

## ITEMS Digital Module 06: Posterior Predictive Model Checking

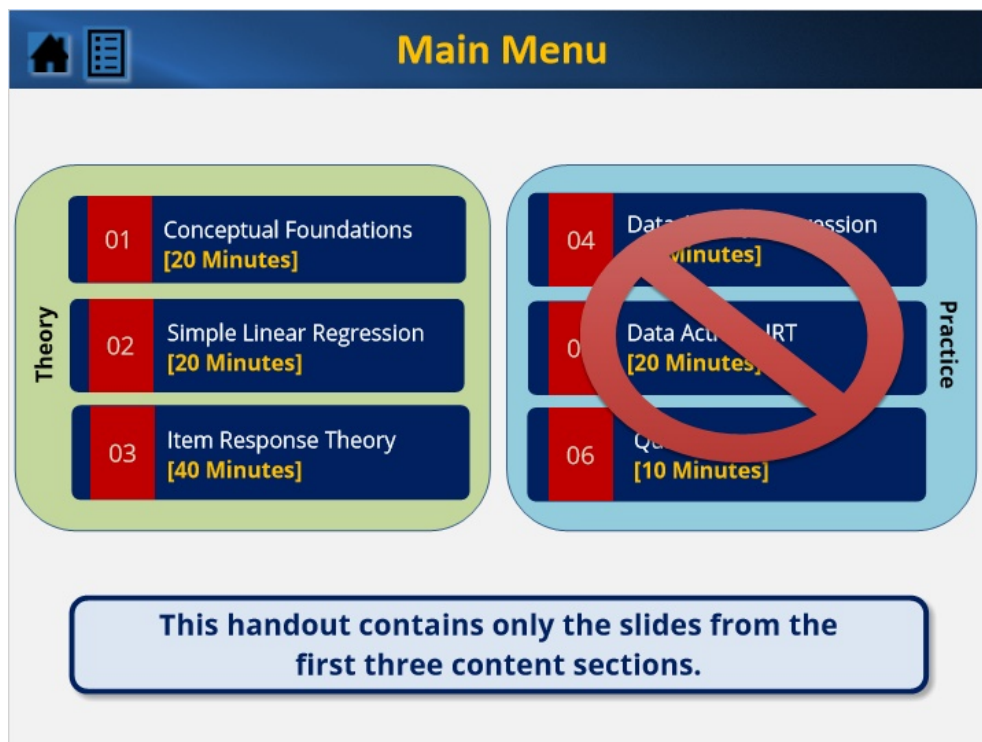
---

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


Meet the instructional design team:



André A. Rupp




Xi Lu




Click on the images to get to know them a little bit!

### 1.4 Support Team



...for supporting this project through staff time and other resources!

## 1.5 Welcome



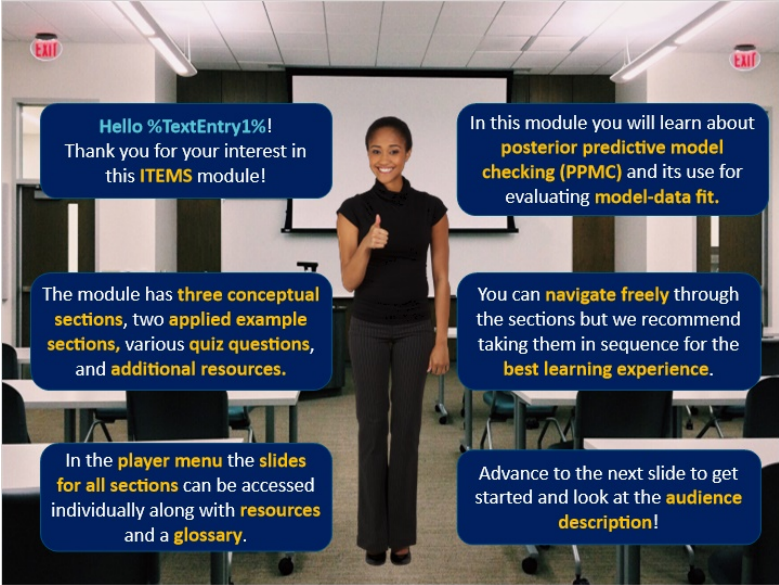
**Welcome to the ITEMS Module!**

The woman to the left is Laura!

Along with the content developers she will be guiding you through the module content.

Please type your name in the text box below:

## 1.6 Overview



Hello %TextEntry1%! Thank you for your interest in this ITEMS module!

In this module you will learn about **posterior predictive model checking (PPMC)** and its use for evaluating **model-data fit**.

The module has **three conceptual sections**, two **applied example sections**, various **quiz questions**, and **additional resources**.

You can **navigate freely** through the sections but we recommend taking them in sequence for the **best learning experience**.

In the **player menu** the **slides for all sections** can be accessed individually along with **resources** and a **glossary**.

Advance to the next slide to get started and look at the **audience description!**

## 1.7 Target Audience

### Target Audience

Anyone who would like a **gentle statistical introduction** to this topic:

- graduate students and faculty in Master's, Ph.D., or certificate programs
- psychometricians and other measurement professionals
- data scientists / analysts
- research assistants or research scientists
- technical project directors
- assessment developers



However, we hope that you find the information in this module **useful no matter what your official title or role** in an organization is!

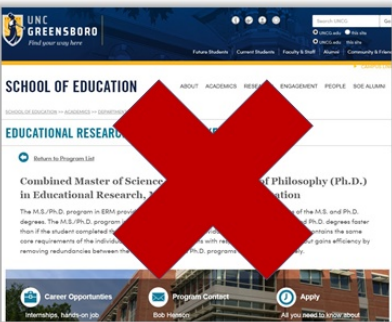
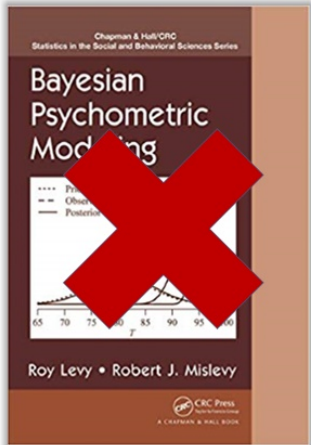
## 1.8 Expectations (I)



**Let's discuss expectations....**


## 1.9 Expectations (II)

**ITEMS Modules in Context**



## 1.10 Learning Objectives

**Learning Objectives**



1. Explain the scientific rationale behind posterior predictive model checking (PPMC)
2. Describe the general computational procedures for PPMC
3. Interpret graphical and statistical output to evaluate the linearity and constant variance assumptions of a simple linear regression model using PPMC
4. Interpret graphical and statistical output to evaluate key fit assumptions for a unidimensional IRT model using PPMC
5. Apply the PPMC approach for simple linear regression and unidimensional IRT using commonly available statistical software

## 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**.

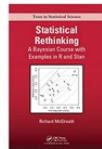
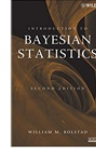
**Links to purchase key reference books on Amazon** are provided via the button below.

**Reference Books**

## Resources 1 (Slide Layer)

### Resources: Books

Bolstad, W. (2007). *Introduction to Bayesian statistics*.



McElreath, R. (2016). *Statistical rethinking: A Bayesian Course with R and Stan*.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*.



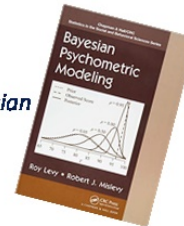
More  
Books

Back to  
Main Slide

## Resources 2 (Slide Layer)

### Resources: Books

Levy, R., & Mislevy, R. (2016). *Bayesian psychometric modeling*.



Jackman, S. (2009). *Bayesian analysis for the social sciences*.

Back to  
Other Books

Back to  
Main Slide

## 1.13 Main Menu



The image shows a 'Main Menu' interface. At the top, there is a dark blue header with a home icon, a list icon, and the text 'Main Menu'. Below the header, there are two columns of content. The left column is labeled 'Theory' and contains three items: '01 Conceptual Foundations [20 Minutes]', '02 Simple Linear Regression [20 Minutes]', and '03 Item Response Theory [40 Minutes]'. The right column is labeled 'Practice' and contains three items: '04 Data Acquisition [20 Minutes]', '05 Data Acquisition IRT [20 Minutes]', and '06 Question [10 Minutes]'. A large red prohibition sign is overlaid on the 'Practice' column. Below the columns, there is a light blue box with the text: 'This handout contains only the slides from the first three content sections.'

## 2. Conceptual Foundations


### 2.1 Cover: Foundations

The image shows a cover slide for 'Section 1: Conceptual Foundations'. On the left, there is a photograph of a classroom with desks, chairs, and a projector screen. On the right, there is a dark blue vertical overlay with the text: 'Section 1: Conceptual Foundations [20 Minutes]'.

## 2.2 Objectives: Foundations





### Learning Objectives



1. Compare and contrast Bayesian inference with maximum likelihood
2. List and describe key steps in the application of Bayesian inference
3. List and describe key steps of Posterior Predictive Model Checking (PPMC)
4. Describe the computational logic underlying PPMC procedures


## 2.3 Motivation



### Why Learn about PPMC?



- Bayesian estimation for **item response theory (IRT)** modeling and **other complex models** has been increasing rapidly
- **Posterior predictive model checking (PPMC)** is a prominent Bayesian approach to assessing model-data fit:

- Simple
- Flexible
- Sound
- Consistent
- Intuitive
- Powerful



(Sinharay, 2005)

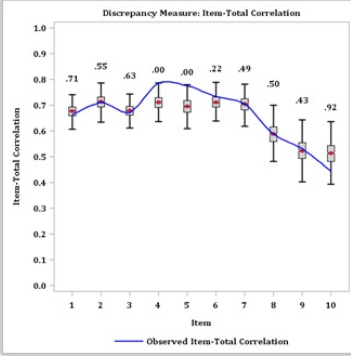
## 2.4 Sample Context



### Sample Context

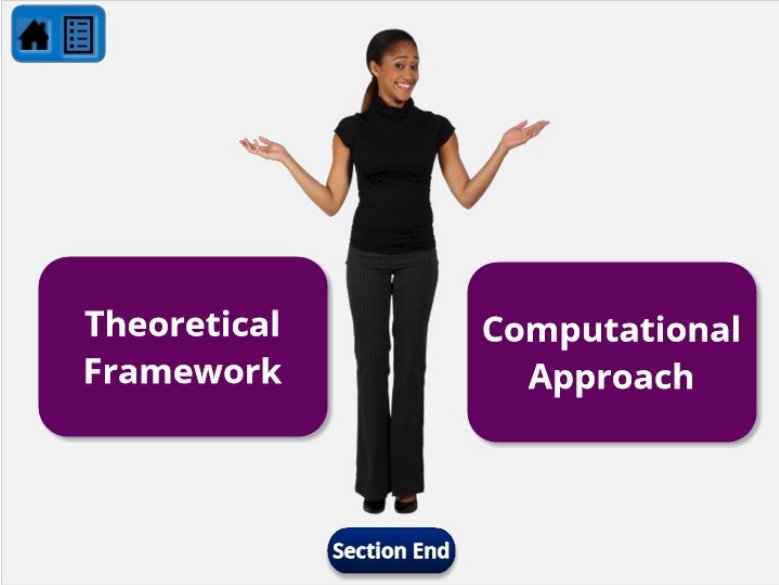


**In this section we will show you how to:**

- Apply the PPMC method to two common modeling approaches (SLR, IRT)
- Display the PPMC results using useful graphics
- Adapt the flexible PPMC method to your own research scenario



Item	Observed Item-Total Correlation
1	.71
2	.55
3	.63
4	.00
5	.00
6	.22
7	.49
8	.50
9	.43
10	.92

## 2.5 Topic Selection

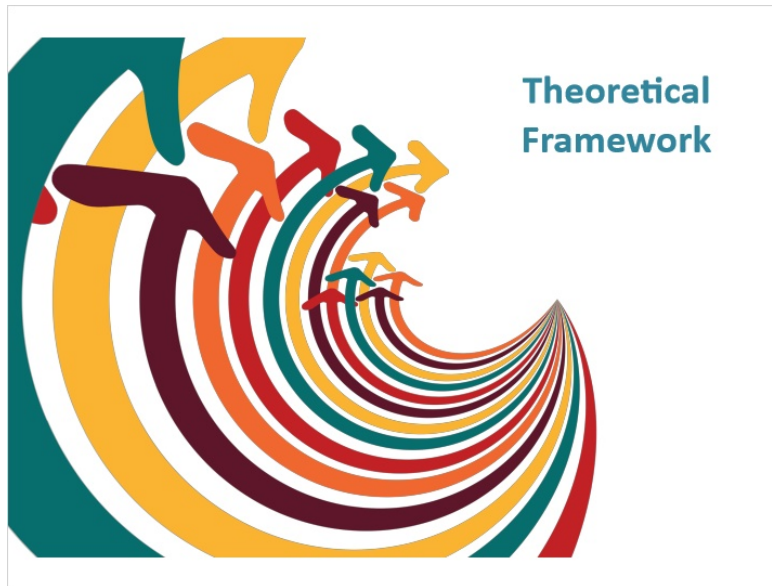


**Theoretical Framework**



**Computational Approach**

**Section End**


## 2.6 Bookmark: Theoretical Framework




## 2.7 Overview: Model Fit

  **Overview: Model Fit**



Statistical models are simplified approximations to the complexities underlying real data sets that help to tell empirical stories about learners and tasks



Drawing valid inferences from data through statistical models requires a good fit of the data to the model



## 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**.

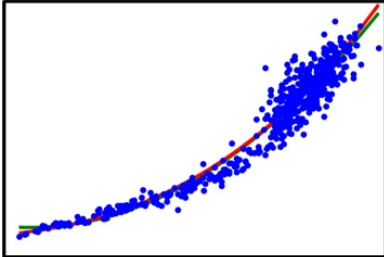
## 2.9 Consequences of Misfit (I)



### 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

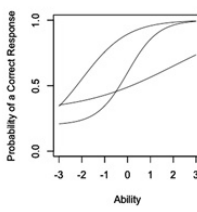


## 2.10 Consequences of Misfit (II)

**Overview: Consequences of Misfit**

**Consequences** of model-data misfit specific to **item response theory**:

- Biased person parameter estimates
- Biased item parameter estimates
- Incorrect ranking of individuals
- Incorrectly equated test scores
- Misinterpretations about latent traits or abilities



The graph plots the Probability of a Correct Response (y-axis, 0.0 to 1.0) against Ability (x-axis, -3 to 3). Three curves are shown: a standard sigmoidal curve, a curve that is shifted to the right, and a curve that is shifted to the left. The shifted curves represent biased item response functions.

