Objectives of Model Fitting: Inference vs. Prediction

Brady T. West

Two Main Objectives of Model Fitting

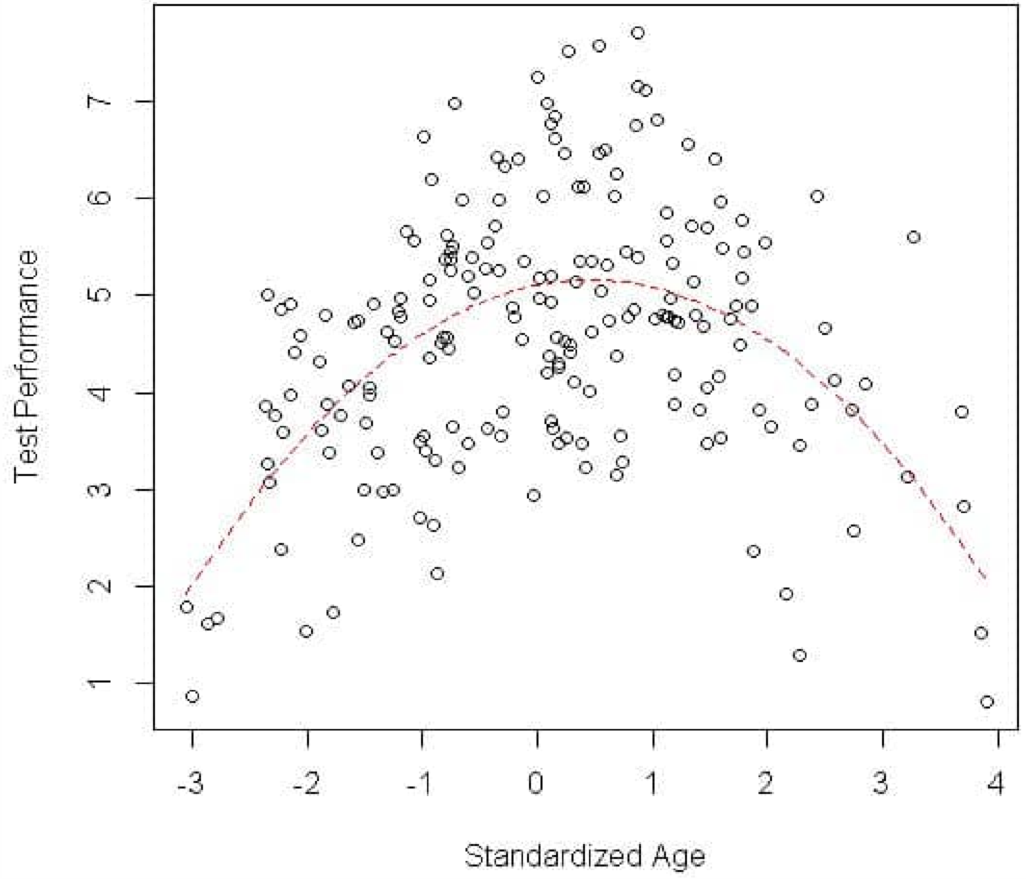
I. Making inference about relationships
between variables in a given data set

II. Making predictions/forecasting future outcomes, based on models estimated using historical data

Objective 1: Making Inference

Predictor
Age (Standardized)

Test Performance
(0 – 8 points)



Objective 1: Making Inference

Predictor
Age (Standardized)

→

Test Performance
(0 – 8 points)

$$\text{Performance} = a + b*\text{age} + c*\text{age}^2 + e$$

Objective 1: Making Inference

Predictor
Age (Standardized)

→

Test Performance
(0 – 8 points)

Performance = a + b*age + c*age^2 +e

- **e** = "error" = actual perf – *predicted* perf using **regression function**

Errors are normally distributed, mean 0, constant variance (given age)

Mean Performance = a + b*age + c*age^2

Objective 1: Making Inference

Make inference about relationship between age and performance
 examining estimates of regression parameters (a, b, and c)

Estimates of parameters + their standard errors we can …

Objective 1: Making Inference

Make inference about relationship between age and performance
 examining estimates of regression parameters (a, b, and c)

Estimates of parameters + their standard errors we can …

Test hypotheses
about whether
parameters equal to 0

Objective 1: Making Inference

Make inference about relationship between age and performance
 examining estimates of regression parameters (a, b, and c)

Estimates of parameters + their standard errors we can …

Test hypotheses
about whether
parameters equal to 0

Form confidence interval
for parameters
~ is 0 in interval?

