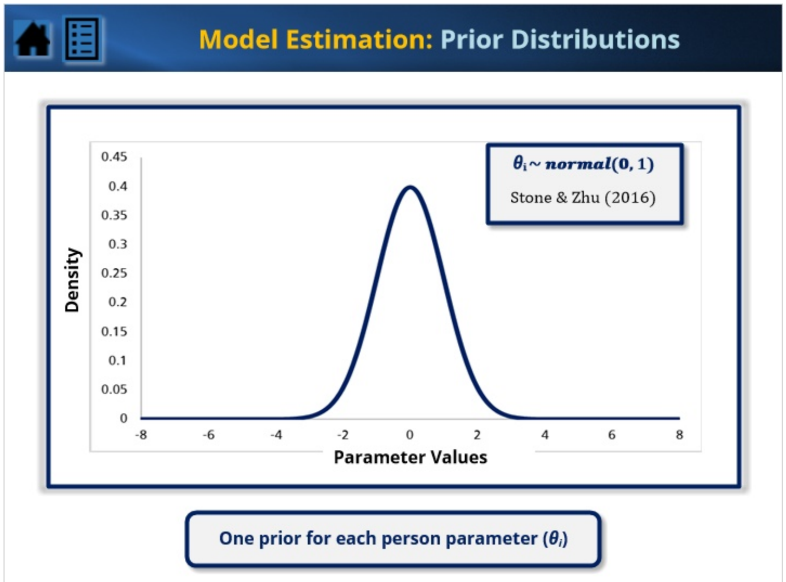
4.13 Model Estimation: Structure & Assumptions



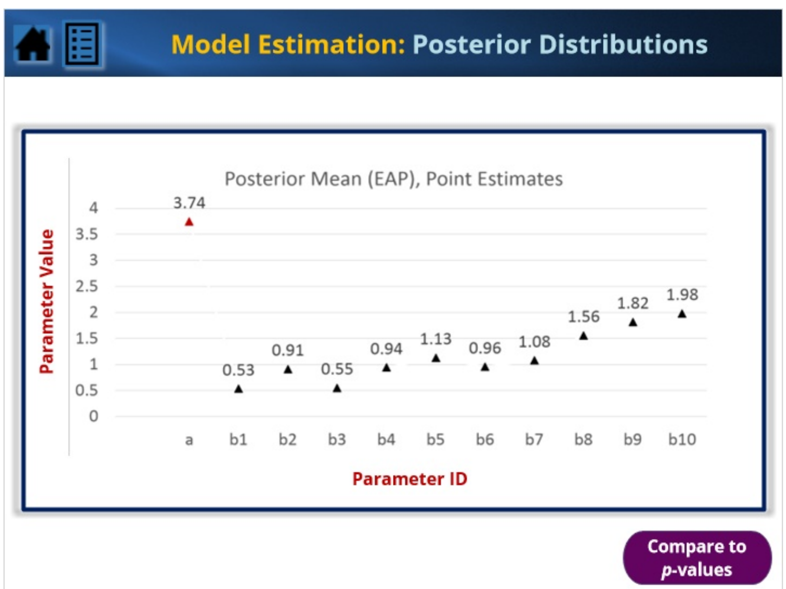
4.14 Model Estimation: Posterior Distributions



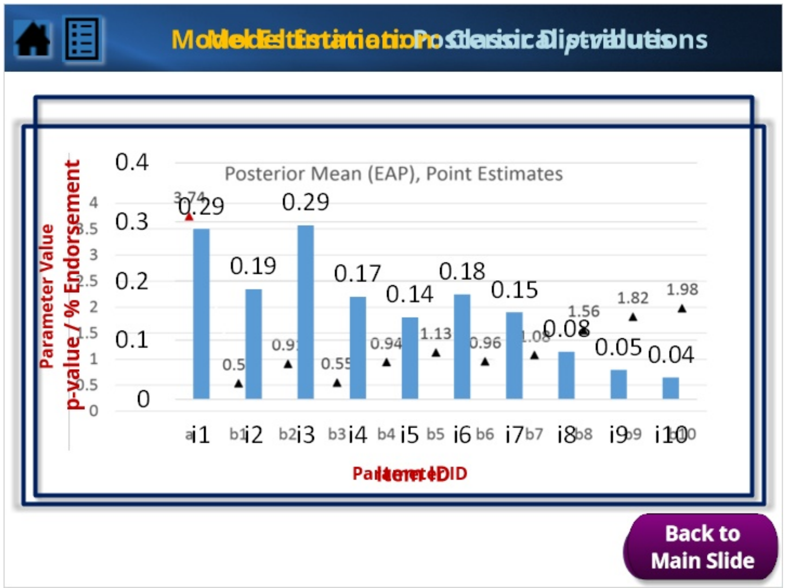
4.15 Model Estimation: Posterior Distributions



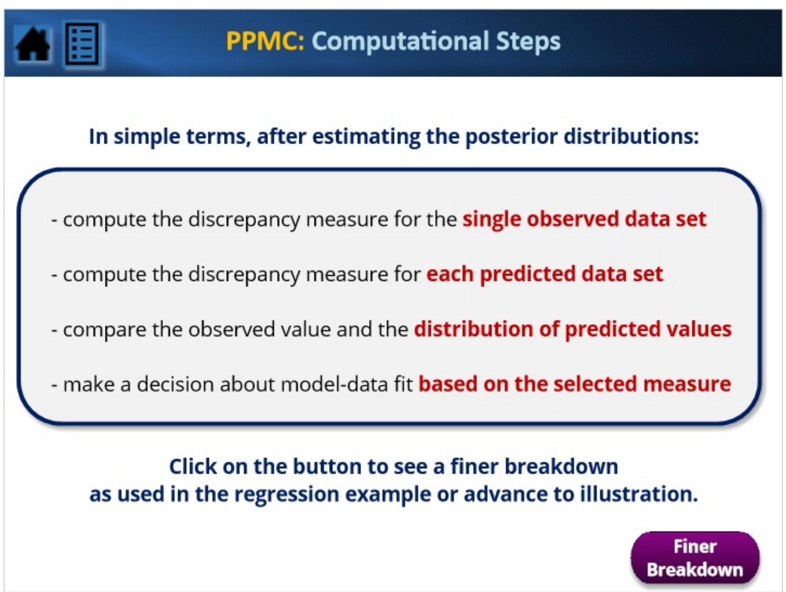
4.16 Model Estimation: Posterior Distributions



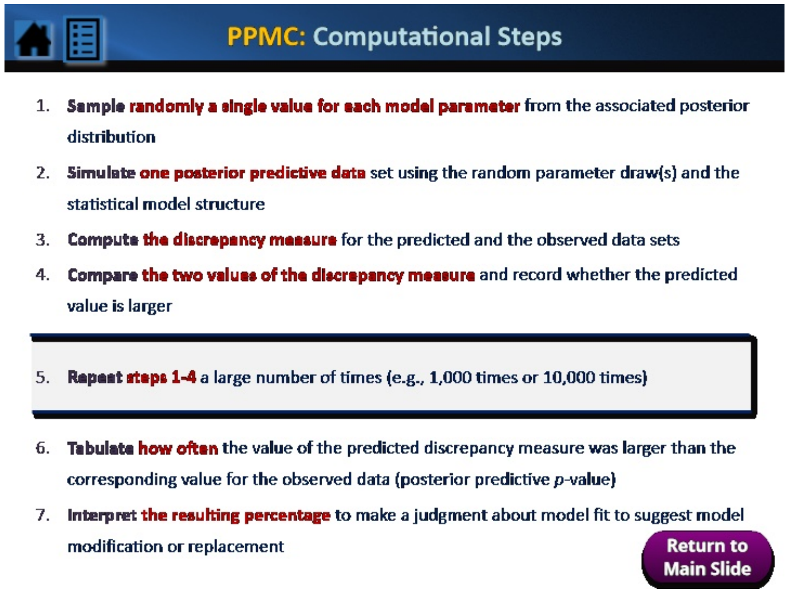
p-values (Slide Layer)



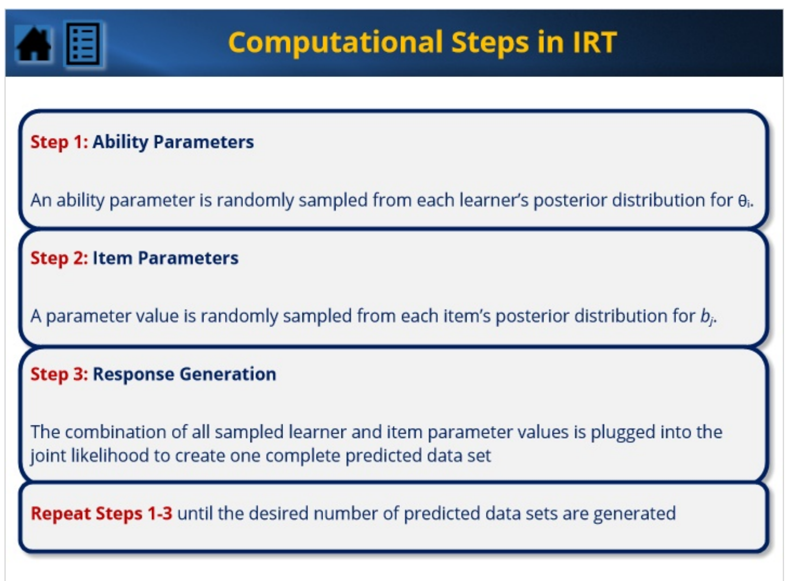
4.17 PPMC: Computational Steps (I)