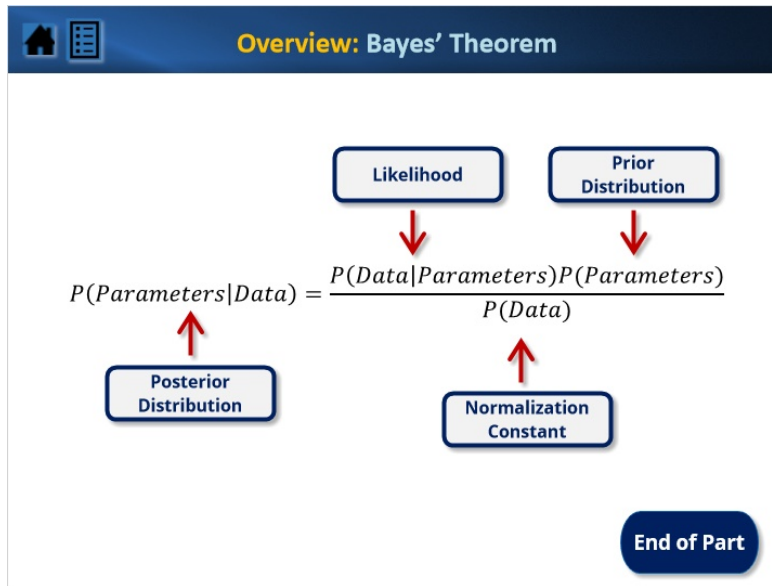
## 2.11 Overview: Inferential Frameworks

**Overview: Inferential Frameworks**

PPMC is a **broad analytic framework** for evaluating **model-data fit** from a **Bayesian perspective**

Maximum Likelihood Inference	Bayesian Inference
Likelihood	Likelihood
Data	Data
	Prior Distributions
	Posterior Distributions
Single most likely value of parameter ( <b>point estimate</b> )	Distribution of plausible parameter values ( <b>posterior</b> )



## 2.12 Overview: Bayes Theorem



## 2.13 Bookmark: Posterior





## 2.14 Posterior Distributions



### Posterior Distributions



Posterior distributions represent an analyst's belief about likely parameter values *after collecting data* and combining *data-drive evidence* with the a priori beliefs from the *prior distribution*

<b>Analytic Solutions</b>	<b>Markov Chain Monte Carlo (MCMC)</b>
<ul style="list-style-type: none"><li>• Via Bayes theorem</li><li>• Involve complicated mathematics</li><li>• Only available for certain models</li></ul>	<ul style="list-style-type: none"><li>• Simulation-based approach</li><li>• Involves "auditioning" values</li><li>• Creates collection of plausible values</li></ul>



[Wikipedia Page for MCMC](#)      [ITEMS Module on MCMC](#)      [Machine Learning Paper on MCMC](#)

## 2.15 Example Selection




**Example 1:**  
Weak Prior

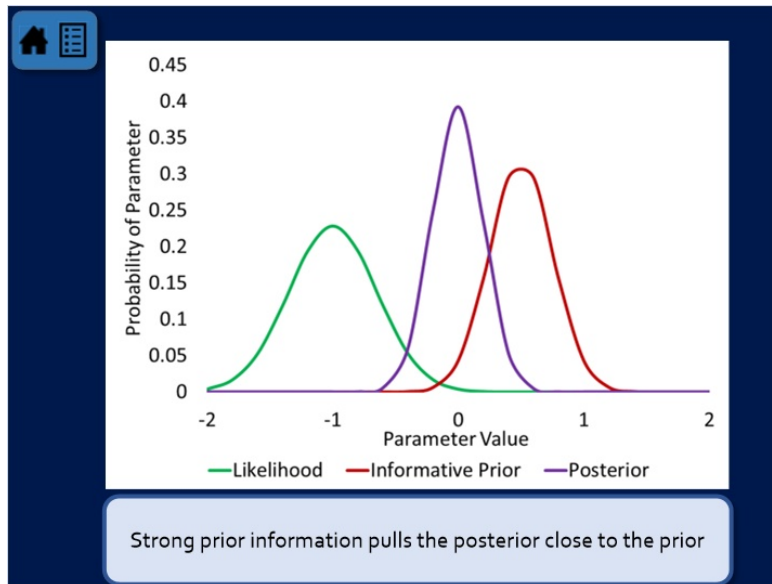
**Example 2:**  
Moderately Informative Prior

**Example 3:**  
Highly Informative Prior

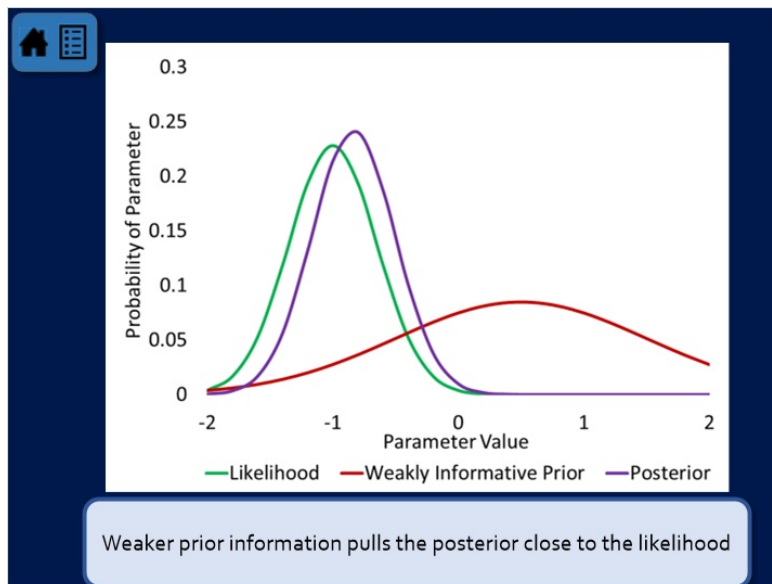
**Informational Value** →



## 2.16 Posterior Example (I)





## 2.17 Posterior Example (II)



## 2.18 Bookmark: Likelihood



## 2.19 Likelihood (I)



  **Likelihood**

Expression of the **joint probability of observing the data** under a given model using the **mathematical structure of the model**

Maximum Likelihood Inference	Bayesian Inference
Determines the single most-likely value of the model parameter(s) for the data	Creates a distribution of plausible values, one for each model parameter
Likelihood completely drives inference	Likelihood is combined with prior distribution to make posterior inference


Given **parameter values** for a model, how well do those values **explain / predict the observed data?**

## 2.20 Likelihood (II)



### Probability Density Functions



- **Probability density function (PDF):** defines the relationship between a set of scores and the probability of observing each score



Example of Normal Distribution


- **Likelihood:** joint probability over all data points for a particular PDF - how likely it is to observe the entire data set is given the PDF parameters

### Example (Slide Layer)



### Normal Distribution Example

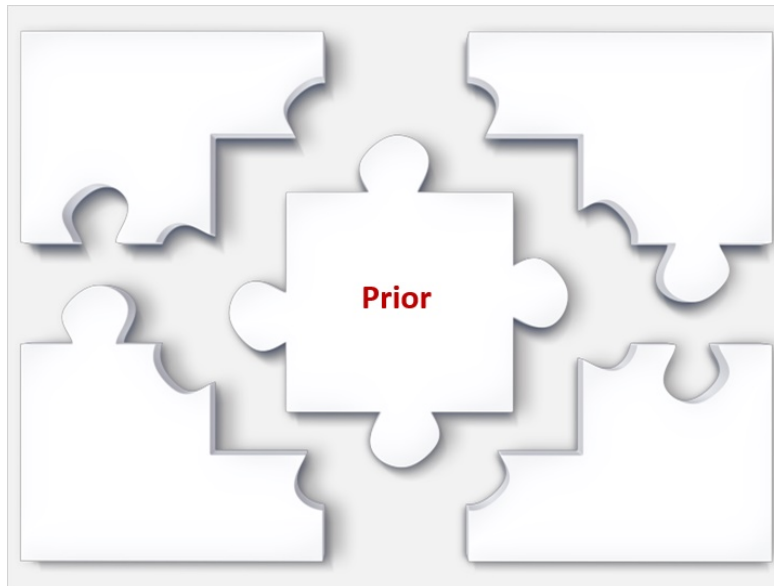
Substituting a value of  $x_i$  into the **normal distribution PDF** gives the **probability (likelihood) of observing this value** given a particular combination of **mean ( $\mu$ )** and **standard deviation ( $\sigma$ )** parameters.

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5 \boxed{\phantom{x_i - \mu}} \sigma^2}$$


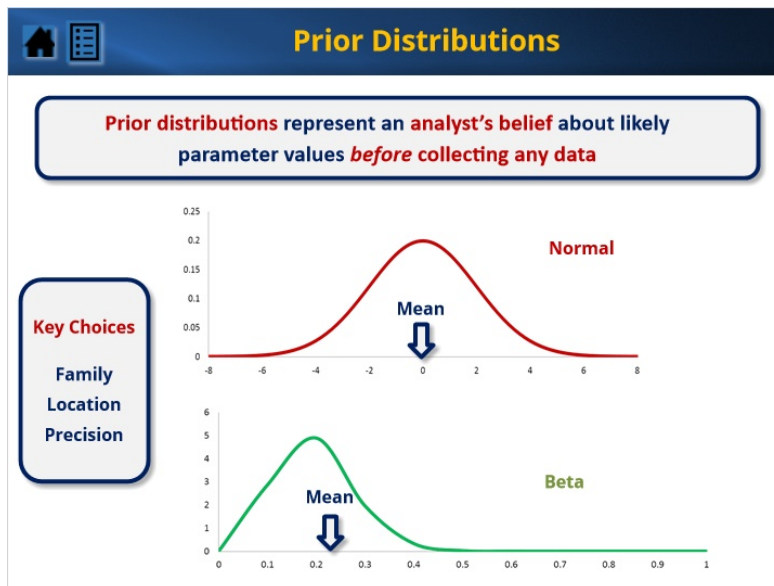
**Different parameter values** or **different distributions** would yield **different probabilities** for  $x_i$ .

Back to Main Slide

## 2.21 Bookmark: Prior



## 2.22 Prior Distributions (I)





## 2.23 Prior Distributions (II)

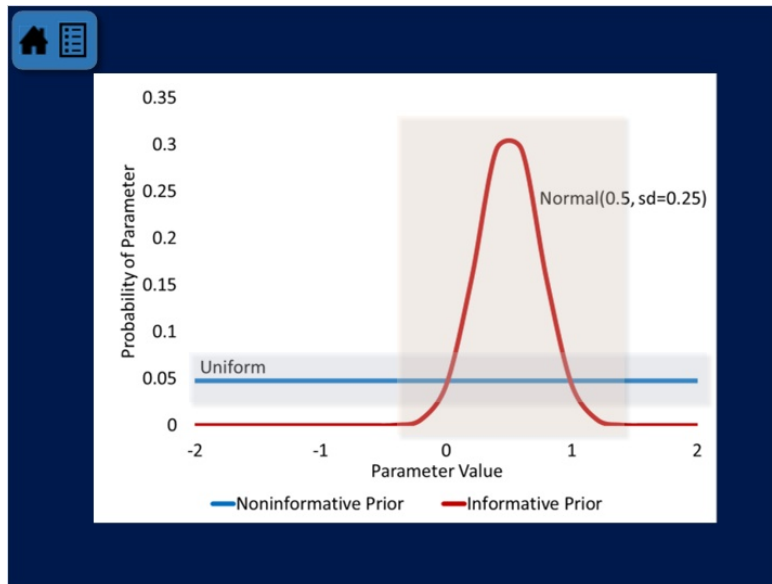
🏠 📄

### Prior Distributions

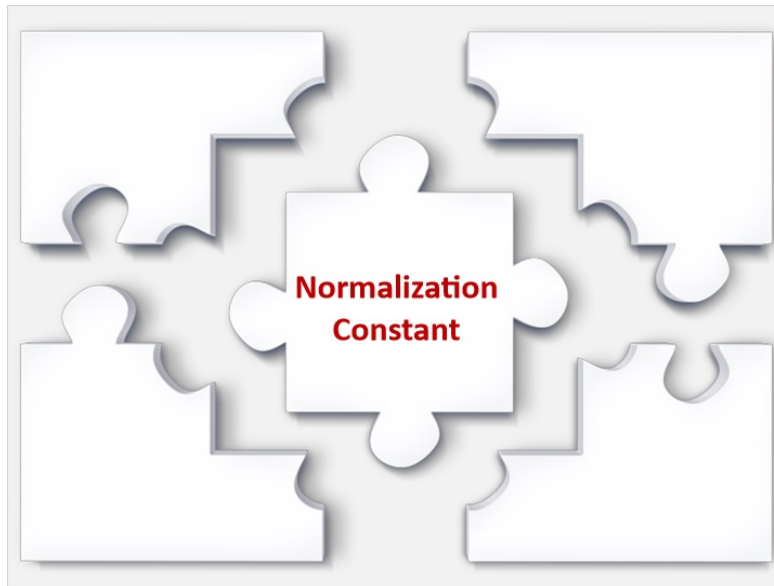
**Prior distributions represent an analyst's belief about parameter values *before* collecting any data**

<p><b>Noninformative (Imprecise)</b></p> <ul style="list-style-type: none"><li>• Little to no prior knowledge</li><li>• Each value of the parameter is equally probable in extreme case</li></ul> <div style="text-align: center;"></div>	<p><b>Informative (Precise)</b></p> <ul style="list-style-type: none"><li>• More specific beliefs about the likelihood of certain parameter values</li><li>• Different distributional shapes can be used to capture belief structure</li></ul> <div style="text-align: center;"></div>
--	--



## 2.24 Prior Distributions (III)



## 2.25 Bookmark: Normalization



## 2.26 Normalization Constant

  **Normalization Constant**

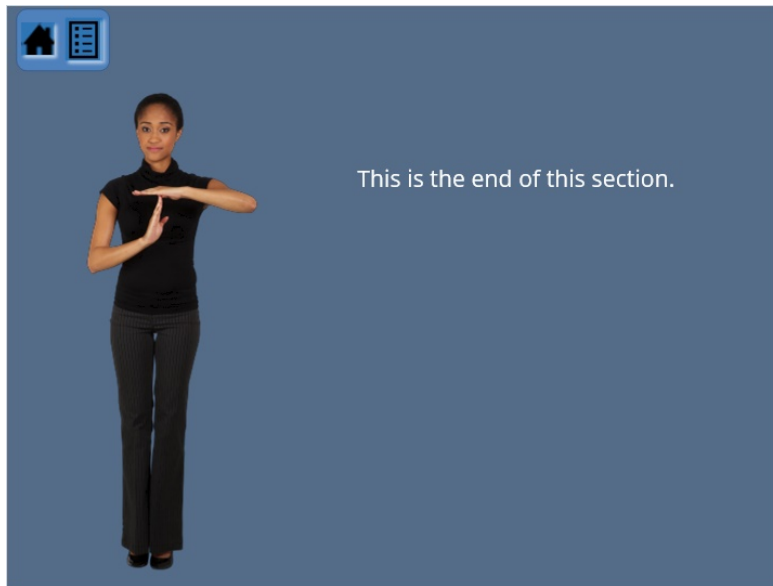
**Normalizing constant:** a scaling factor in the denominator or the posterior that is used to ensure the posterior distribution is a proper distribution (i.e., the area under the distribution must equal '1')

$$P(\text{Parameters}|\text{Data}) \propto P(\text{Data}|\text{Parameters})P(\text{Parameters})$$

**Normal Distribution Example:**

is the normalization constant for the posterior.



