



1

// Module 2 Final Project  
Using multivariate linear regression (MLR) to predict life expectancy  
  
Jonathan Vasquez

2

Contents

- Overall Goal
- Data Gathering / Data Cleaning
- Exploratory Data Analysis
- Hypothesis Testing
- Modeling Data
- Conclusion
- Future work

3

// Overall Goal  
  
Predict life expectancy via:

- Health Behaviors
- Clinical Care Access
- Socioeconomic Factors
- Physical Environment

  
Lasso model's prediction within 0.5587 SD's.  
  
Lasso Training Error: 1.0213  
Cross-Validation Error: 1.0017  
SD: 2.146

4

// Data Gathering / Data Cleaning

**Data Gathering:**  
Download [Country Health Statistics](#)  
Collaboration: [DataCamp](#)  
Health datasets: 3,153 countries/50 states  
Year: 2015

**Data Cleaning:**

- Imported with columns
- Reordered "Year"
- Reordered columns by State
- Filled in NaN's

5

// Exploratory Data Analysis

Relationships in Life Expectancy

- Tobacco Use
- Diet & Exercise
- Alcohol & Drug Use
- Sexual Activity
- Access to Care
- Quality of Care
- Education
- Employment
- Income
- Family & Social Support
- Community Safety
- Air & Water Quality
- Housing & Thriving

6

// Hypothesis Testing

Reject the null hypothesis:

- Correlation: Socioeconomic & Life Expectancy
- Multiple Regression
- Hypothesis Testing

  
Fail to reject the null

// 

# Module 2 Final Project

Using multivariate linear regression (MLR) to predict life expectancy.

Jonathan Vasquez

Click to add speaker notes