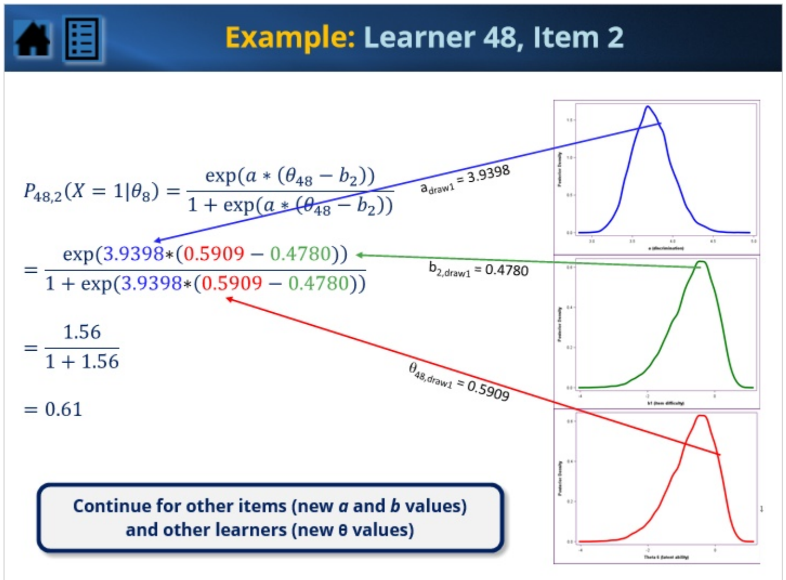
Finer Breakdown (Slide Layer)



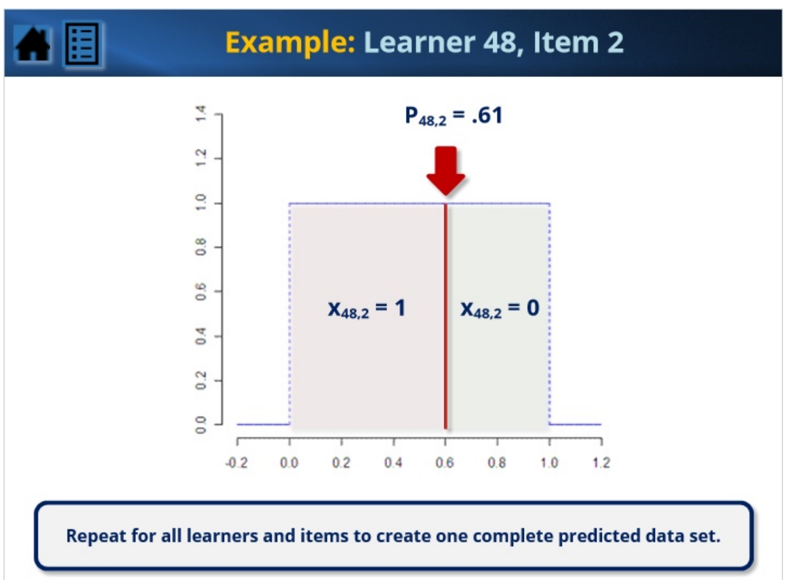
4.18 General Principles (II)



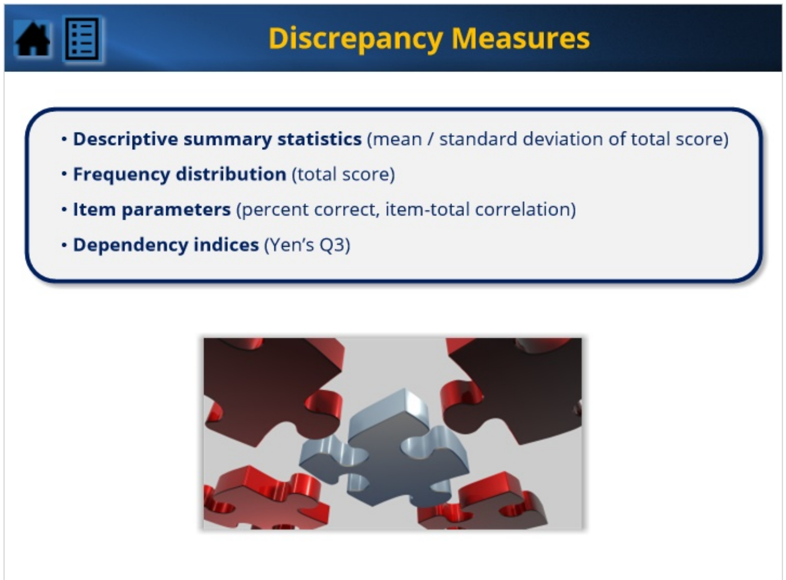
4.19 Computation (I)



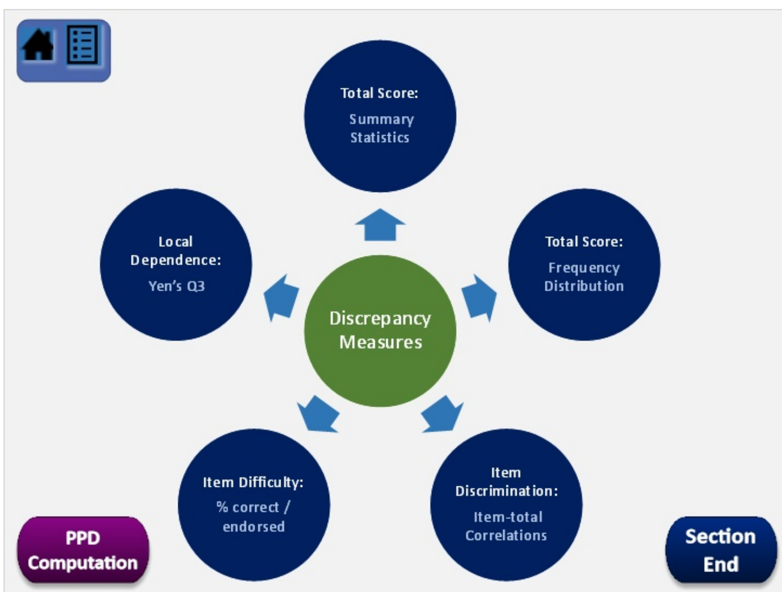
4.20 Computation (II)