Objective 1: Making Inference

$$\textbf{perf} = \textbf{a} + \textbf{b}*\textbf{age} + \textbf{c}*\textbf{age}^2 + \textbf{e}, \quad \textbf{where e} \sim N(0, \sigma^2)$$

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

Objective 1: Making Inference

$$\text{perf} = a + b*\text{age} + c*\text{age}^2 + e, \quad \text{where } e \sim N(0, \sigma^2)$$

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

For each parameter we could calculate a test statistic:

$$\text{Test statistic} = \frac{estimate - 0}{standard\ error}$$

Objective 1: Making Inference

perf = a + b*age + c*age^2 + e,   where e ~ N(0, σ^2)

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

For parameter b:

$$t = \frac{estimate - 0}{standard\ error} = \frac{0.24}{0.06} = 4$$

Objective 1: Making Inference

perf = a + b*age + c*age^2 + e, where e ~ N(0, σ^2)

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

For parameter b:

$$t = \frac{estimate - 0}{standard\ error} = \frac{0.24}{0.06} = 4$$

The estimated coefficient for age is 4 standard errors above 0 ~
A big difference □ H_0: b = 0 would be rejected, significant result!

IVQ … Objective 1: Making Inference

perf = a + b*age + c*age^2 + e, where e ~ N(0, σ^2)

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

Compute test statistics for parameter a and c to assess if significant

$$t = \frac{estimate - 0}{standard\ error}$$

Objective 1: Making Inference

perf = a + b*age + c*age^2 + e, where e ~ N(0, σ^2)

Parameter Estimates

Estimate of a = 5.11 (SE = 0.10)

Estimate of b = 0.24 (SE = 0.06)

Estimate of c = -0.26 (SE = 0.03)

Test Statistic:

a: t = 5.11 / 0.10 = 51.1

b: t = 0.24 / 0.06 = 4.0

c: t = -0.26 / 0.03 = -8.67

For each parameter, test statistic "large distance"

□ H_0: parameter = 0 would be rejected

Relationship between age and performance is significant!

Objective 1: Making Inference

Inferences about relationships!

$$\textbf{perf} = \textbf{a} + \textbf{b*age} + \textbf{c*age}^2 + \textbf{e}, \quad \textbf{where e} \sim \textbf{N}(0, \sigma^2)$$

Estimate of **a** = 5.11 (SE = 0.10)

a represents mean test performance
when age is equal to the mean in the data set

 average test performance at this age is 5.11 points
this is significantly different from 0

Objective 1: Making Inference

Inferences about relationships!

$$\mathbf{perf = a + b{*}age + c{*}age^2 + e, \quad where\ e \sim N(0, \sigma^2)}$$

Estimate of **b** = 0.24 (SE = 0.06)

b represents expected rate of increase in performance when standardized age is zero

 This is positive and significantly different from 0

Objective 1: Making Inference

Inferences about relationships!

perf = a + b*age + c*age^2 + e,   where e ~ N(0, σ^2)

Estimate of **c** = -0.26 (SE = 0.03)

c represents **non-linear acceleration** in performance as function of age, captures extent of non-linear relationship

Negative value after initial acceleration, additional increases in age **reduce** test performance, This aspect of relationship is significantly different from 0

Objective 1: Making Inference

Inferences about relationships!

$$\text{perf} = a + b*\text{age} + c*\text{age}^2 + e, \quad \text{where } e \sim N(0, \sigma^2)$$

Think about it …

What if estimate of **c** was
not significantly different from 0?

What might this **indicate about the relationship** between performance and age?

Objective 1: Making Inference

Inferences about relationships!

$$\mathbf{perf = a + b*age + c*age^2 + e, \quad where\ e \sim N(0, \sigma^2)}$$

Think about it …

What if estimate of **c** was
not significantly different from 0?