## 2.27 Bookend: Theoretical Framework



## 2.28 Bookmark: Computational Approach



## 2.29 Posterior Predictive Distributions

  **PPMC: Posterior Predictive Distributions (PPDs)**

**PPMC is used to check for model-data fit *before* making parameter inferences:**



- Highlight violation of model assumptions
- Illuminate features of observed data that the model fails to capture


---


**Posterior predictive distribution (PPD) are used for this purpose:**

- Conditional upon the model being evaluated
- Multiple predicted data sets are simulated to reflect the full range of plausible parameter values
- Observed model-fit statistics for data are compared to distributions of model-fit statistics from predicted data sets

## 2.30 Model-fit Logic

  **PPMC: Model-fit Logic**






If the model fits the observed data, then the PPD sets will resemble the observed data



**The PPD captures important features of the data**

If the model does not fit the observed data, then the PPD sets will not resemble the observed data

**The PPD does not capture important features of the data**





## 2.31 Evaluation Procedures

**PPMC: Evaluation Approaches**

- ✓ **Fit Statistics (Discrepancy Measures)**
  - Indices that highlight a feature of the data that is important to the analyst / researcher / practitioner
  - Selected or developed from model assumptions or substantive considerations / interpretations
  - Computed for the observed data (one value) and the PPD sets (as many values as there are simulated data sets)
- ✓ **Inferential Procedures (Posterior Predictive  $p$ -values [PPPs])**
  - Related to the number of simulated data sets that produce fit statistics that are different from the observed one
  - Large or small values are used to flag model-data misfit
- ✓ **Graphical Displays**

## 2.32 Discrepancy Measures

**PPMC: Discrepancy Measures**



**Linear Regression**

- **Linearity:** Scatterplots, descriptive statistics
- **Independence:** Autocorrelation statistics
- **Homoscedasticity:** Breusch-Pagan statistic

**Item Response Theory**

- **Monotonicity:** Conditional residual plots
- **Dimensionality:** Factor-analytic methods, DIMTEST, odds ratio
- **Complexity:** Information indices, LR test

## 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**

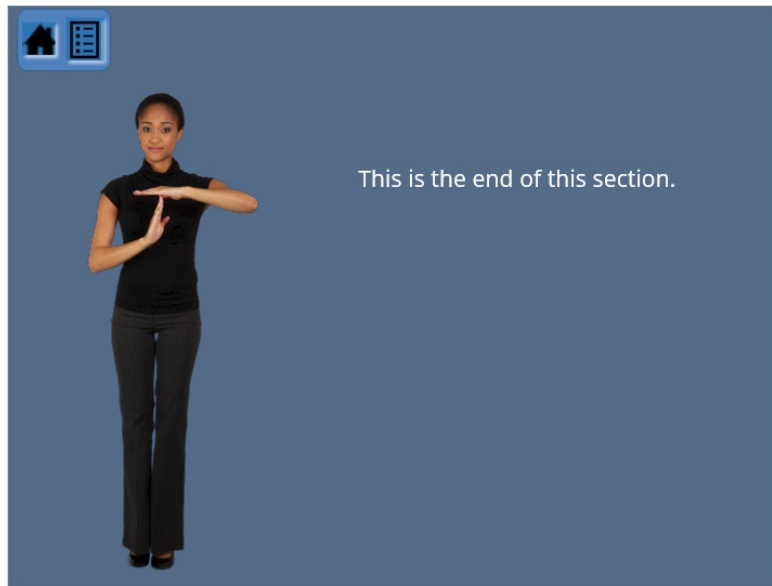
There are **two extended examples** in this digital module that you can access from the **main menu** or the **final slide in this section**:

- Simple linear regression
- Item response theory

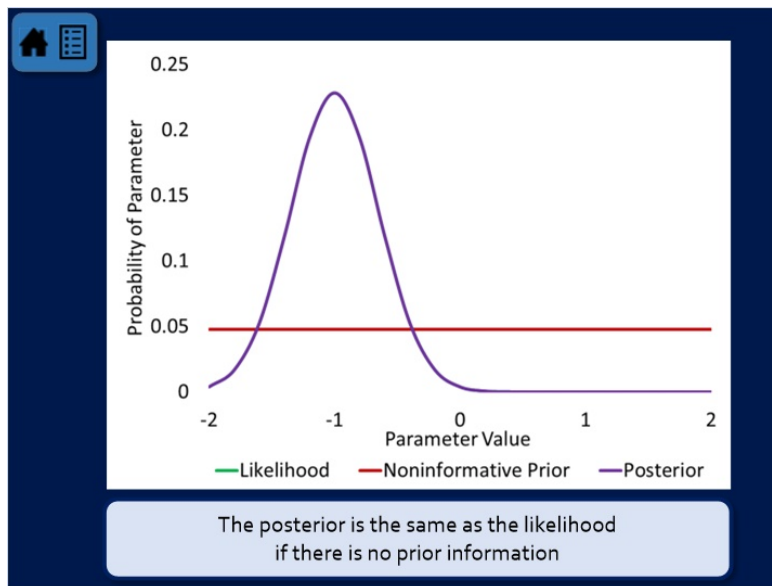


The general steps for **Bayesian inference and PPMC** are the same but the **statistical model** and associated **discrepancy measures** vary!

## 2.35 Bookend: Computational Approach



## 2.36 Posterior Example (III)




## 3. Simple Linear Regression

### 3.1 Cover: Regression




**Section 2:**  
**Simple Linear Regression**  
**[20 Minutes]**

### 3.2 Learning Objectives: Regression







**Learning Objectives**





1. Describe the basic assumptions of a simple linear regression model
2. Identify discrepancy measures that can be used to evaluate the basic model assumptions
3. Describe the basic computational steps for evaluating the basic model assumptions
4. Describe the rationale for making adjustments to the regression model based on the evaluations

### 3.3 Model-data Fit (III)

  **Recap: Model-data Fit**

<b>Adequate model-data fit:</b>	<b>Poor model-data fit:</b>
<ul style="list-style-type: none"><li>• Model-predicted data should be similar to most observed data points</li><li>• Difference between the observed data and predicted data should be small</li></ul> 	<ul style="list-style-type: none"><li>• Model-predicted data will be different from some observed data points</li><li>• Difference between the observed data and predicted data will be large</li></ul> 

### 3.4 Example: Overview

  **Statistical Model: Structure & Assumptions**


**Statistical Model:**  
Simple linear regression model of the form  $Y = a + bX$

**Model Parameters:**

- Intercept ( $a$ )
- Slope ( $b$ )

**Assumptions:**

- Linearity
- Homoscedasticity
- Independence



**Regression Formula**

## Regression Formula (Slide Layer)

### Regression Formula

#### Structure

$$Y_i = a + bX_i + E_i$$

$$i = 1, \dots, n$$

$n$  = number of observations

#### Assumptions

$$E_i \sim \text{Normal}(0, \sigma_E^2)$$

Error are mutually independent and identically normally distributed with a constant variance

Back to  
Main Slide

## 3.5 Regression Example (II)

### Statistical Model: Example Data

- Mean of the **Normal distribution**,  $\mu_i$ , is specified as a **linear function** of the predictor variable and parameters of interest:

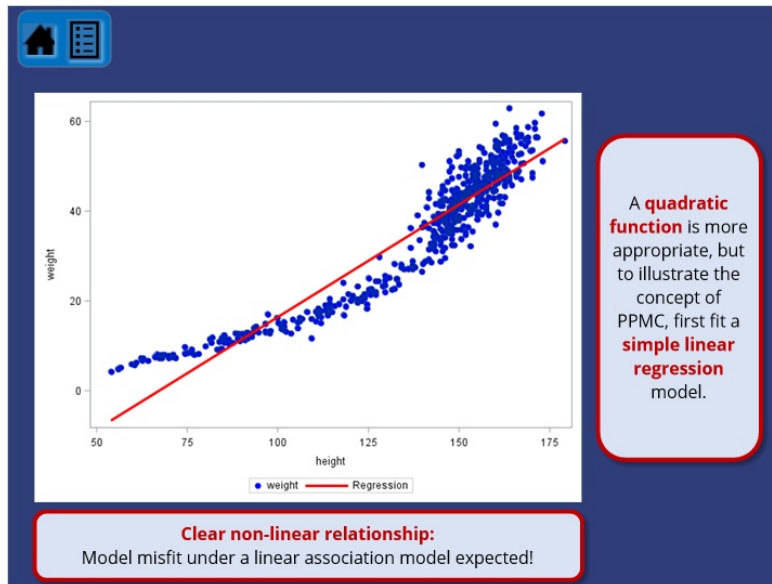
$$\text{weight}_i \sim \text{Normal}(\mu_i, \sigma^2)$$

$$\mu_i = a + b * \text{height}_i$$

$$\text{weight}_i \sim \text{Normal}(a + b * \text{height}_i, \sigma^2)$$

- $\text{weight}_i$  is the **dependent variable (outcome)** for the  $i^{\text{th}}$  individual ( $i = 1, 2, \dots, n$ )
- $\text{height}_i$  is the **independent variable (predictor)** for the  $i^{\text{th}}$  individual
- $a$  is the **intercept**
- $b$  is the **slope**
- $\sigma^2$  is the **variance** (spread) of the data

### 3.6 Regression Example (I)



### 3.7 Regression Example (III)

**Statistical Model: Prior Specification**

**Model**  $weight_i \sim Normal(a + b * height_i, \sigma^2)$

**Priors**

- $a \sim Normal(45, sd = 100)$
- $b \sim Uniform(0, 10)$
- $\sigma \sim Uniform(0, 50)$

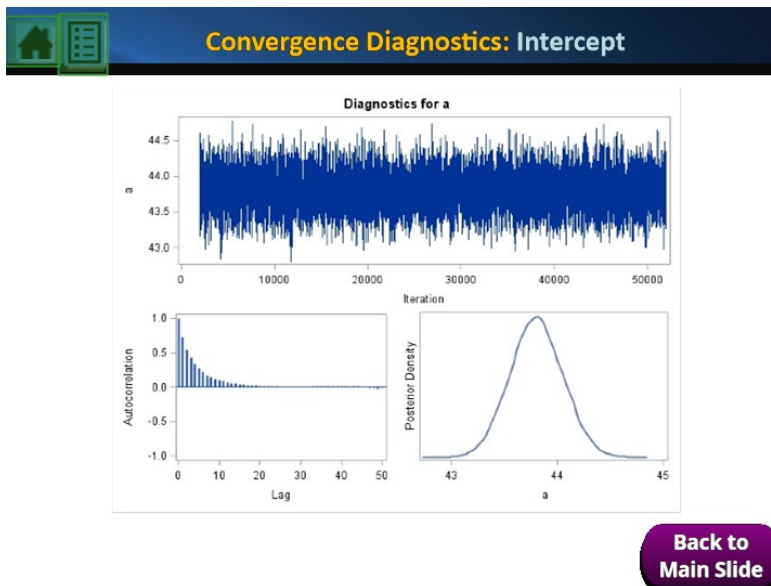
**Convergence** assessed via trace plots and formal statistical tests (i.e., Geweke's diagnostic)

### 3.8 Regression Example (IV)

**Statistical Model: Posterior Distributions**

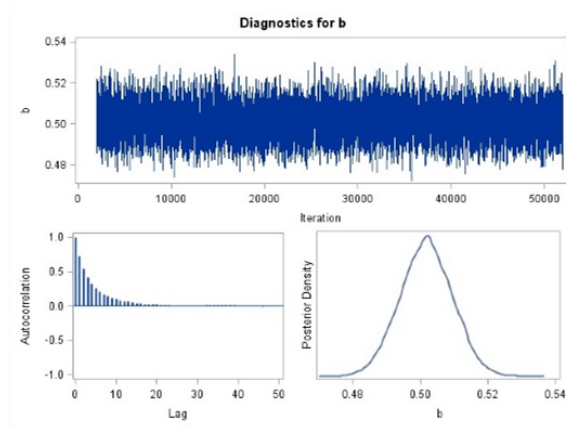
Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
a	50000	43.80	0.25	43.30	44.27
b	50000	0.50	0.01	0.49	0.52
s	50000	5.00	0.15	4.70	5.31