4.21 Untitled Slide

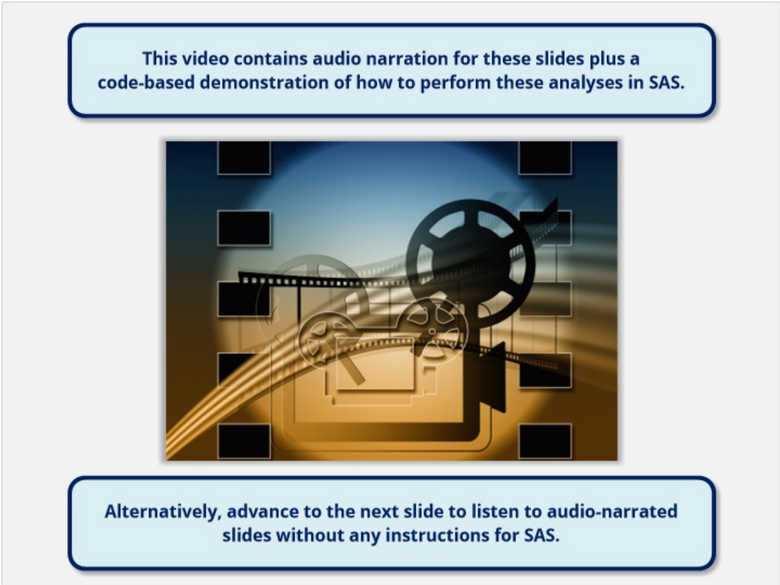


4.22 Discrepancy Measure Selection

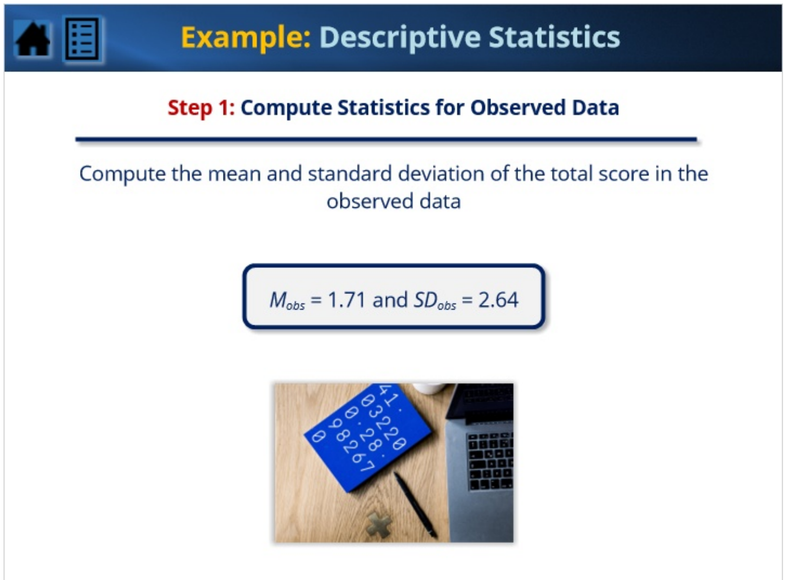


4.23 Bookmark: Descriptive Statistics

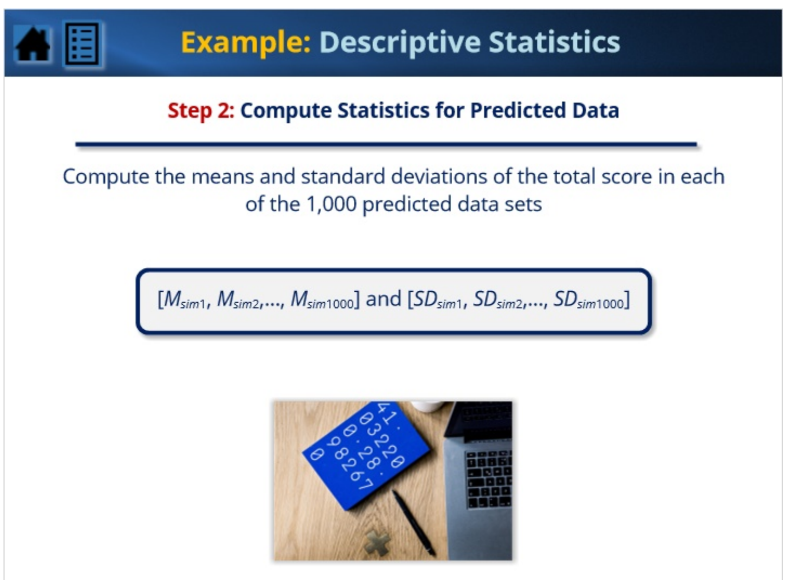
4.24 Video: Descriptive Statistics



4.25 Descriptive Statistics (I)



4.26 Descriptive Statistics (II)



4.27 Total Score Distribution (II)

Example: Total Score Distribution

Step 2a: Compute Frequencies for Predicted Data

Compute and save the total score frequencies for each of the 1,000 simulated data sets

4.28 Descriptive Statistics (III)

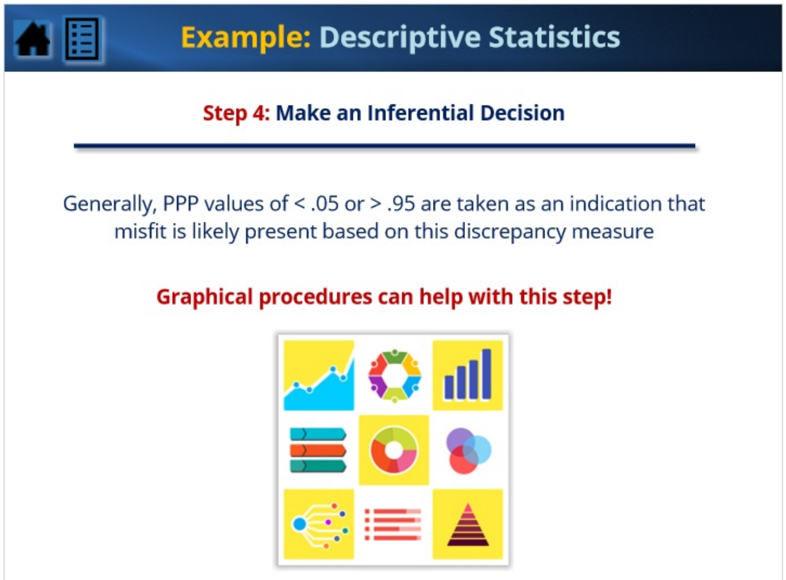
Example: Descriptive Statistics

Step 3: Compare Statistics for Observed and Predicted Data

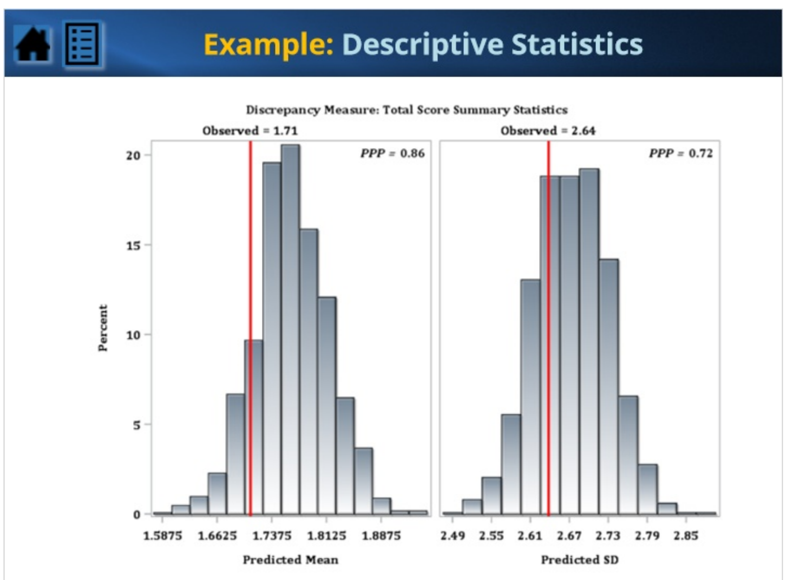
Compare the distribution of the 1,000 means and standard deviations from the predicted data to the single observed mean and standard deviation to obtain the Bayesian posterior predictive *p*-value (PPP)

$$PPP = Proportion(T(x^{sim}) > T(x)|x)$$

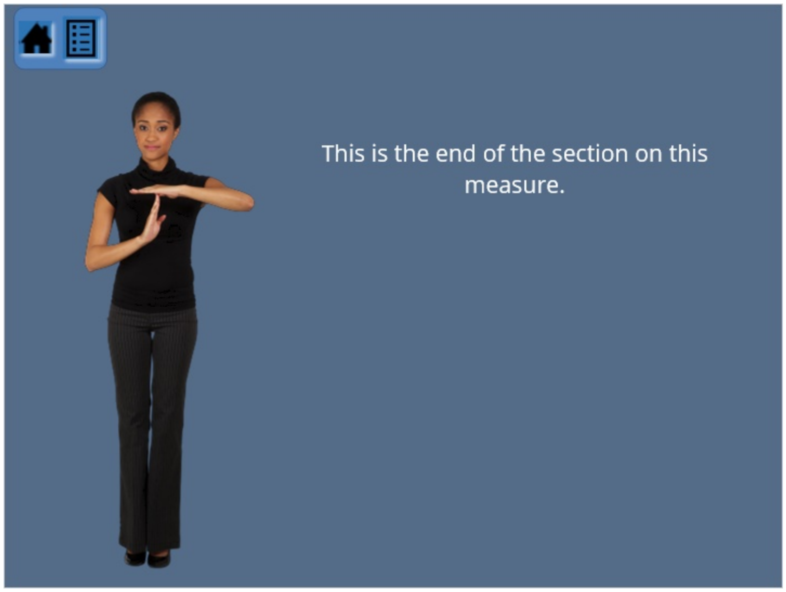
4.29 Descriptive Statistics (IV)



4.30 Descriptive Statistics (V)

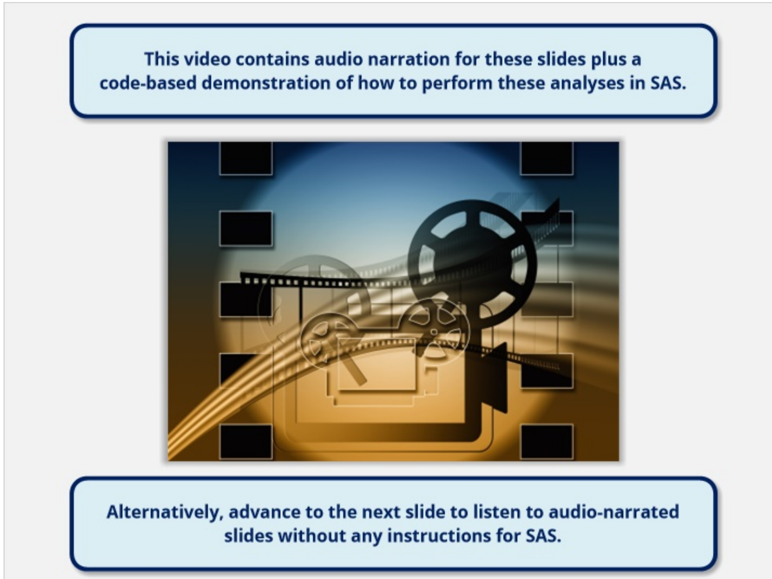


4.31 Bookend: Descriptive Statistics

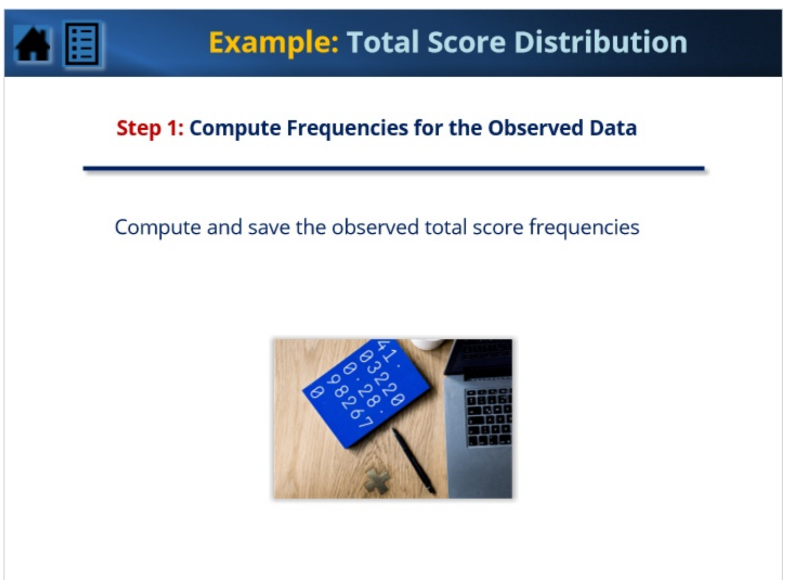


4.32 Bookmark: Total Score Distribution

4.33 Video: Total Score Distribution



4.34 Total Score Distribution (I)



4.35 Total Score Distribution (III)

Example: Total Score Distribution

Step 2b: Compute Percentiles for Predicted Data

Compute and save the 5^{th}, 50^{th}, and 95^{th} percentiles associated with each frequency distribution at each total score point (collapsing across replicated data sets)