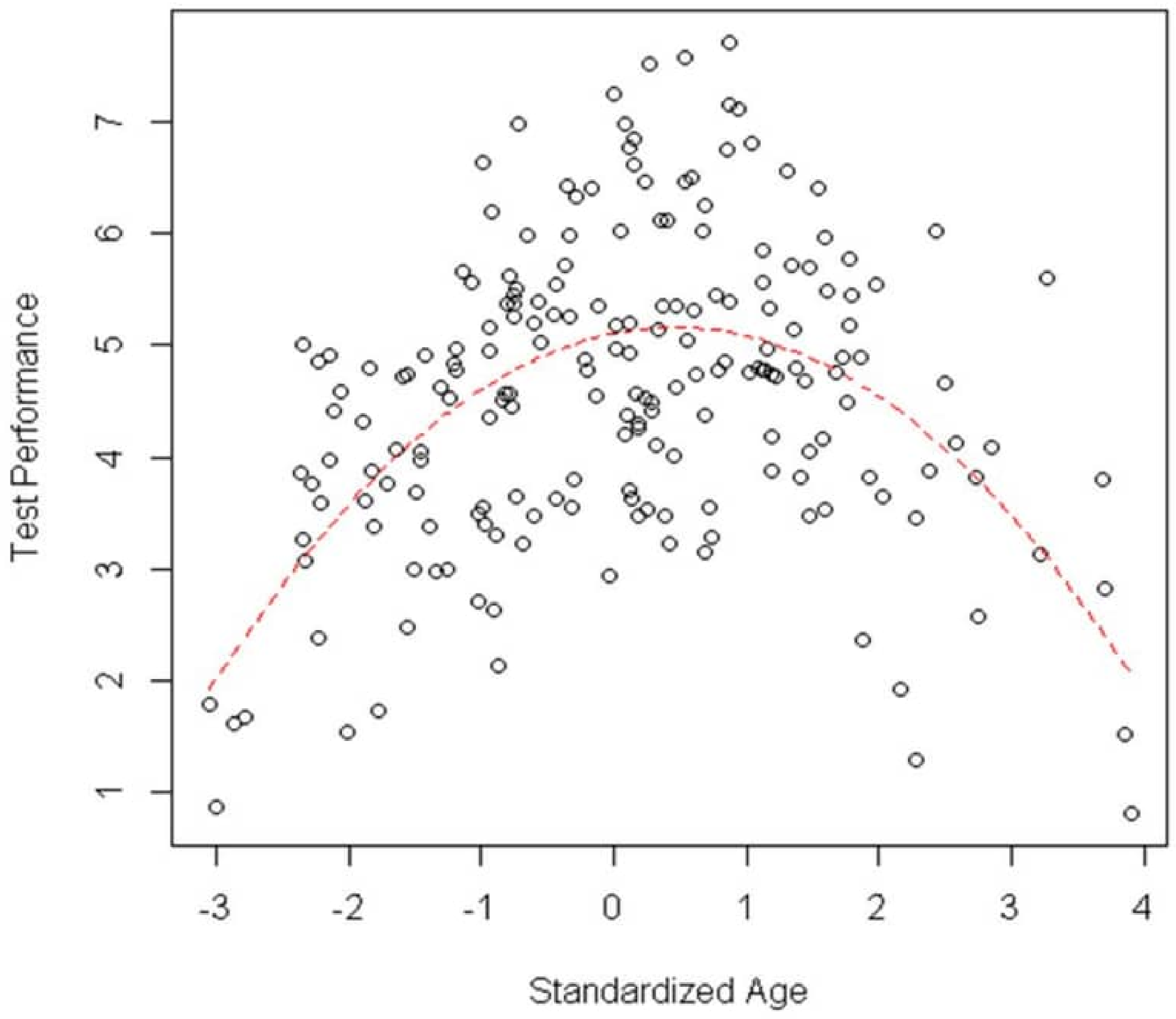
 evidence of **strictly LINEAR relationship** between performance and age

Objective 2: Making Predictions

Scatterplot shows **predicted values** of test performance as a function of age, based on fitted regression model:

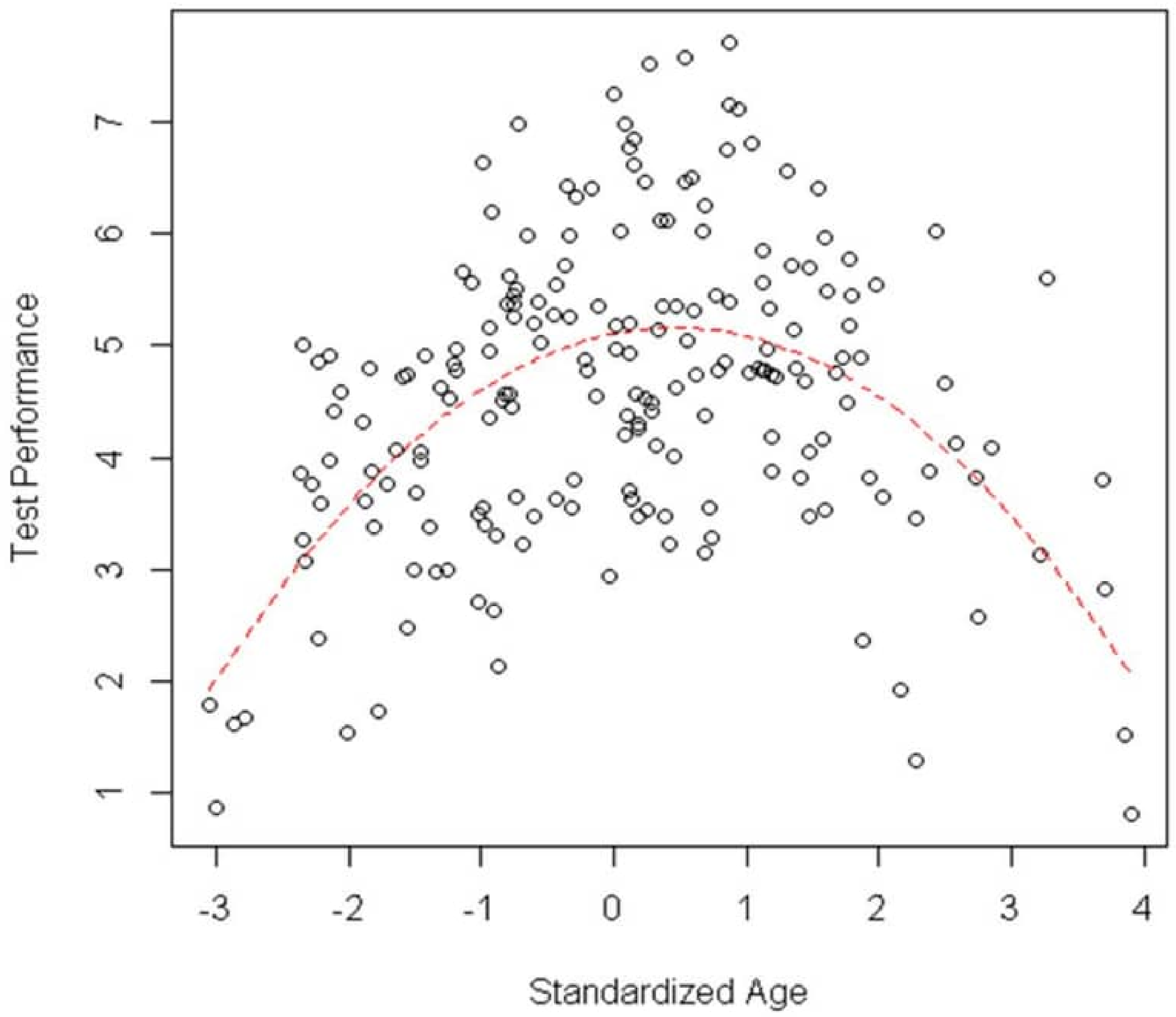
$$perf = 5.11 + 0.24*age - 0.26*age^2 + e$$

Could "plug in" values of age to compute **predictions** of performance!