#### Diag a (Slide Layer)



## Diag b (Slide Layer)

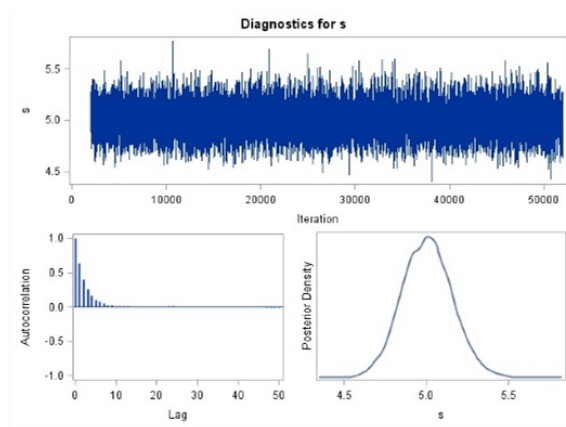
### Convergence Diagnostics: Slope



[Back to Main Slide](#)



## Diag s (Slide Layer)

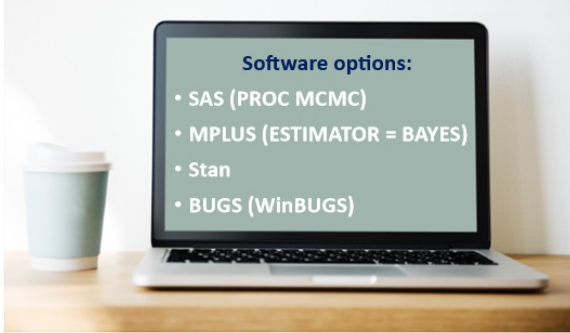
### Convergence Diagnostics: Standard Deviation



[Back to Main Slide](#)

### 3.9 PPMC: Computational Steps (II)



  **PPMC: Note on Software**




- SAS (PROC MCMC)
- MPLUS (ESTIMATOR = BAYES)
- Stan
- BUGS (WinBUGS)

- PPMC is **not built in** to many of these programs
- Knowledge of the **programming language** is required
- We have provided **sample SAS code** for your use

### 3.10 Topic Selection





**Linearity** **Homoscedasticity**

**Section End**


### 3.11 Bookmark: Linearity



### 3.12 Untitled Slide

  **Example: Overview**



**Violations of the linearity assumption are serious:**  
Your predictions will have considerable error!



**Models should capture the entire range of the data**

- Misfit due to non-linearity may occur in specific regions, such as the high or low regions of the distribution
- The predicted mean, minimum, and maximum will help determine whether the model's linearity assumption is adequate for the entire range of the data

### 3.13 Example: Overview




#### Example: Overview

**Analytic Goal:**  
Determine whether the model assumption of linearity is being met  
-> **linear association between X and Y**



**Discrepancy Measures:**  
Mean, standard deviation, minimum, maximum

**Description:**  
Descriptive summary measures with different properties can help evaluate whether the model is a good fit for the entire range of the data



**Regression Formula**

### Regression Formula (Slide Layer)



#### Regression Formula

**Structure**

$$Y_i = a + bX_i + E_i$$

$i = 1, \dots, n$   
 $n = \text{number of observations}$

**Assumptions**



$$E_i \sim \text{Normal}(0, \sigma_E^2)$$

Error are mutually independent and identically normally distributed with a constant variance

**Wikipedia Page**

**Back to Main Slide**

### 3.14 PPMC: Computational Steps (I)





#### PPMC: Computational Steps

In simple terms, after estimating the posterior distributions:

- compute the discrepancy measure for the **single observed data set**
- compute the discrepancy measure for **each predicted data set**
- compare the observed value and the **distribution of predicted values**
- make a decision about model-data fit **based on the selected measure**

The next slide shows the computational steps in more detail.

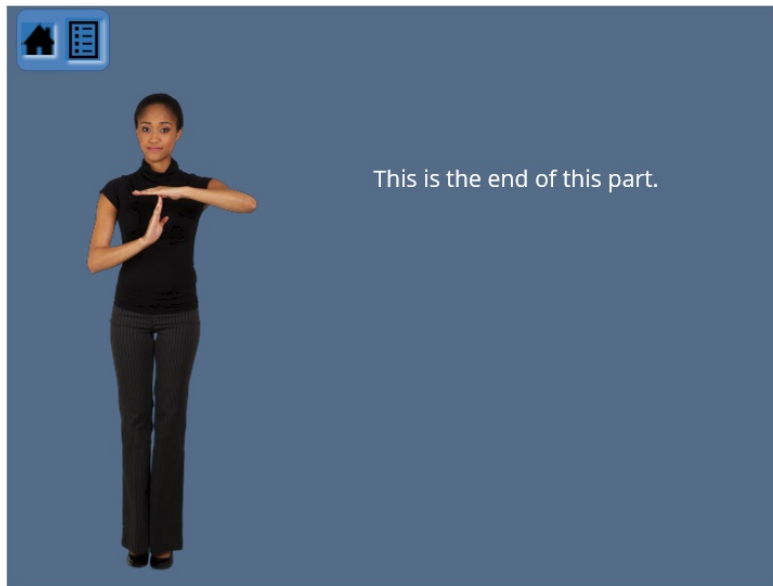
### 3.15 PPMC: Computational Steps (II)



#### 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

### 3.16 Bookend: Linearity



### 3.17 Bookmark: Homoscedasticity



### 3.18 Example: Overview


**Motivating Example: Overview**

**Analytic Goal:**  
Determine whether the model assumption of homoscedasticity (equal error variance) is being met

**Discrepancy Measure:**  
Breusch-Pagan statistic

**Description:**

- The Breusch-Pagan test regresses the squared residuals on the independent variables.
- With homoscedasticity this regression should NOT explain the variation in the squared residuals.



**Breusch-Pagan Formula**

**Regression Formula**

### Regression Formula (Slide Layer)

**Regression Formula**

**Structure**

$$Y_i = a + bX_i + E_i$$

$i = 1, \dots, n$   
 $n = \text{number of observations}$

**Assumptions**



$$E_i \sim \text{Normal}(0, \sigma_E^2)$$

Error are mutually independent and identically normally distributed with a constant variance

**Wikipedia Page**

**Back to Main Slide**

## Breusch-Pagan Formula (Slide Layer)



### Breusch-Pagan Statistic

**Regression Model**  $weight_i = a + b * height_i + e_i$

**Error Regression**  $e_i^2 = a_{BP} + b_{BP} * height_i + e_{iBP}$

**Breusch-Pagan Statistic**  $n * R_e^2$



$R_e^2$  is the **coefficient of determination** (squared multiple correlation) from the regression of the **independent variables** on the **squared residuals**

**Small Value**  $\Rightarrow$  variances are likely to be all equal (**homoscedasticity**)

**Large Value**  $\Rightarrow$  variances are likely to be NOT all equal (**heteroscedasticity**)

[Wikipedia Page](#) [Back to Main Slide](#)

### 3.19 PPMC: Computational Steps (I)





### PPMC: Computational Steps

In simple terms, after estimating the posterior distributions:

- compute the discrepancy measure for the **single observed data set**
- compute the discrepancy measure for **each predicted data set**
- compare the observed value and the **distribution of predicted values**
- make a decision about model-data fit **based on the selected measure**

The next slide shows the computational steps in more detail.

### 3.20 PPMC: Computational Steps (II)


  **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

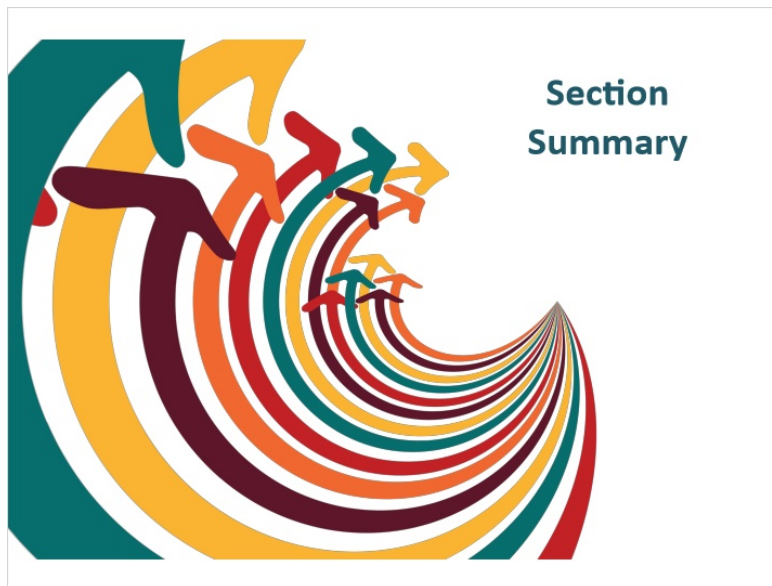
### 3.21 Bookend: Homoscedasticity





This is the end of this part.

### 3.22 Bookmark: Summary



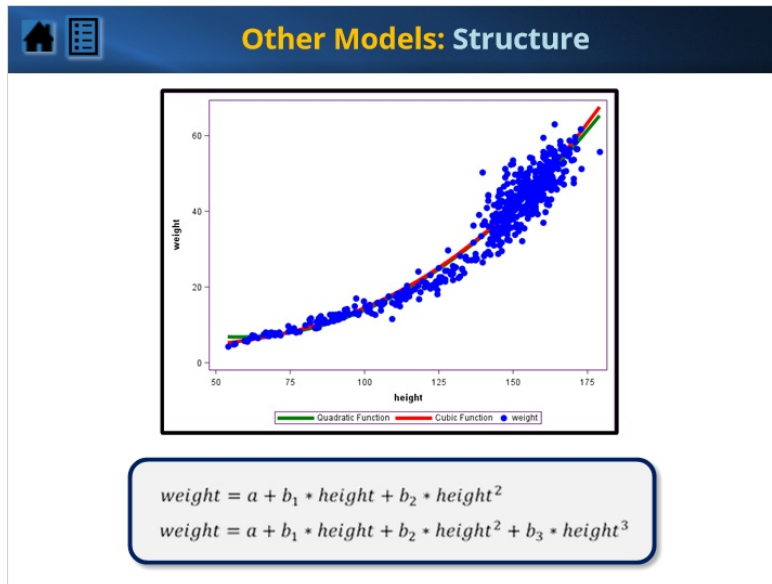
### 3.23 Summary

  **Summary**

Using the linear regression model for our data,  
PPMC indicated:

- **model-data misfit at the lower end of the distribution**  
(ppp < .05 for minimum)
- **a violation of the homoscedasticity assumption**  
(ppp < .05 for Breusch-Pagan statistic)

### 3.24 Other Models: Structure





### 3.25 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

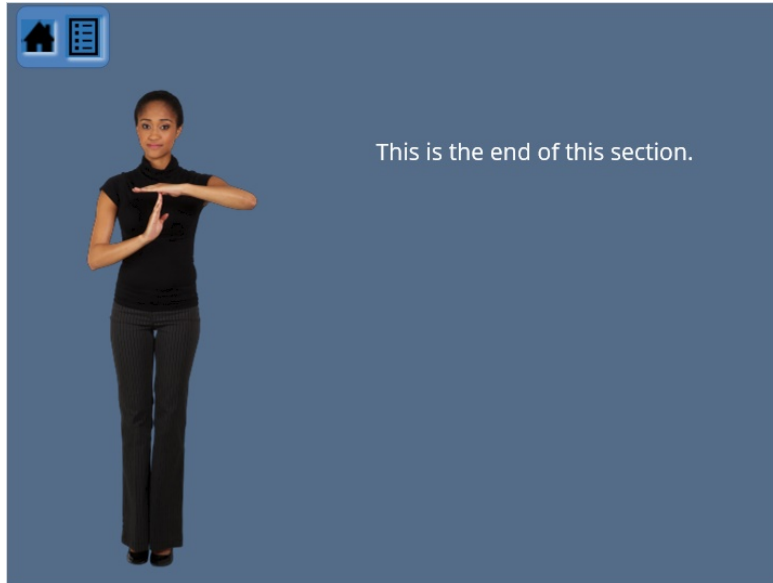
##### Choosing the appropriate statistic matters!

- 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.**

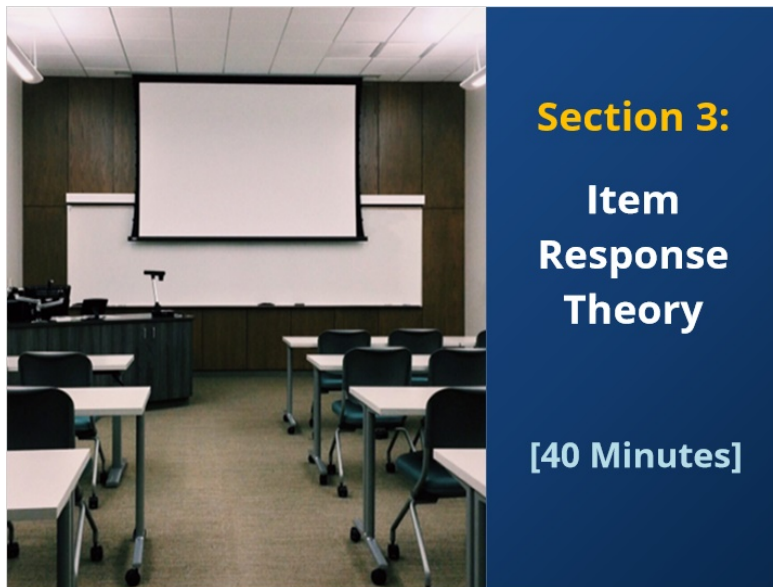


### 3.28 Bookend: Regression





## 4. Item Response Theory


### 4.1 Cover: IRT



## 4.2 Objectives: IRT





### Learning Objectives





1. Describe the basic assumptions of a unidimensional IRT model
2. Identify discrepancy measures that help evaluate IRT model assumptions
3. Describe the basic computational steps to evaluate the assumptions
4. Interpret visual summaries of the discrepancy measures to make decisions about model fit

## 4.3 IRT Models (I)



### Item Response Theory

- A **family of statistical models** commonly used in **large-scale testing**
- **Parameters** for respondents and items are placed on **common scales**
- **Response probabilities** depend upon a respondent's **latent trait** and the **item characteristics** via **model parameters**
- Respondents with a **higher value** on the latent trait will have a **higher probability** of responding correctly to items and **vice versa**
- One **popular method** of **estimating** the item and respondent parameters is **Bayesian inference**; the other is **maximum likelihood**