4.36 Total Score Distribution (IV)

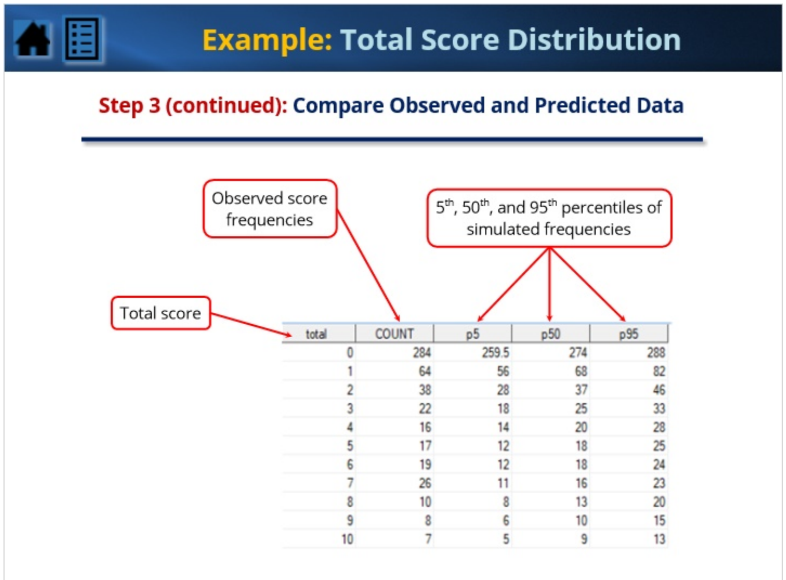
Example: Total Score Distribution

Step 3: Compare Statistics for Observed and Predicted Data

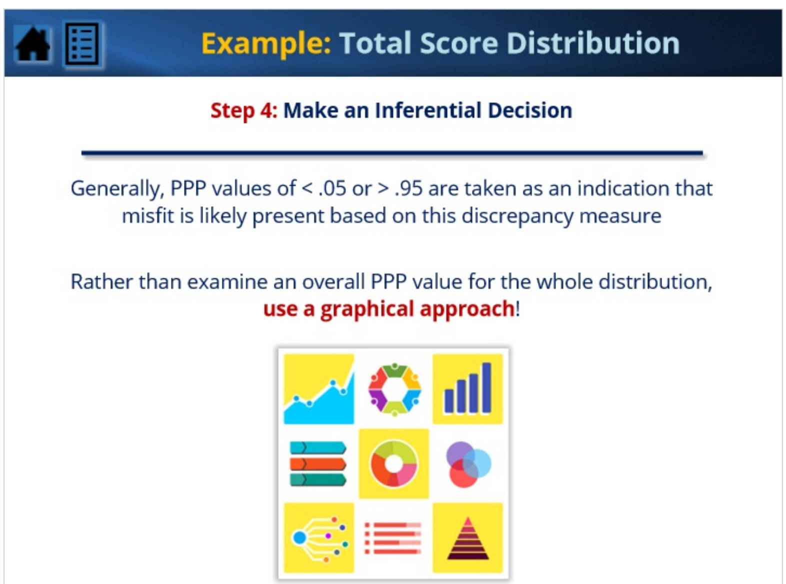
Compare the 1,000 frequency distributions from the predicted data to the single observed frequency distribution to obtain the Bayesian posterior predictive *p*-value (PPP)

$$PPP = Proportion(T(x^{sim}) > T(x)|x)$$

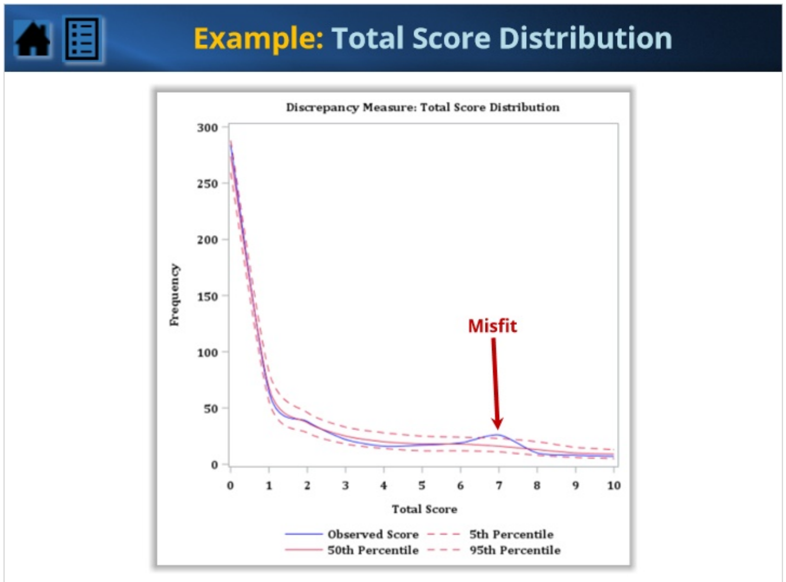
4.37 Total Score Distribution (V)



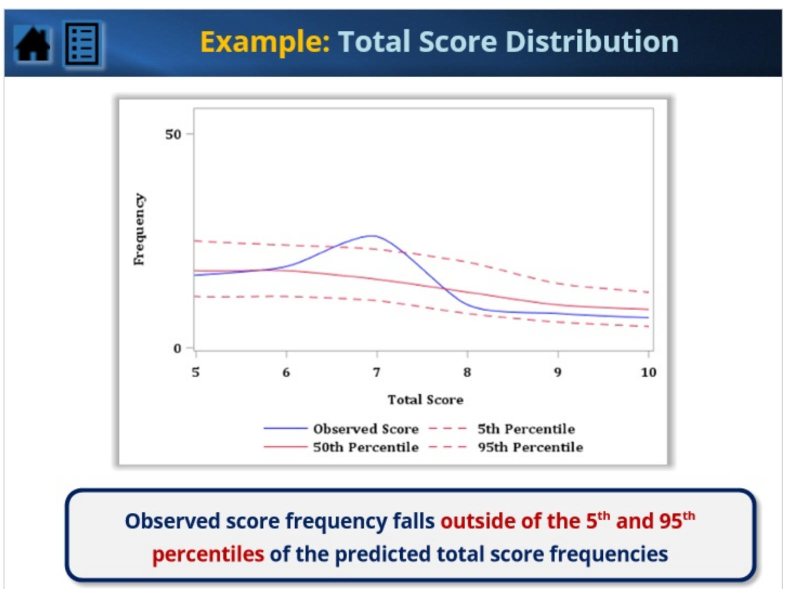
4.38 Total Score Distribution (VI)



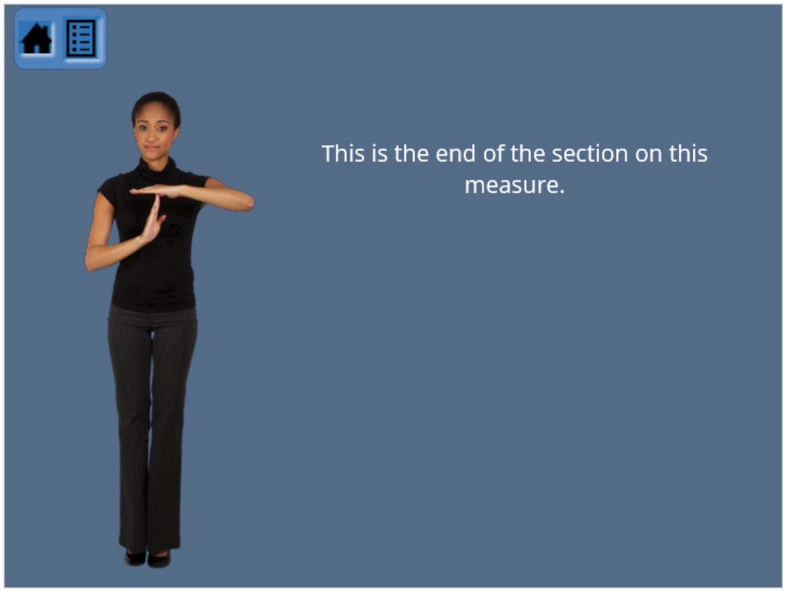
4.39 Total Score Distribution (VII)



4.40 Total Score Distribution (VIII)

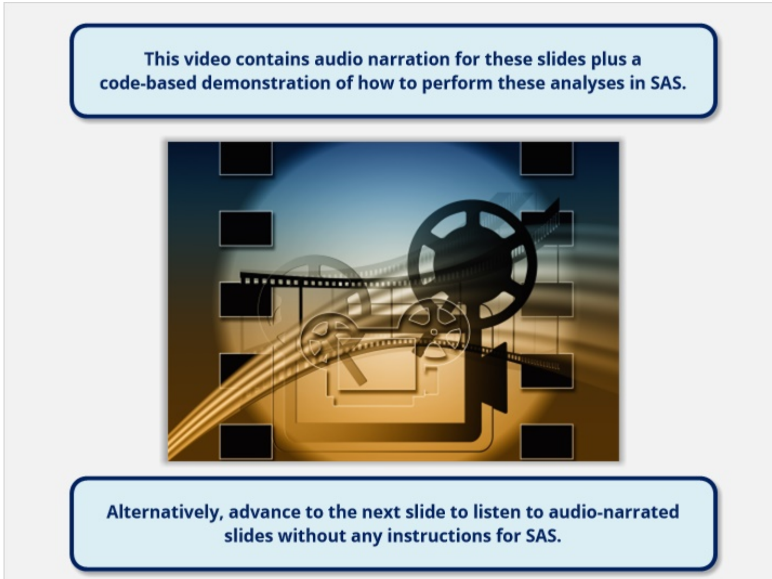


4.41 Bookend: Total Score Distribution

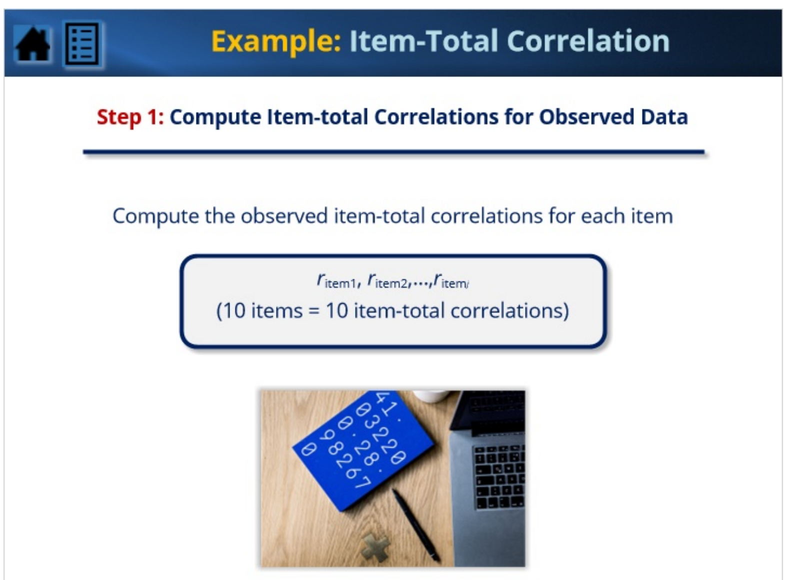


4.42 Bookmark: Item-total Correlations

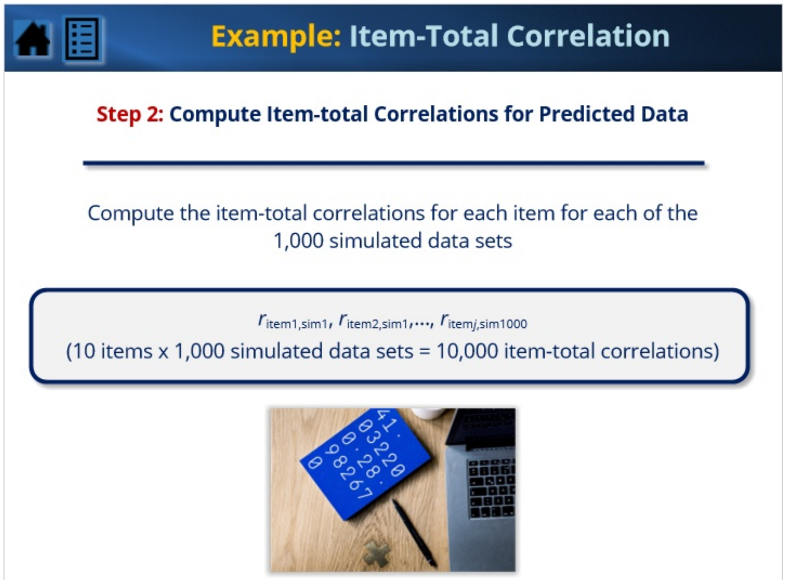
4.43 Video: Item-total Correlations



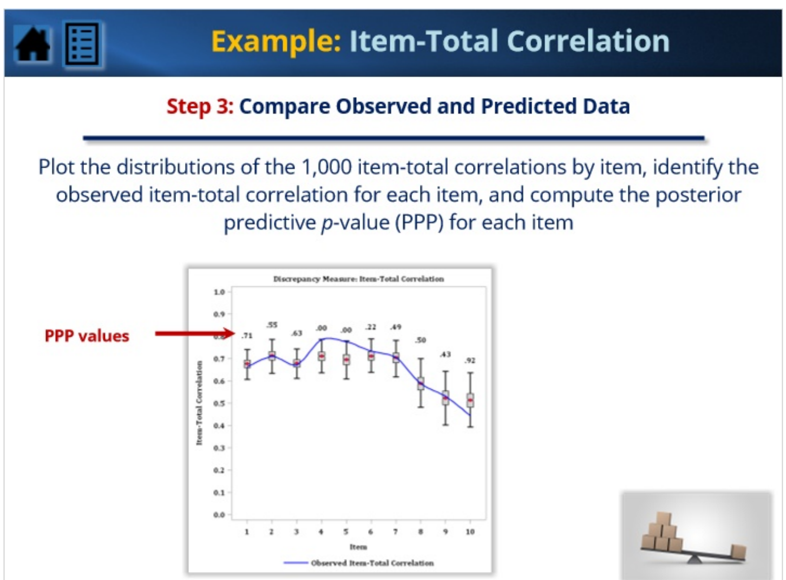
4.44 Item-total Correlations (I)



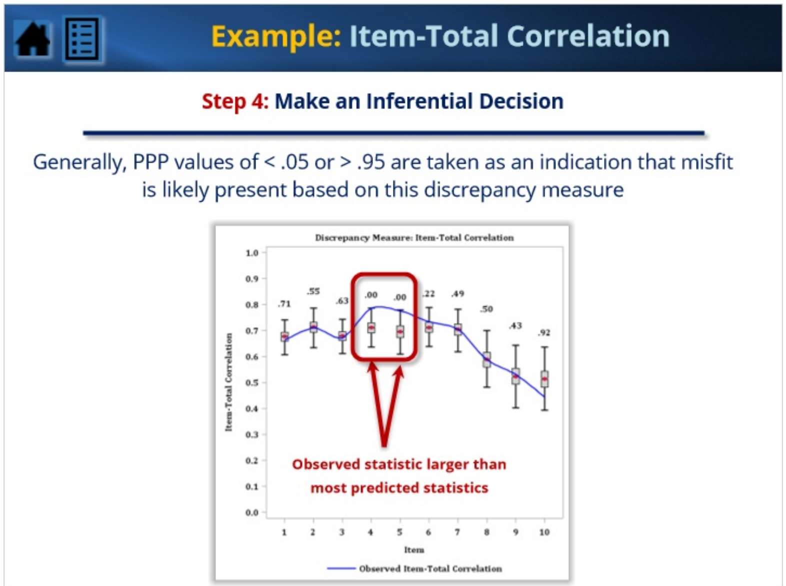
4.45 Item-total Correlations (II)



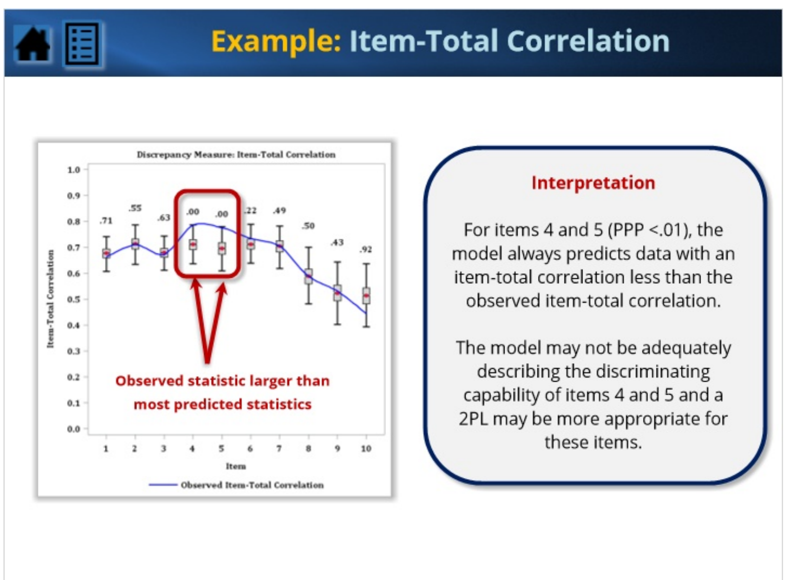
4.46 Item-total Correlations (III)



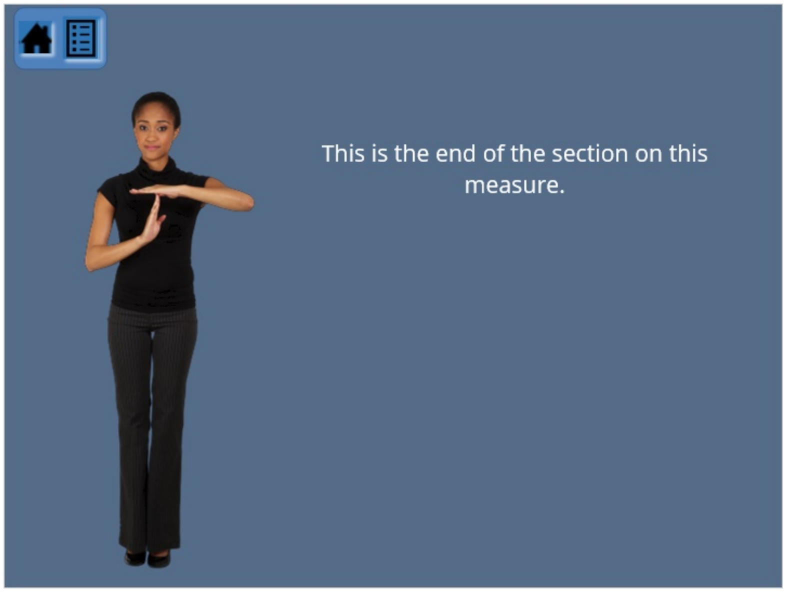
4.47 Item-total Correlations (IV)