Wikipedia Refresher

## 4.4 IRT Models (II)

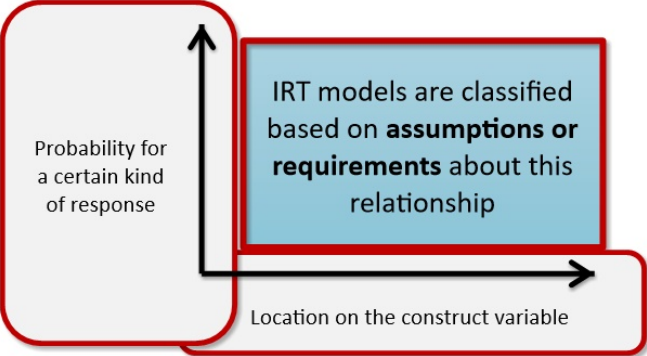
**Item-Response Function (IRF)**

Probability for a certain kind of response

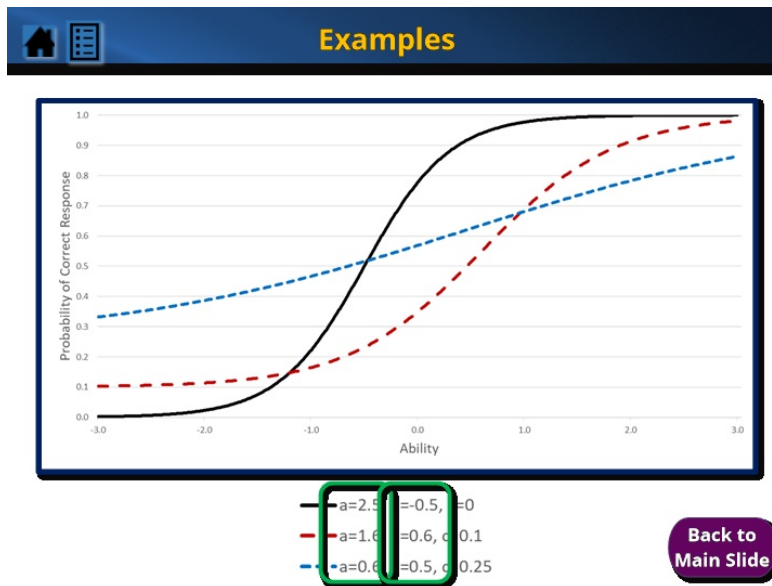
IRT models are classified based on **assumptions or requirements** about this relationship

Location on the construct variable

Examples



### Examples (Slide Layer)




## 4.5 Parametric Item Response Theory

### Model Types

- The **functional relationship** between the probability for a response and the latent trait is the **item response function (IRF)**
- Two common choices for the IRF are the **logistic** and the **probit** function
- Depending on the model that is chosen, different **item parameters** are available to influence the **shape of the function** within and across items:

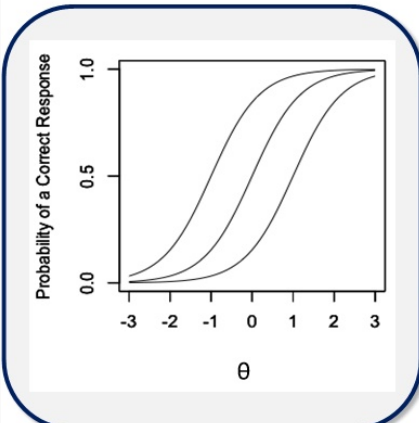
- ✓ **One-parameter / Rasch model:** Difficulty
- ✓ **Two-parameter model:** Difficulty, Discrimination
- ✓ **Three-parameter model:** Difficulty, Discrimination, Guessing

One-parameter / Rasch Model      Two-parameter Model      Three-parameter Model



## 4.6 One-parameter Model

### One-parameter Logistic (1PL) / Rasch Model



$$P_{ij} = \frac{\exp(a(\theta_i - b_j))}{1 + \exp(a(\theta_i - b_j))}$$

Difficulty Parameter

Latent Variable

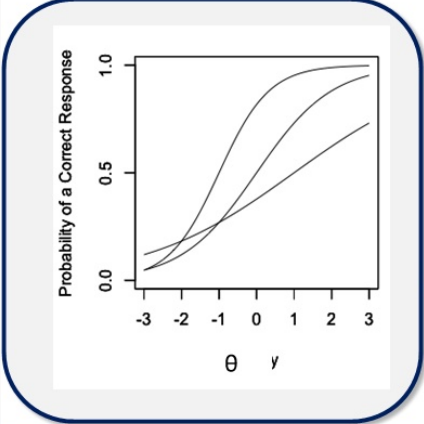
Rasch model:  $a = 1$

One-parameter / Rasch Model      Two-parameter Model      Three-parameter Model



## 4.7 Two-parameter Model

**Two-parameter Logistic (2PL) Model**



$$P_{ij} = \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}$$

Difficulty Parameter

Discrimination Parameter

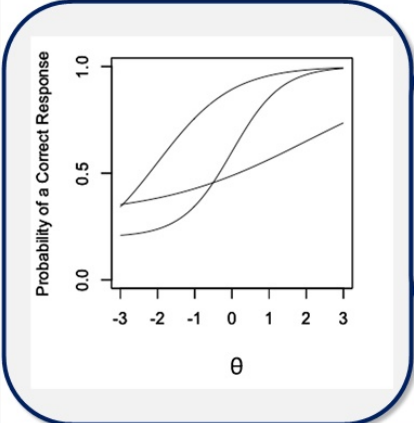
Latent Variable

One-parameter / Rasch Model      Two-parameter Model      Three-parameter Model

WIKIPEDIA

## 4.8 Three-parameter Model

**Three-parameter Logistic (3PL) Model**



$$P_{ij} = c_j + (1 - c_j) \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}$$

Pseudo-guessing Parameter

Difficulty Parameter

Discrimination Parameter

Latent Variable

One-parameter / Rasch Model      Two-parameter Model      Three-parameter Model

WIKIPEDIA

## 4.9 Prior Distributions (III)

### Prior Distributions

Parameter	Meaning	Range	Distribution
a	Discrimination	positive values	Lognormal
b	Difficulty	any value	Normal
c	Guessing	between 0 and 1	Beta
$\theta$	Ability / Trait	any value	Normal

Prior distributions are selected to match the possible value range of the parameters

Difficulty      Discrimination      Guessing

### Prior Difficulty (Slide Layer)

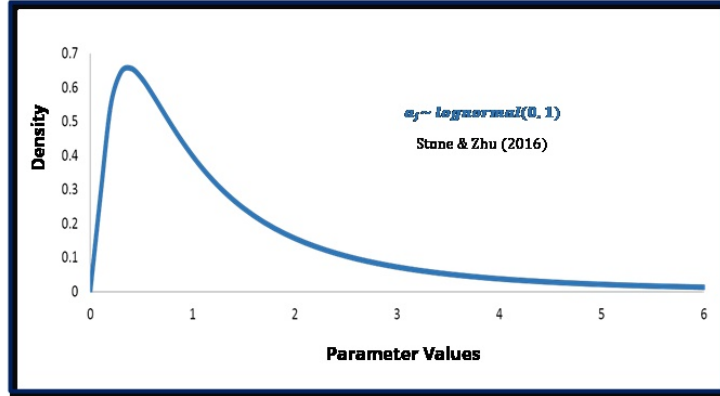
### Prior for Difficulty Parameter

$b_j \sim \text{normal}(0, 4)$   
Stone & Zhu (2016)

Prior for Difficulty Parameter      Prior for Discrimination Parameter      Prior for Guessing Parameter

## Prior Discrimination (Slide Layer)

### Prior for Discrimination Parameter



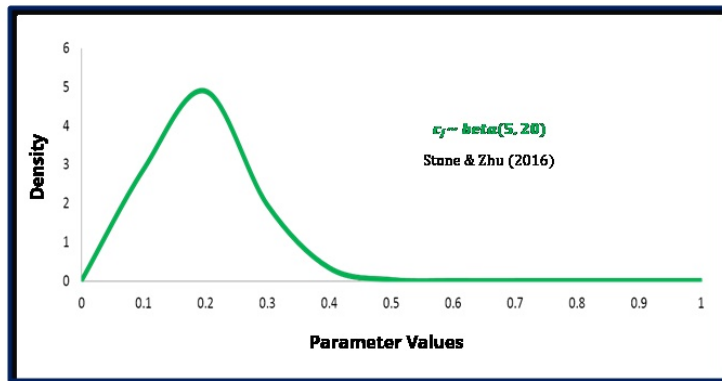
Prior for  
Difficulty Parameter

Prior for  
Discrimination Parameter

Prior for  
Guessing Parameter

## Prior Guessing (Slide Layer)

### Prior for Guessing Parameter





Prior for  
Difficulty Parameter

Prior for  
Discrimination Parameter

Prior for  
Guessing Parameter

## 4.10 Example: Instrumentation

**Example: Instrumentation**

**Instrument**

- 10-item USDA *Adult Food Security Scale Module* (AFSSM)
- Frequency and severity of experiencing food insecurity
- Affirmative responses = 1, non-affirmative responses = 0



**Respondents**

- Data collected at a public university in the South ( $n = 462$ )

**Statistical Model**

- Unidimensional 1PL model
- Higher scale scores indicate more severe food insecurity

## 4.11 Example: Sample Items

**Example: Item Types & Properties**

**Sample Item:**

*In the last 12 months, did you ever eat less than you felt you should because there wasn't enough money for food?*

Yes (1)  
 No (0)  
 Don't Know (0)

**Full Survey:** <https://www.ers.usda.gov/media/8282/short2012.pdf>

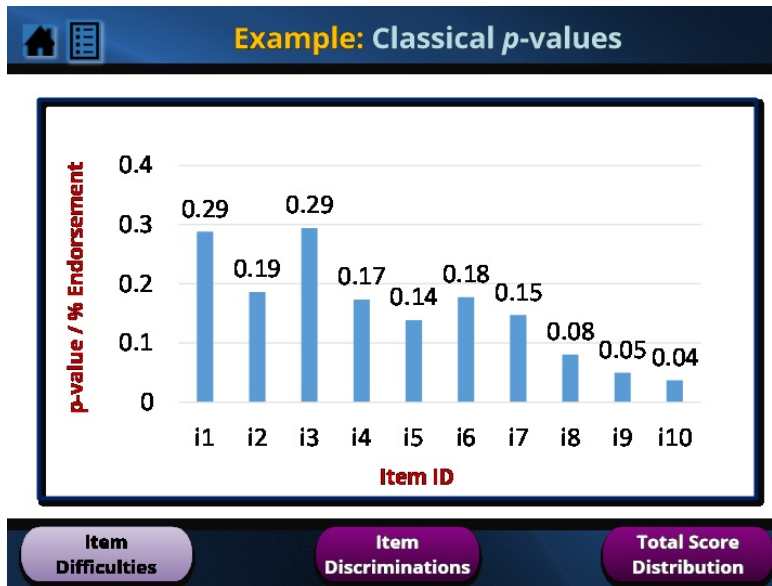
**Sample Statistics:** Access via buttons below

Item  
Difficulties

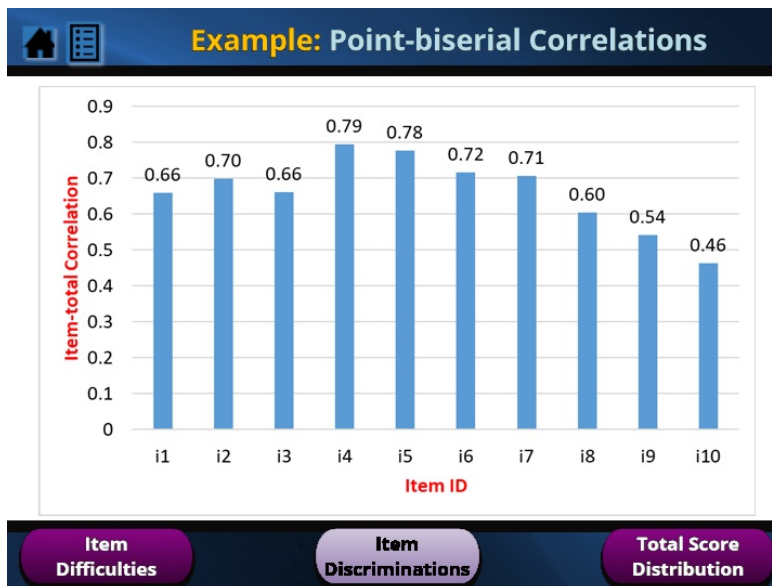
Item  
Discriminations

Total Score  
Distribution

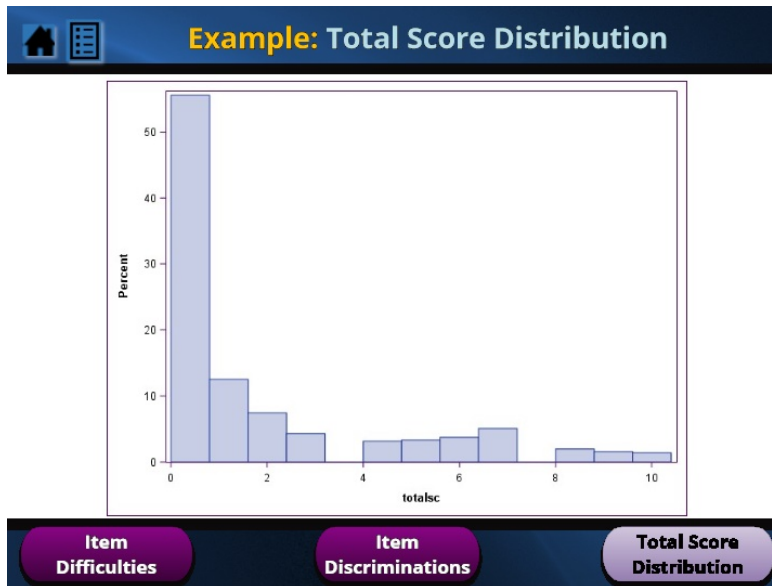
## Difficulty (Slide Layer)