4.48 Item-total Correlations (V)

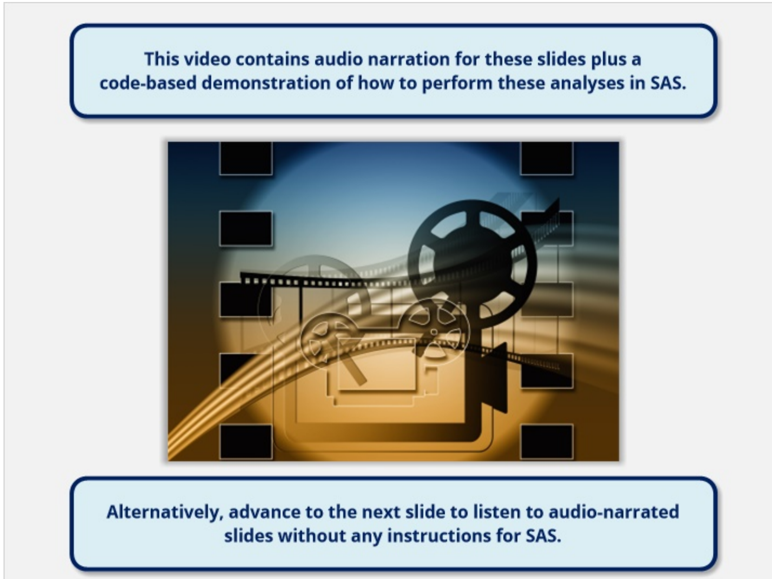


4.49 Bookend: Item-total Correlations

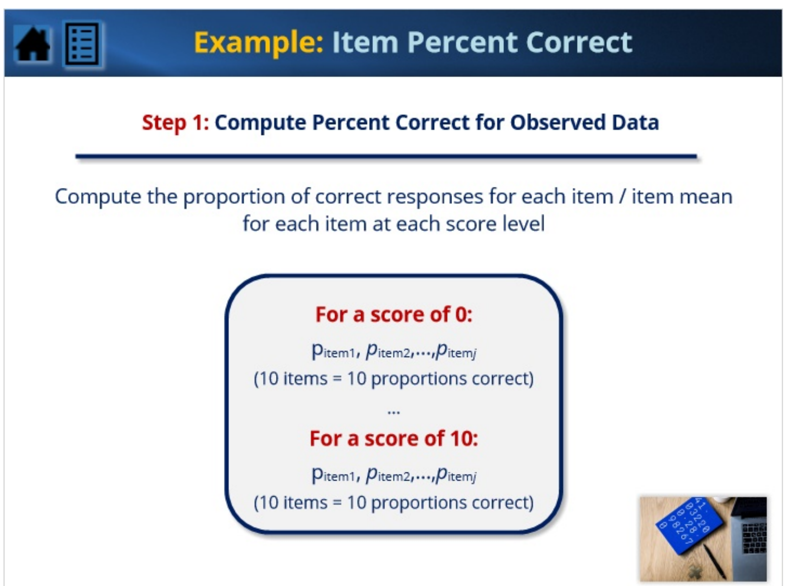


4.50 Bookmark: Percent Correct

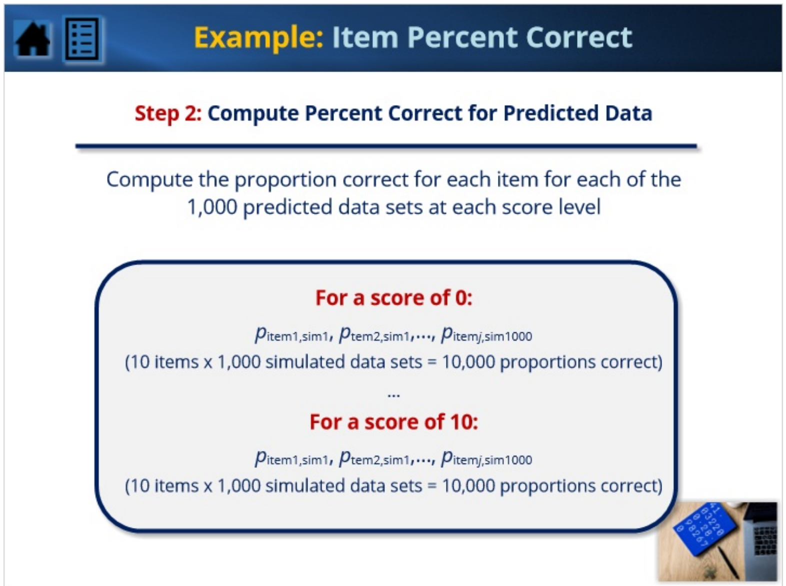
4.51 Video: Percent Correct



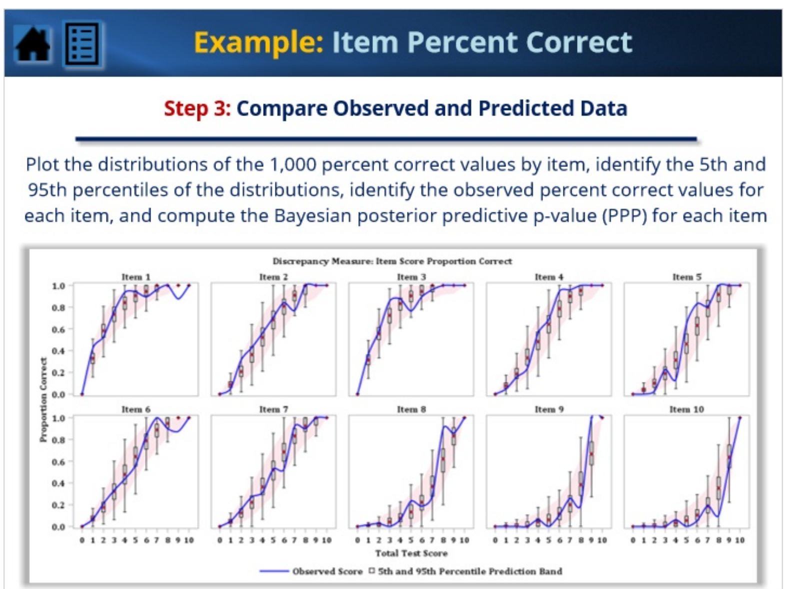
4.52 Percent Correct (I)



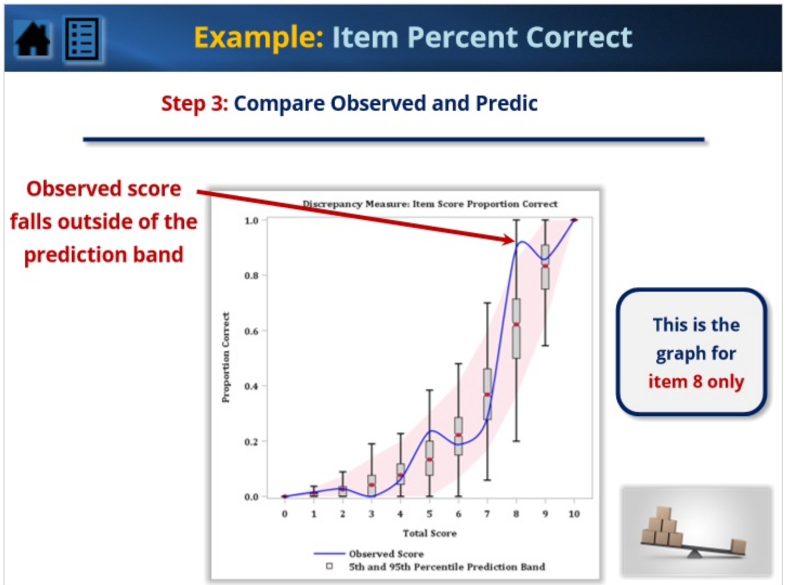
4.53 Percent Correct (II)



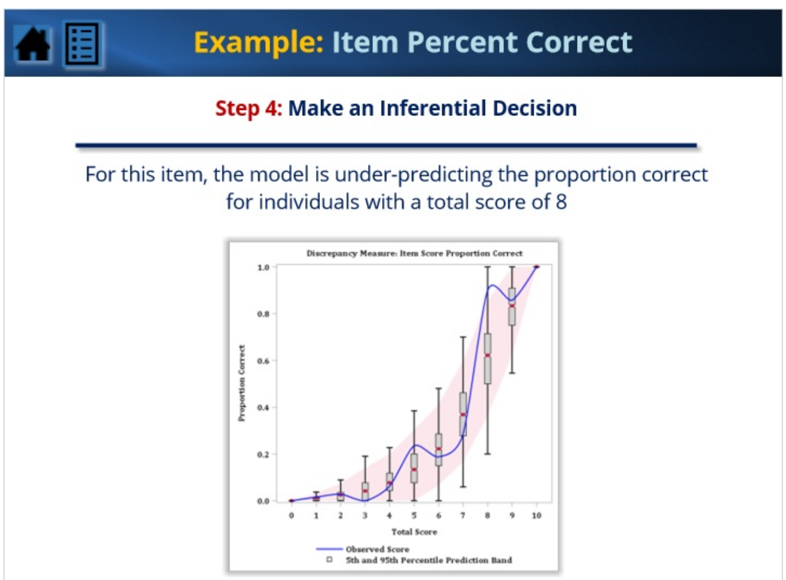
4.54 Percent Correct (III)



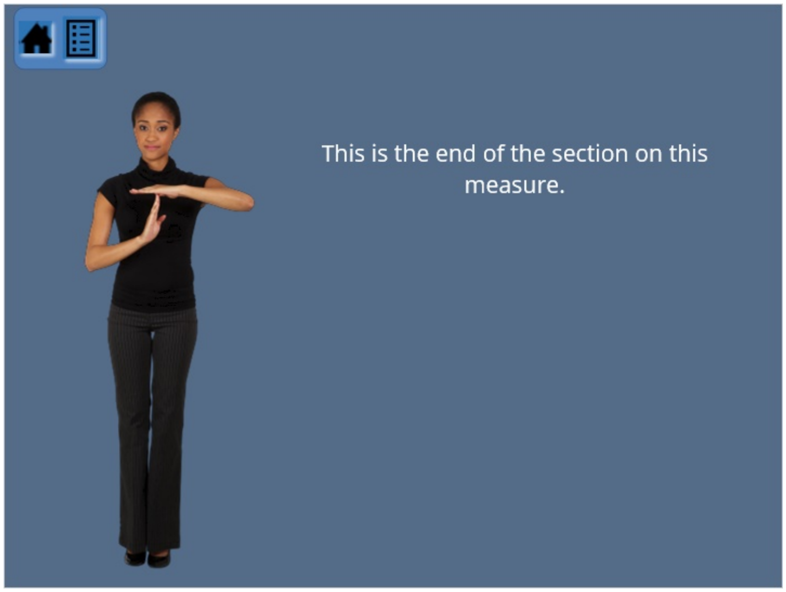
4.55 Percent Correct (IV)



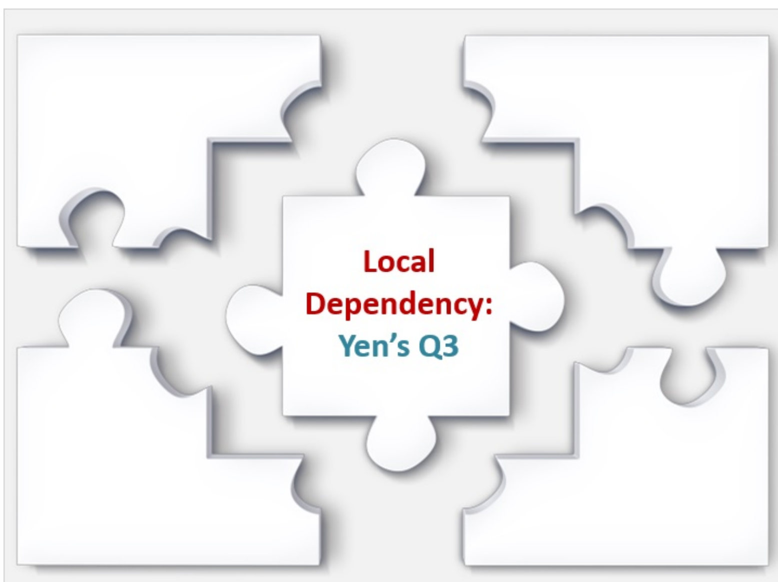
4.56 Percent Correct (V)



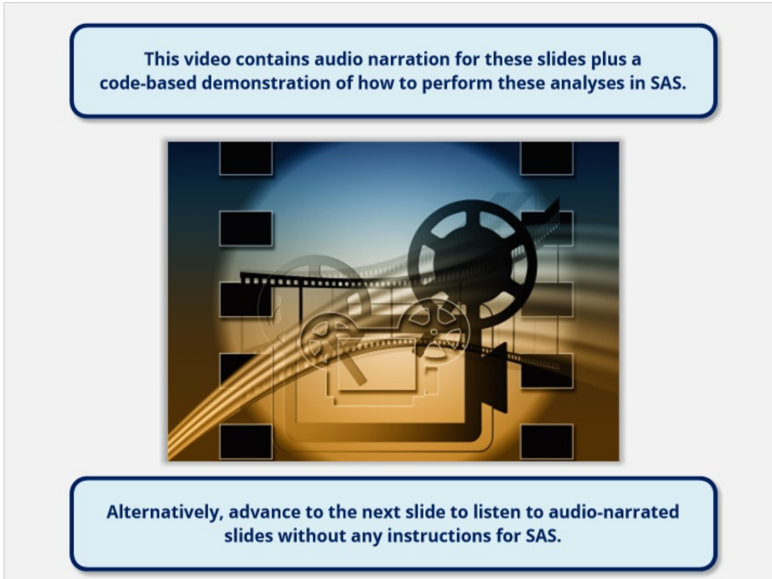
4.57 Bookend: Percent Correct



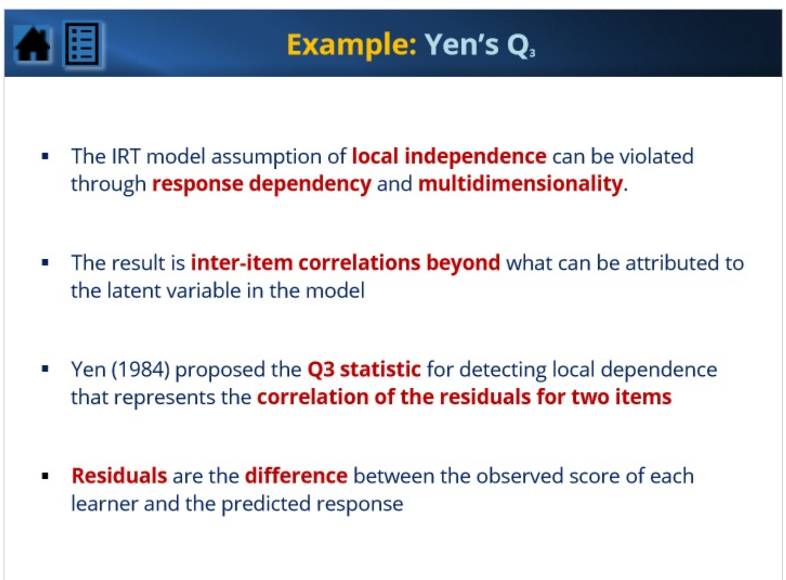
4.58 Bookmark: Yen's Q3



4.59 Video: Yen's Q3



4.60 Yen's Q3 (I)



4.61 Yen's Q3 (II)

Example: Yen's Q_3

Yen's Q3 is the correlation between two item residuals:

$$r_i = X_i - E(X_i|\theta)$$

$$r_i = Observed - Predicted$$

$$Q3_{ij} = corr(r_i, r_j)$$

4.62 Yen's Q3 (III)

Example: Yen's Q_3

Step 1: Compute Yen's Q3 for Observed Data

$Q3_{item12}, Q3_{item13}, \ldots, Q3_{item(j-1)j}$

There are as many Q3 statistics as there are item pairs:

$$\frac{\#Items \times (\#Items - 1)}{2} = \frac{10 \times (10 - 1)}{2} = 45$$

4.63 Yen's Q3 (IV)

Example: Yen's Q₃

Step 2: Compute Yen's Q3 for Predicted Data

Compute Yen's Q3 statistic for each of the 1,000 simulated data sets for each item pair

$Q3_{item12,sim1}, Q3_{tem13,sim1},\ldots,Q3_{tem(j\text{-}1,j),sim1},\ldots,Q3_{item(j\text{-}1,j),sim1000}$
(45 item pairs x 1,000 simulated data sets = 45,000 Q3 statistics)

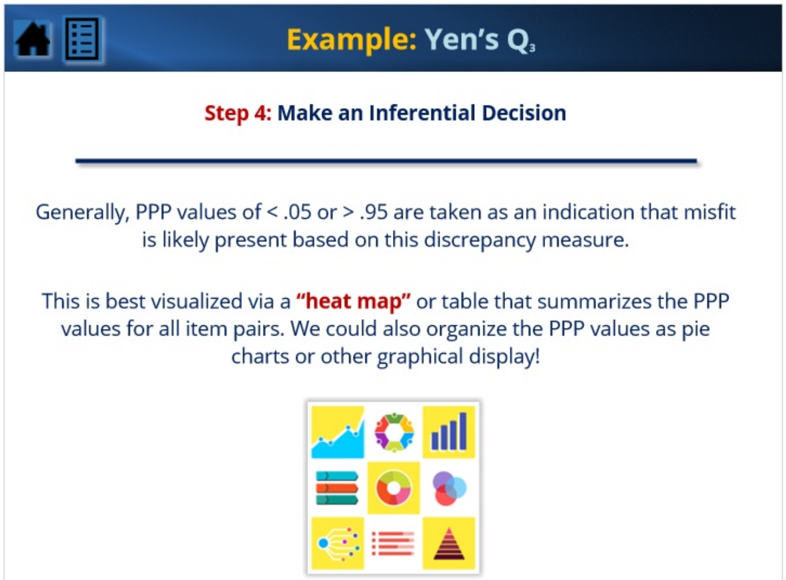
4.64 Yen's Q3 (V)