## Discrimination (Slide Layer)



## Total score distribution (Slide Layer)



## 4.12 Model Estimation: Structure & Assumptions

**Model Estimation: Likelihoods**

$$\text{Likelihood} = \prod_i \prod_j P_{ij}^{x_{ij}} (1 - P_{ij})^{1 - x_{ij}}$$

$$\text{Log-Likelihood} = \sum_i \sum_j (x_{ij} * \log(P_{ij}) + (1 - x_{ij}) * \log(1 - P_{ij}))$$

- $P_{ij}$  = probability of correct response for the  $i^{\text{th}}$  learner and  $j^{\text{th}}$  item
- $x_{ij}$  = observed response (0 or 1) for the  $i^{\text{th}}$  learner and  $j^{\text{th}}$  item

---

**Example:** Response string [1 0 1]

$$P_{i1} * (1 - P_{i2}) * P_{i3}$$

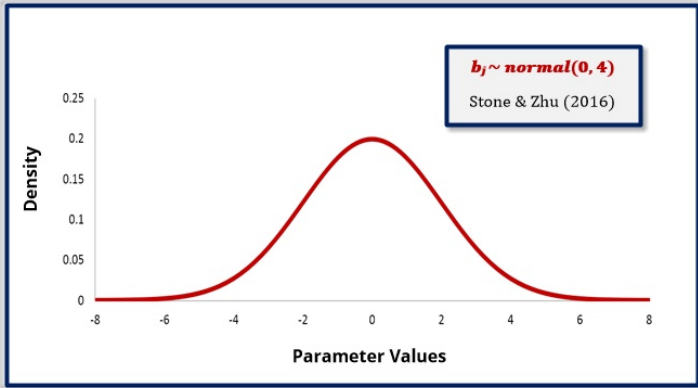
### 4.13 Model Estimation: Structure & Assumptions

**Model Estimation: Assumptions**

- **Monotonicity (non-decreasing response function)**  
Respondents that are more insecure about food have a higher probability of endorsing an item than respondents who are more secure
- **Unidimensionality (one dimension is appropriate for data)**  
Only the food insecurity trait is driving responses.
- **Local independence (no residual dependencies given model)**  
A response to one item is independent of the response to another item, after controlling for level of food insecurity

### 4.14 Model Estimation: Posterior Distributions

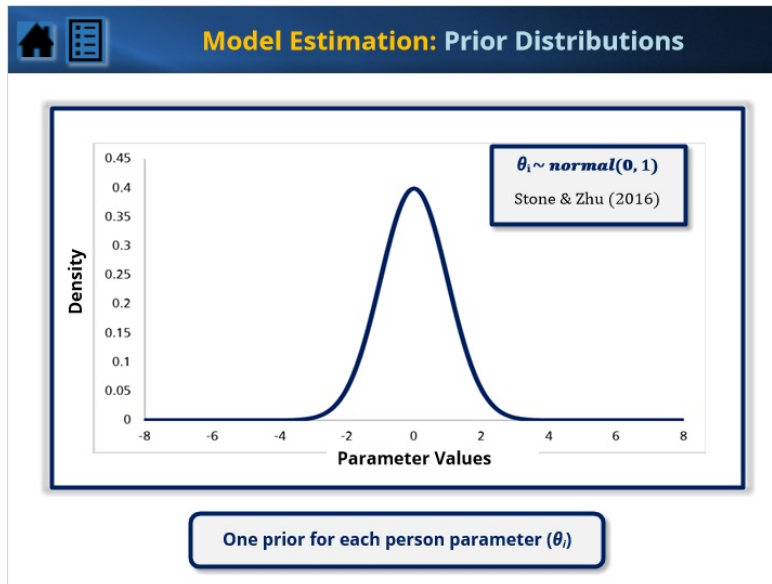
**Model Estimation: Prior Distributions**



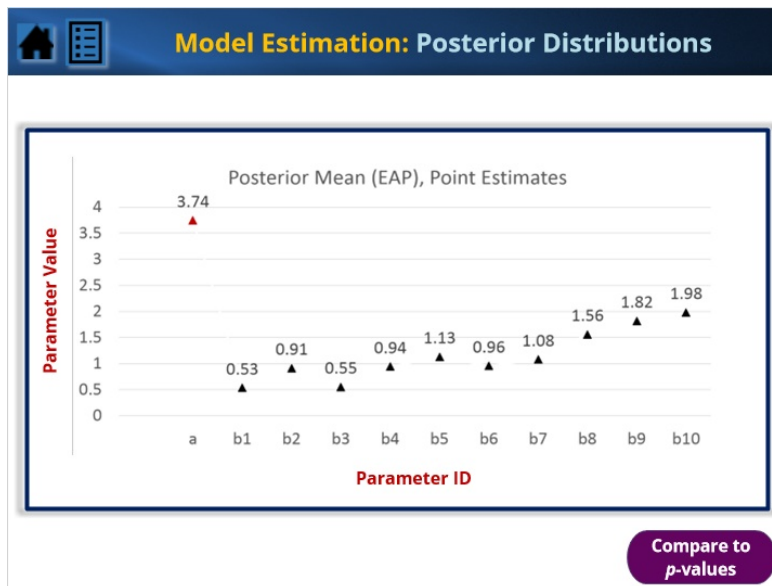
$b_j \sim \text{normal}(0, 4)$   
Stone & Zhu (2016)

One prior for each difficulty parameter (one per item)

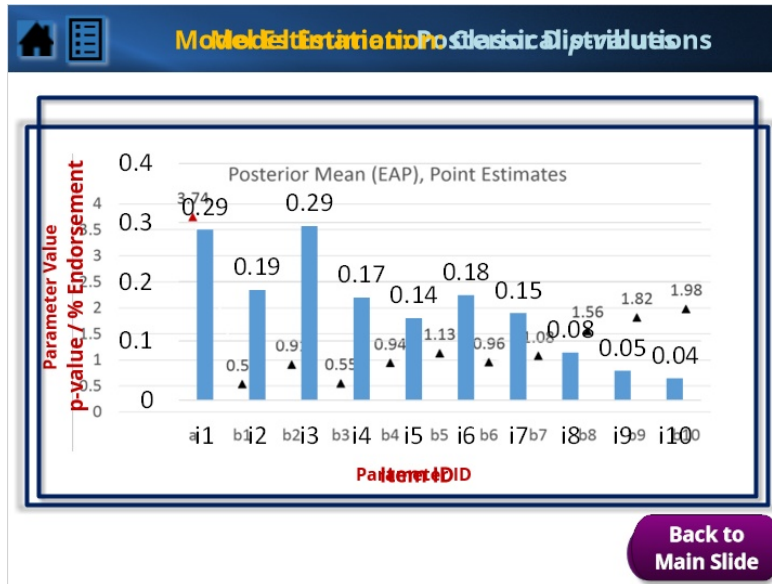
#### 4.15 Model Estimation: Posterior Distributions



#### 4.16 Model Estimation: Posterior Distributions



## p-values (Slide Layer)



### 4.17 PPMC: Computational Steps (I)

PPMC: Computational Steps



In simple terms, after estimating the posterior distributions:

- compute the discrepancy measure for the **single observed data set**
- compute the discrepancy measure for **each predicted data set**
- compare the observed value and the **distribution of predicted values**
- make a decision about model-data fit **based on the selected measure**

Click on the button to see a finer breakdown as used in the regression example or advance to illustration.

Finer Breakdown

## Finer Breakdown (Slide Layer)

**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

**Return to Main Slide**

## 4.18 General Principles (II)

**Computational Steps in IRT**

**Step 1: Ability Parameters**

An ability parameter is randomly sampled from each learner's posterior distribution for  $\theta_i$ .

**Step 2: Item Parameters**

A parameter value is randomly sampled from each item's posterior distribution for  $b_j$ .

**Step 3: Response Generation**

The combination of all sampled learner and item parameter values is plugged into the joint likelihood to create one complete predicted data set

**Repeat Steps 1-3** until the desired number of predicted data sets are generated

## 4.19 Computation (I)

**Example: Learner 48, Item 2**

$$P_{48,2}(X = 1|\theta_8) = \frac{\exp(a * (\theta_{48} - b_2))}{1 + \exp(a * (\theta_{48} - b_2))}$$

$$= \frac{\exp(3.9398 * (0.5909 - 0.4780))}{1 + \exp(3.9398 * (0.5909 - 0.4780))}$$

$$= \frac{1.56}{1 + 1.56}$$

$$= 0.61$$



Continue for other items (new  $a$  and  $b$  values) and other learners (new  $\theta$  values)

## 4.20 Computation (II)

**Example: Learner 48, Item 2**


Repeat for all learners and items to create one complete predicted data set.

## 4.21 Untitled Slide



### Discrepancy Measures

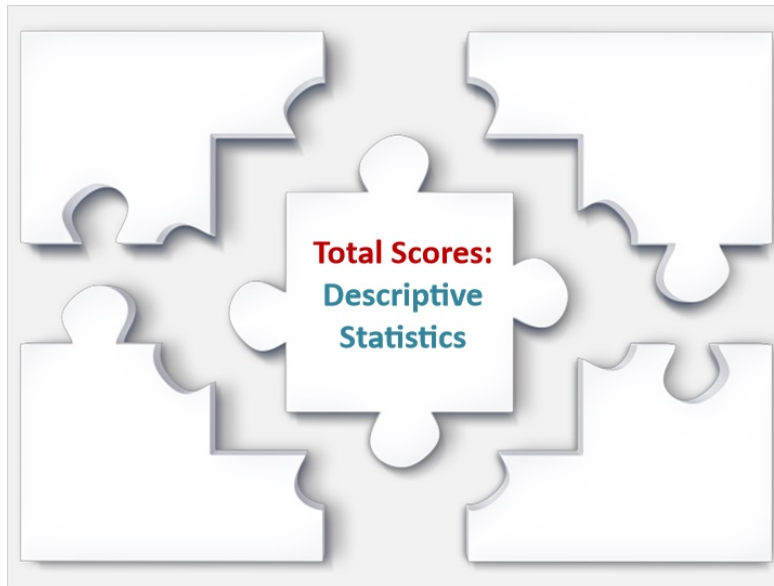
- **Descriptive summary statistics** (mean / standard deviation of total score)
- **Frequency distribution** (total score)
- **Item parameters** (percent correct, item-total correlation)
- **Dependency indices** (Yen's Q3)



## 4.22 Discrepancy Measure Selection




### **4.23 Bookmark: Descriptive Statistics**





### **4.24 Video: Descriptive Statistics**

This video contains audio narration for these slides plus a code-based demonstration of how to perform these analyses in SAS.



Alternatively, advance to the next slide to listen to audio-narrated slides without any instructions for SAS.


## 4.25 Descriptive Statistics (I)

  **Example: Descriptive Statistics**



**Step 1: Compute Statistics for Observed Data**

Compute the mean and standard deviation of the total score in the observed data

$M_{obs} = 1.71$  and  $SD_{obs} = 2.64$




## 4.26 Descriptive Statistics (II)

  **Example: Descriptive Statistics**



**Step 2: Compute Statistics for Predicted Data**

Compute the means and standard deviations of the total score in each of the 1,000 predicted data sets

$[M_{sim1}, M_{sim2}, \dots, M_{sim1000}]$  and  $[SD_{sim1}, SD_{sim2}, \dots, SD_{sim1000}]$



## 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)

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



## 4.29 Descriptive Statistics (IV)


  **Example: Descriptive Statistics**

**Step 4: Make an Inferential Decision**

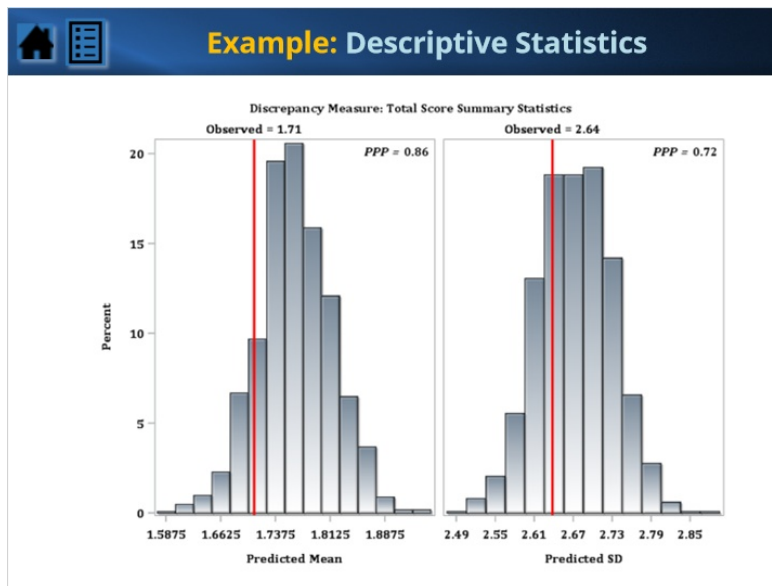
---

Generally, PPP values of  $< .05$  or  $> .95$  are taken as an indication that misfit is likely present based on this discrepancy measure

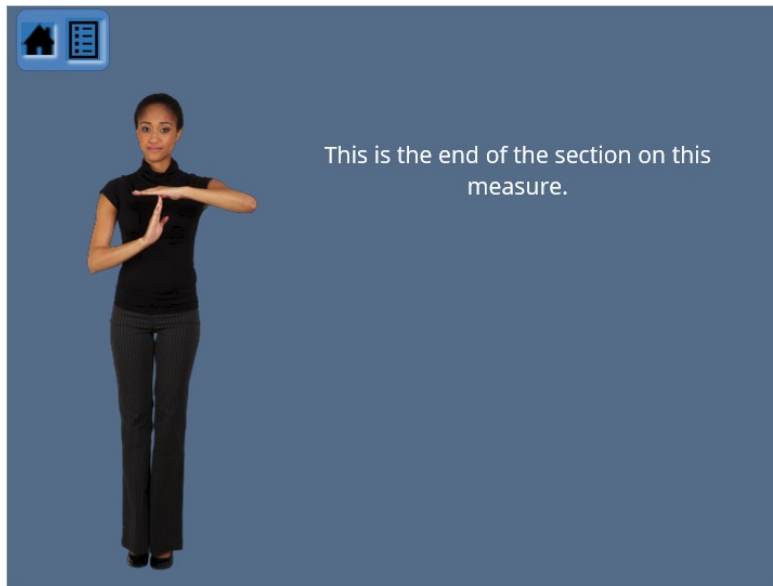
**Graphical procedures can help with this step!**



## 4.30 Descriptive Statistics (V)



### 4.31 Bookend: Descriptive Statistics



### 4.32 Bookmark: Total Score Distribution





### 4.33 Video: Total Score Distribution

This video contains audio narration for these slides plus a code-based demonstration of how to perform these analyses in SAS.