Example: Yen's Q₃

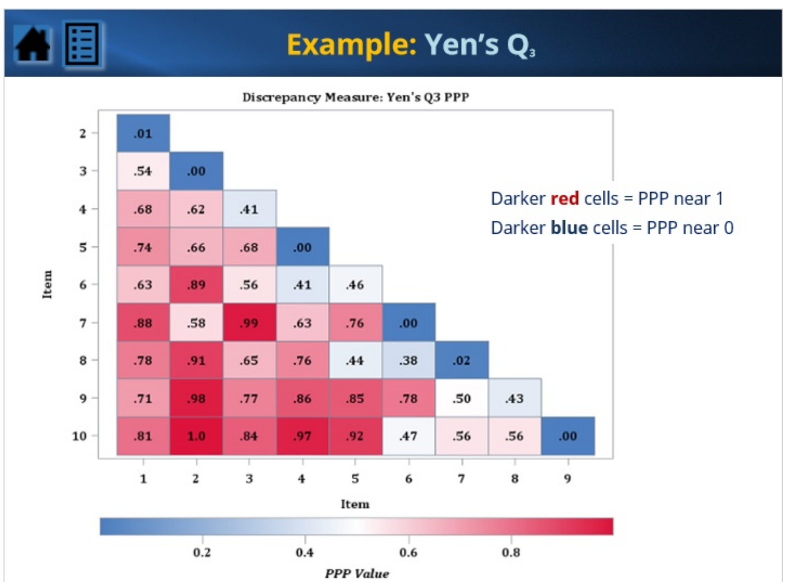
Step 3: Compare Observed and Predicted Data

I. Identify the 5th and 95th percentiles of the distribution of Q3 values for each item pair,
II. Identify the observed Q3 value within this distribution, and
III. Compute the Bayesian posterior predictive p-value (PPP) for each item pair

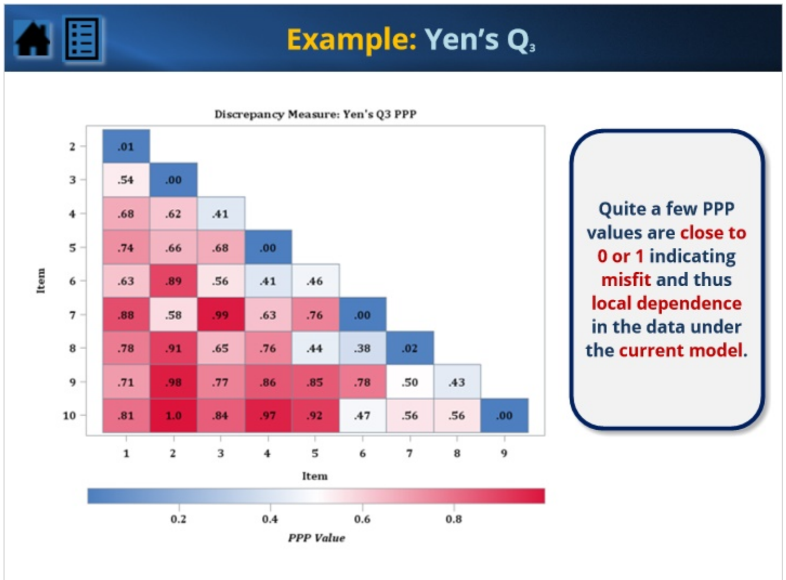
4.65 Yen's Q3 (VI)



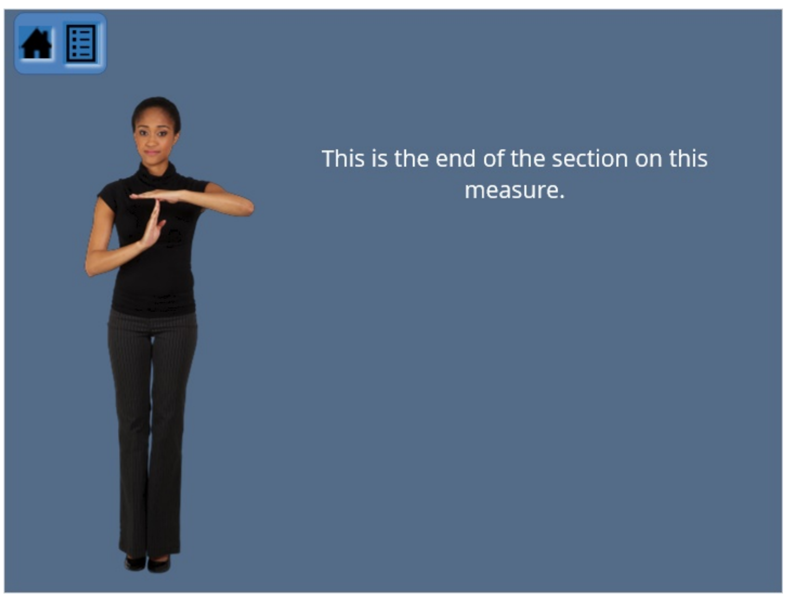
4.66 Yen's Q3 (VII)



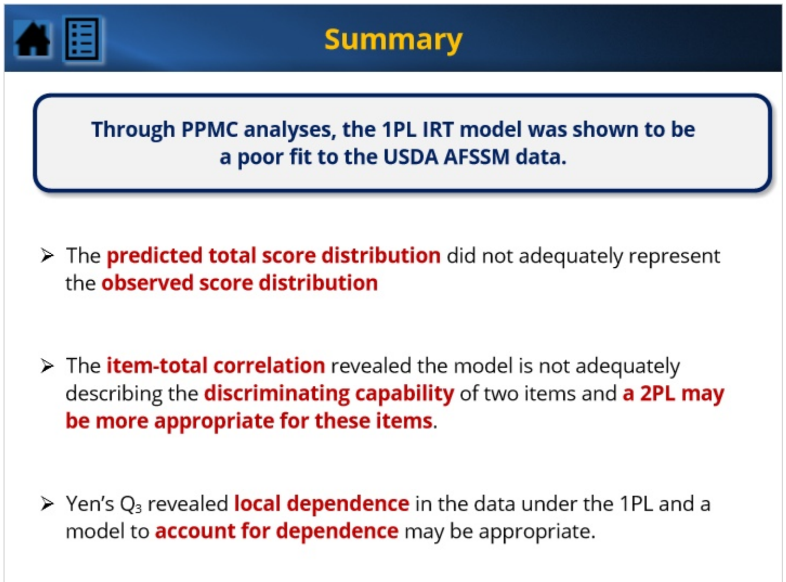
4.67 Yen's Q3 (VIII)



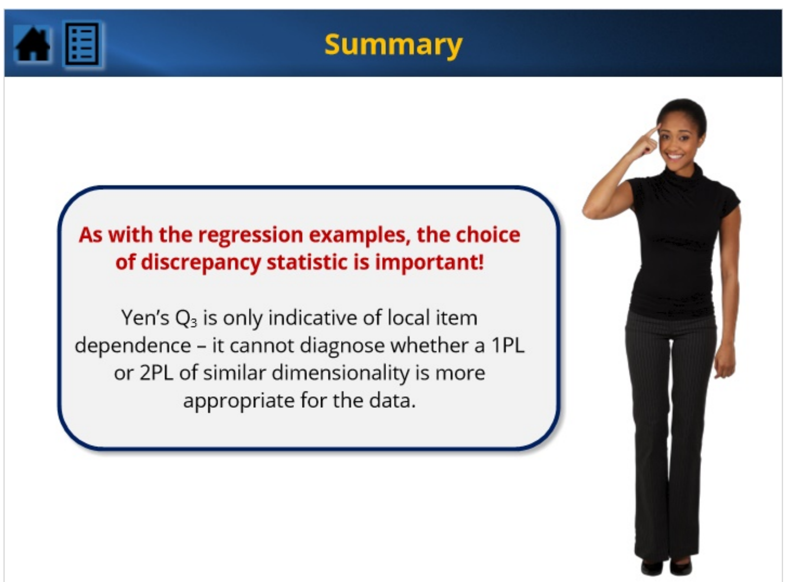
4.68 Bookend: Yen's Q3



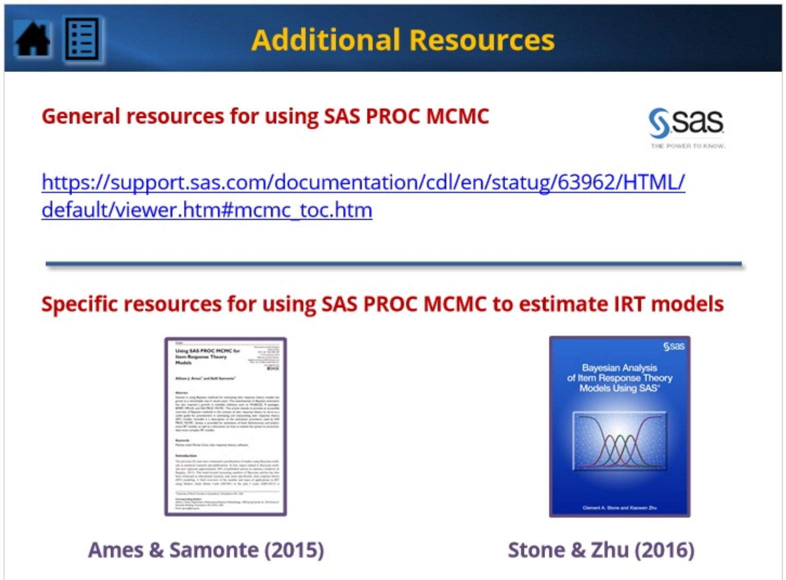
4.69 Summary (I)



4.70 Summary (II)



4.71 SAS Resources



4.72 Bookend: IRT