Alternatively, advance to the next slide to listen to audio-narrated slides without any instructions for SAS.


### 4.34 Total Score Distribution (I)

  **Example: Total Score Distribution**



**Step 1: Compute Frequencies for the Observed Data**

---

Compute and save the observed total score frequencies




### 4.35 Total Score Distribution (III)

  **Example: Total Score Distribution**



**Step 2b: Compute Percentiles for Predicted Data**

---

Compute and save the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> 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)

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


### 4.37 Total Score Distribution (V)

**Example: Total Score Distribution**

**Step 3 (continued): Compare Observed and Predicted Data**

Observed score frequencies

5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of simulated frequencies

Total score

total	COUNT	p5	p50	p95
0	284	259	5	288
1	64	56	68	82
2	38	28	37	46
3	22	18	25	33
4	16	14	20	28
5	17	12	18	25
6	19	12	18	24
7	26	11	16	23
8	10	8	13	20
9	8	6	10	15
10	7	5	9	13


### 4.38 Total Score Distribution (VI)

**Example: Total Score Distribution**

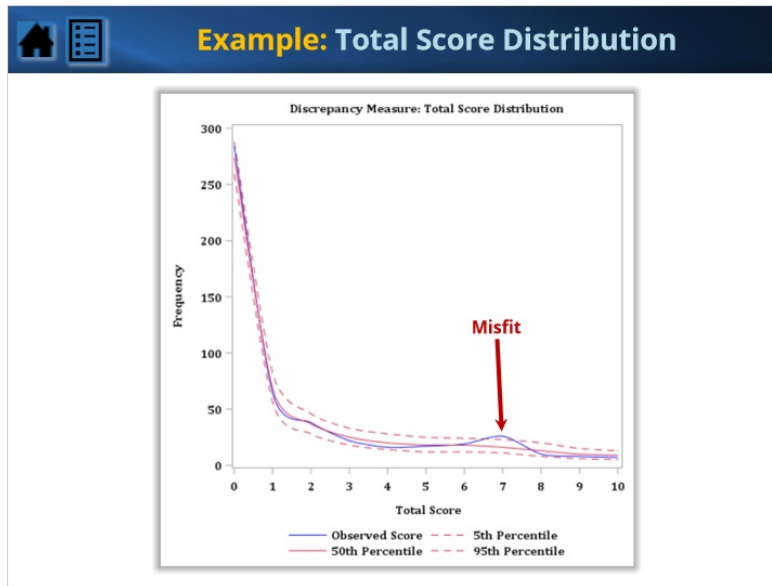
**Step 4: Make an Inferential Decision**

Generally, PPP values of  $< .05$  or  $> .95$  are taken as an indication that misfit is likely present based on this discrepancy measure

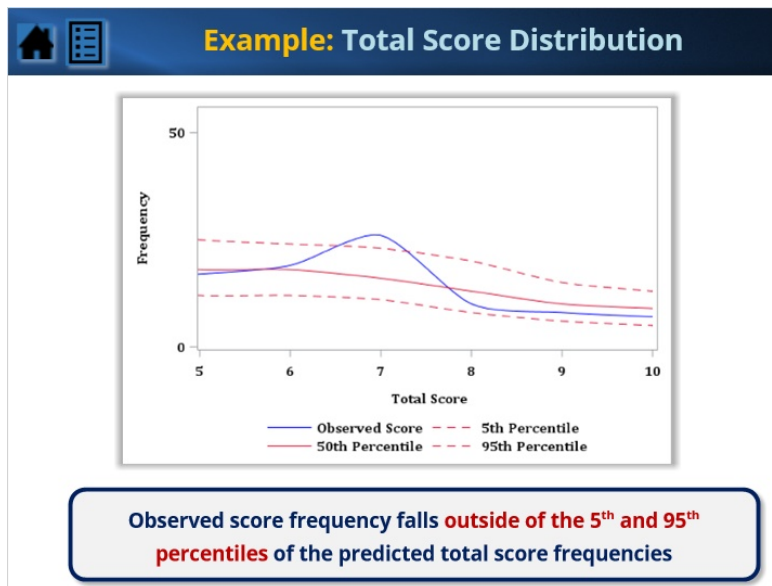
Rather than examine an overall PPP value for the whole distribution, **use a graphical approach!**



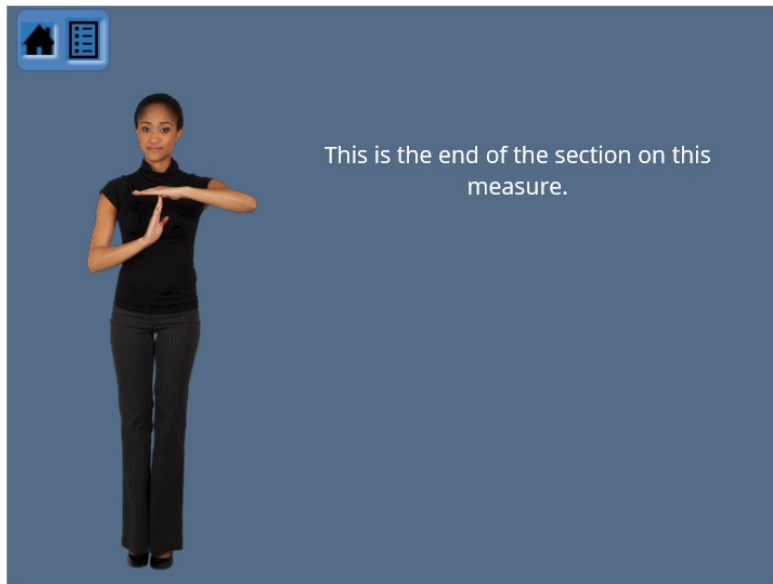
### 4.39 Total Score Distribution (VII)



### 4.40 Total Score Distribution (VIII)



#### ***4.41 Bookend: Total Score Distribution***



#### ***4.42 Bookmark: Item-total Correlations***





## 4.43 Video: Item-total Correlations

This video contains audio narration for these slides plus a code-based demonstration of how to perform these analyses in SAS.



Alternatively, advance to the next slide to listen to audio-narrated slides without any instructions for SAS.

## 4.44 Item-total Correlations (I)


  **Example: Item-Total Correlation**

**Step 1: Compute Item-total Correlations for Observed Data**



Compute the observed item-total correlations for each item

$$r_{\text{item}1}, r_{\text{item}2}, \dots, r_{\text{item}i}$$

(10 items = 10 item-total correlations)



## 4.45 Item-total Correlations (II)

### Example: Item-Total Correlation


**Step 2: Compute Item-total Correlations for Predicted Data**

---



Compute the item-total correlations for each item for each of the 1,000 simulated data sets

$$r_{\text{item}1, \text{sim}1}, r_{\text{item}2, \text{sim}1}, \dots, r_{\text{item}j, \text{sim}1000}$$

(10 items x 1,000 simulated data sets = 10,000 item-total correlations)



## 4.46 Item-total Correlations (III)

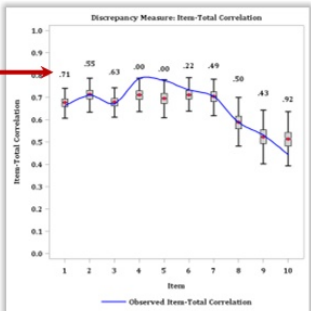
### Example: Item-Total Correlation

**Step 3: Compare Observed and Predicted Data**


---

Plot the distributions of the 1,000 item-total correlations by item, identify the observed item-total correlation for each item, and compute the posterior predictive  $p$ -value (PPP) for each item



**PPP values** →



Item	PPP value
1	.71
2	.55
3	.63
4	.60
5	.60
6	.22
7	.49
8	.50
9	.43
10	.92



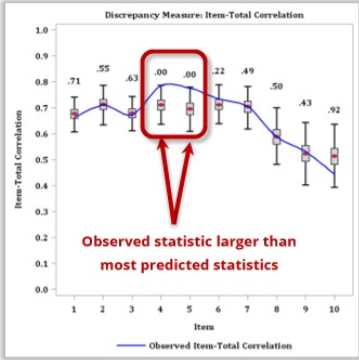
## 4.47 Item-total Correlations (IV)

### Example: Item-Total Correlation



Step 4: Make an Inferential Decision

Generally, PPP values of  $< .05$  or  $> .95$  are taken as an indication that misfit is likely present based on this discrepancy measure

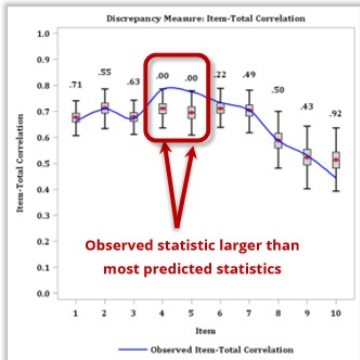


Item	Observed Item-Total Correlation	PPP Value
1	0.71	
2	0.55	
3	0.63	
4	0.80	.00
5	0.80	.00
6	0.72	.22
7	0.49	
8	0.50	
9	0.43	
10	0.52	.92

## 4.48 Item-total Correlations (V)

### Example: Item-Total Correlation



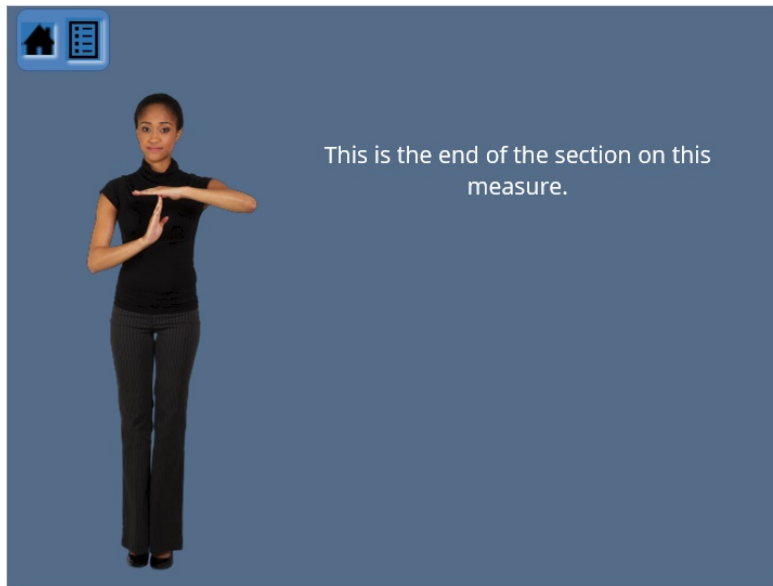
Item	Observed Item-Total Correlation	PPP Value
1	0.71	
2	0.55	
3	0.63	
4	0.80	.00
5	0.80	.00
6	0.72	.22
7	0.49	
8	0.50	
9	0.43	
10	0.52	.92

Interpretation

For items 4 and 5 (PPP  $< .01$ ), the model always predicts data with an item-total correlation less than the observed item-total correlation.

The model may not be adequately describing the discriminating capability of items 4 and 5 and a 2PL may be more appropriate for these items.

#### 4.49 Bookend: Item-total Correlations



This is the end of the section on this measure.

#### 4.50 Bookmark: Percent Correct



**Item  
Difficulty:**  
% Correct at  
Score Level

## 4.51 Video: Percent Correct

This video contains audio narration for these slides plus a code-based demonstration of how to perform these analyses in SAS.



Alternatively, advance to the next slide to listen to audio-narrated slides without any instructions for SAS.

## 4.52 Percent Correct (I)



### Example: Item Percent Correct

#### Step 1: Compute Percent Correct for Observed Data

Compute the proportion of correct responses for each item / item mean for each item at each score level

**For a score of 0:**

$p_{\text{item}1}, p_{\text{item}2}, \dots, p_{\text{item}j}$   
(10 items = 10 proportions correct)

...

**For a score of 10:**

$p_{\text{item}1}, p_{\text{item}2}, \dots, p_{\text{item}j}$   
(10 items = 10 proportions correct)



## 4.53 Percent Correct (II)

**Example: Item Percent Correct**


**Step 2: Compute Percent Correct for Predicted Data**

Compute the proportion correct for each item for each of the 1,000 predicted data sets at each score level

**For a score of 0:**  
 $p_{\text{item1},\text{sim1}}, p_{\text{item2},\text{sim1}}, \dots, p_{\text{itemj},\text{sim1000}}$   
(10 items x 1,000 simulated data sets = 10,000 proportions correct)

...

**For a score of 10:**  
 $p_{\text{item1},\text{sim1}}, p_{\text{item2},\text{sim1}}, \dots, p_{\text{itemj},\text{sim1000}}$   
(10 items x 1,000 simulated data sets = 10,000 proportions correct)

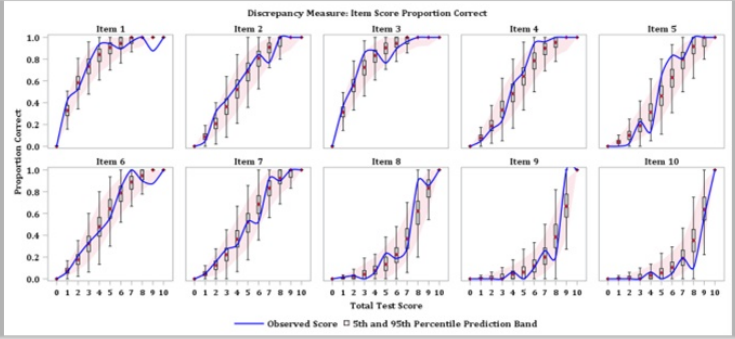


## 4.54 Percent Correct (III)

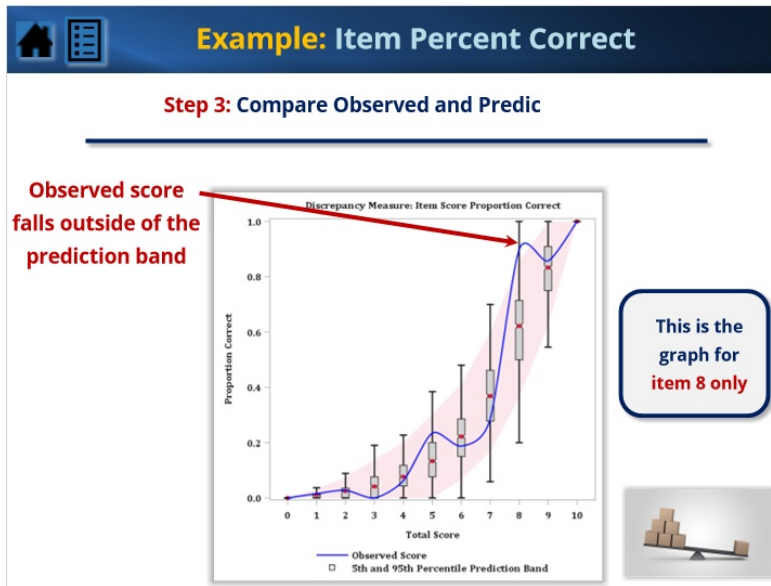
**Example: Item Percent Correct**

**Step 3: Compare Observed and Predicted Data**

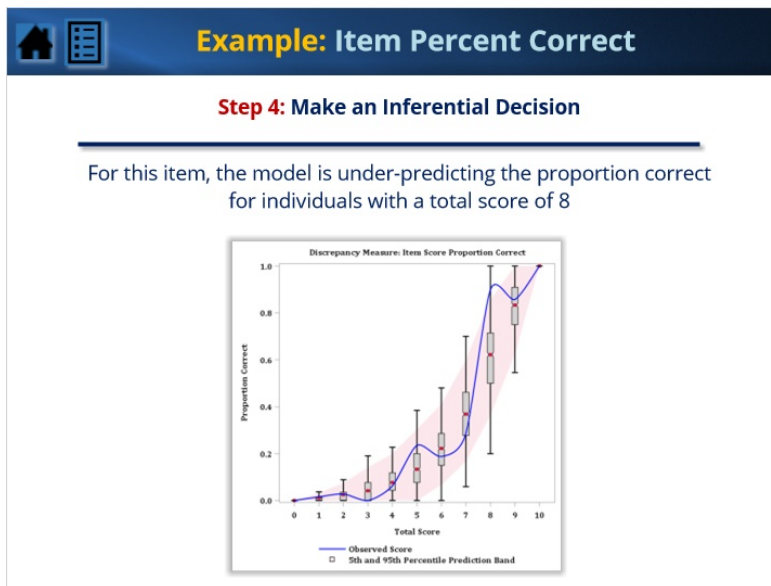
Plot the distributions of the 1,000 percent correct values by item, identify the 5th and 95th percentiles of the distributions, identify the observed percent correct values for each item, and compute the Bayesian posterior predictive p-value (PPP) for each item



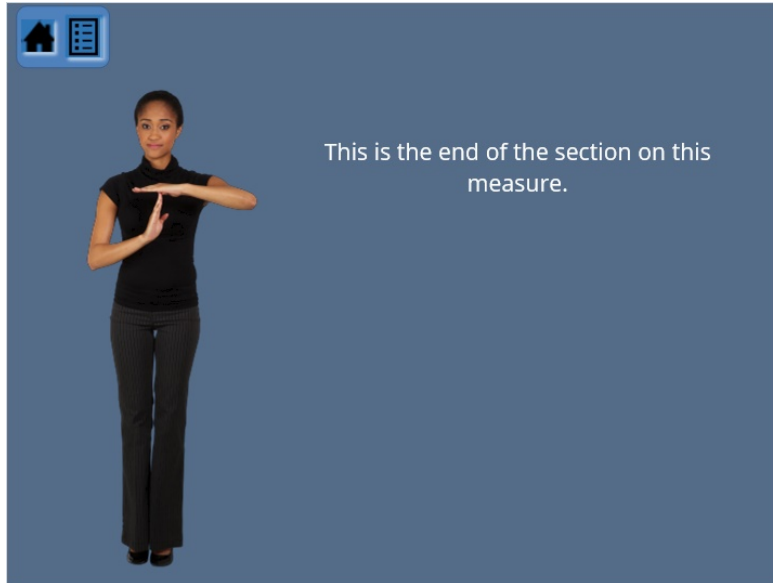
## 4.55 Percent Correct (IV)



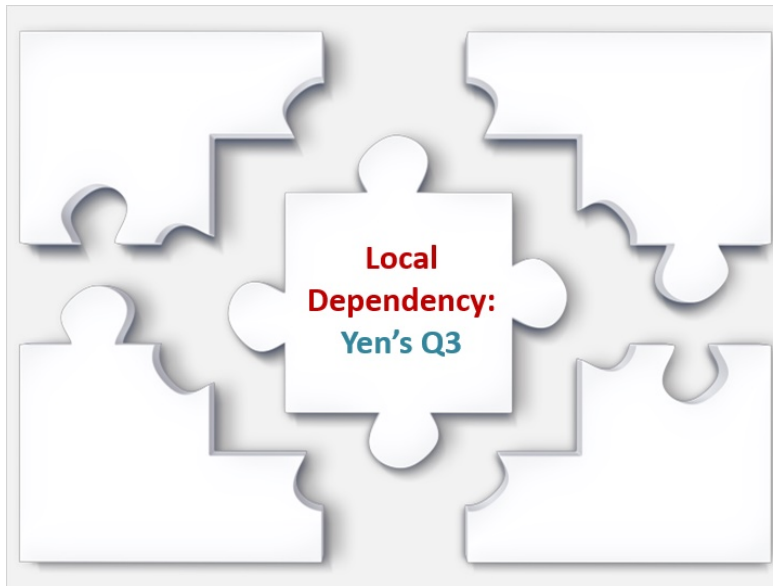
## 4.56 Percent Correct (V)



### 4.57 Bookend: Percent Correct



### 4.58 Bookmark: Yen's Q3





## 4.59 Video: Yen's Q3

This video contains audio narration for these slides plus a code-based demonstration of how to perform these analyses in SAS.



Alternatively, advance to the next slide to listen to audio-narrated slides without any instructions for SAS.



## 4.60 Yen's Q3 (I)



### Example: Yen's Q<sub>3</sub>

- The IRT model assumption of **local independence** can be violated through **response dependency** and **multidimensionality**.
- The result is **inter-item correlations beyond** what can be attributed to the latent variable in the model
- Yen (1984) proposed the **Q3 statistic** for detecting local dependence that represents the **correlation of the residuals for two items**
- **Residuals** are the **difference** between the observed score of each learner and the predicted response



## 4.61 Yen's Q3 (II)

  **Example: Yen's Q<sub>3</sub>**

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

$$r_i = X_i - E(X_i|\theta)$$
$$r_i = \text{Observed} - \text{Predicted}$$
$$Q3_{ij} = \text{corr}(r_i, r_j)$$

## 4.62 Yen's Q3 (III)


  **Example: Yen's Q<sub>3</sub>**

**Step 1: Compute Yen's Q3 for Observed Data**



---

$$Q3_{\text{item}12}, Q3_{\text{item}13}, \dots, Q3_{\text{item}(j-1)j}$$

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 $Q_3$


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

$$Q3_{\text{item}12, \text{sim}1}, Q3_{\text{item}13, \text{sim}1}, \dots, Q3_{\text{item}(i-1), \text{sim}1}, \dots, Q3_{\text{item}(i-1), \text{sim}1000}$$

(45 item pairs x 1,000 simulated data sets = 45,000 Q3 statistics)



## 4.64 Yen's Q3 (V)


 

### Example: Yen's $Q_3$

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)

### Example: Yen's Q<sub>3</sub>

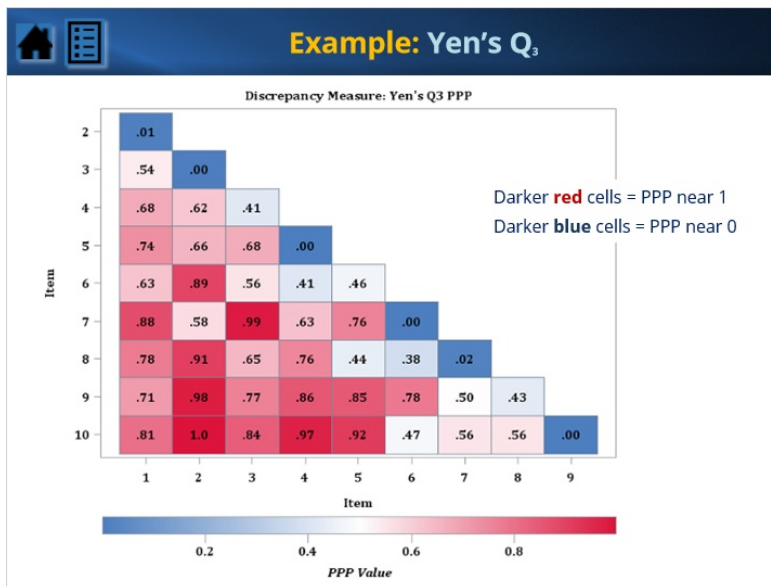
Step 4: Make an Inferential Decision

---

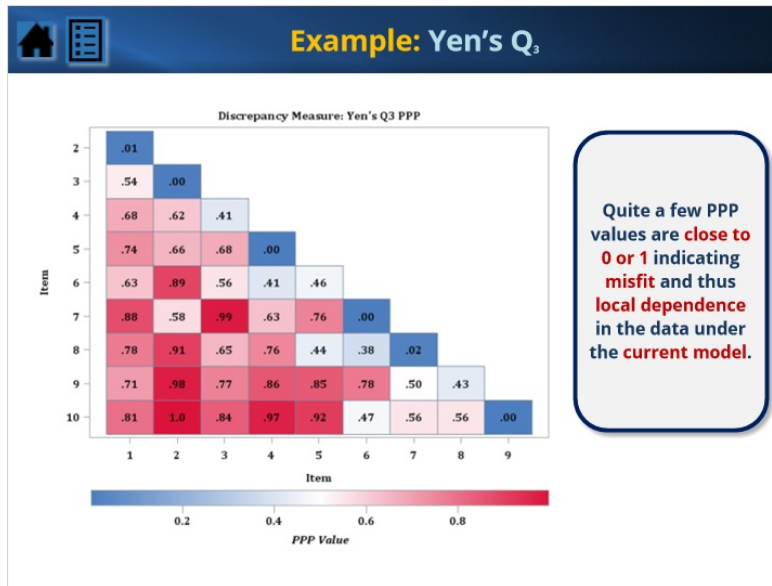
Generally, PPP values of  $< .05$  or  $> .95$  are taken as an indication that misfit is likely present based on this discrepancy measure.

This is best visualized via a **"heat map"** or table that summarizes the PPP values for all item pairs. We could also organize the PPP values as pie charts or other graphical display!

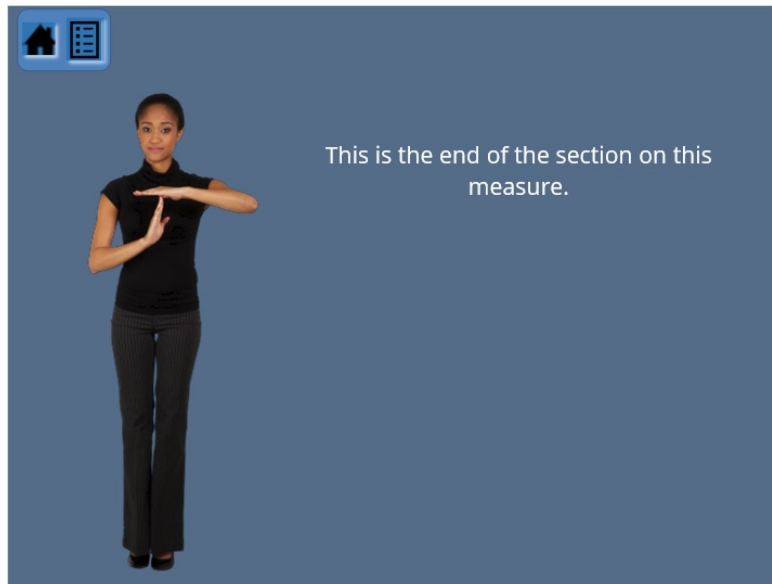
## 4.66 Yen's Q3 (VII)





## 4.67 Yen's Q3 (VIII)



## 4.68 Bookend: Yen's Q3



## 4.69 Summary (I)





### Summary

Through PPMC analyses, the 1PL IRT model was shown to be a poor fit to the USDA AFSSM data.

- The **predicted total score distribution** did not adequately represent the **observed score distribution**
- The **item-total correlation** revealed the model is not adequately describing the **discriminating capability** of two items and **a 2PL may be more appropriate for these items.**
- Yen's  $Q_3$  revealed **local dependence** in the data under the 1PL and a model to **account for dependence** may be appropriate.


## 4.70 Summary (II)





### Summary

**As with the regression examples, the choice of discrepancy statistic is important!**

Yen's  $Q_3$  is only indicative of local item dependence – it cannot diagnose whether a 1PL or 2PL of similar dimensionality is more appropriate for the data.




## 4.71 SAS Resources



### Additional Resources


**General resources for using SAS PROC MCMC**



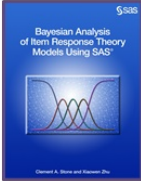
[https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#mcmc\\_toc.htm](https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#mcmc_toc.htm)

---

**Specific resources for using SAS PROC MCMC to estimate IRT models**




Ames & Samonte (2015)



Stone & Zhu (2016)

## 4.72 Bookend: IRT



This is the end of this section.