IVQ …Objective 2: Making Predictions

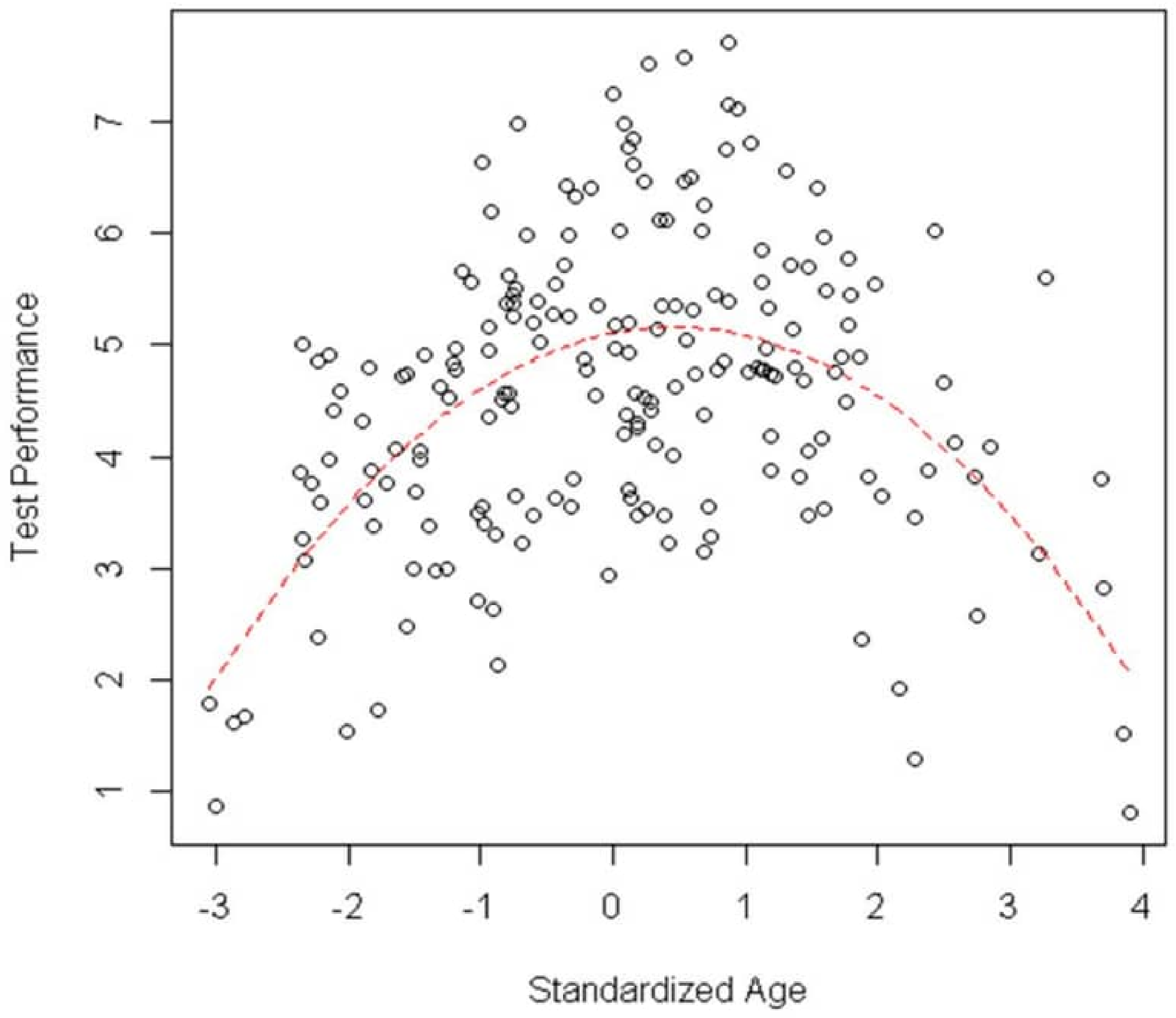
Use the fitted regression model to predict the performance at a standardized age of +1:



Objective 2: Making Predictions

Use the fitted regression model to predict the performance at a standardized age of +1:

predicted performance
$= 5.11 + 0.24*(1) - 0.26*(1)^2$
$= 5.09$ points

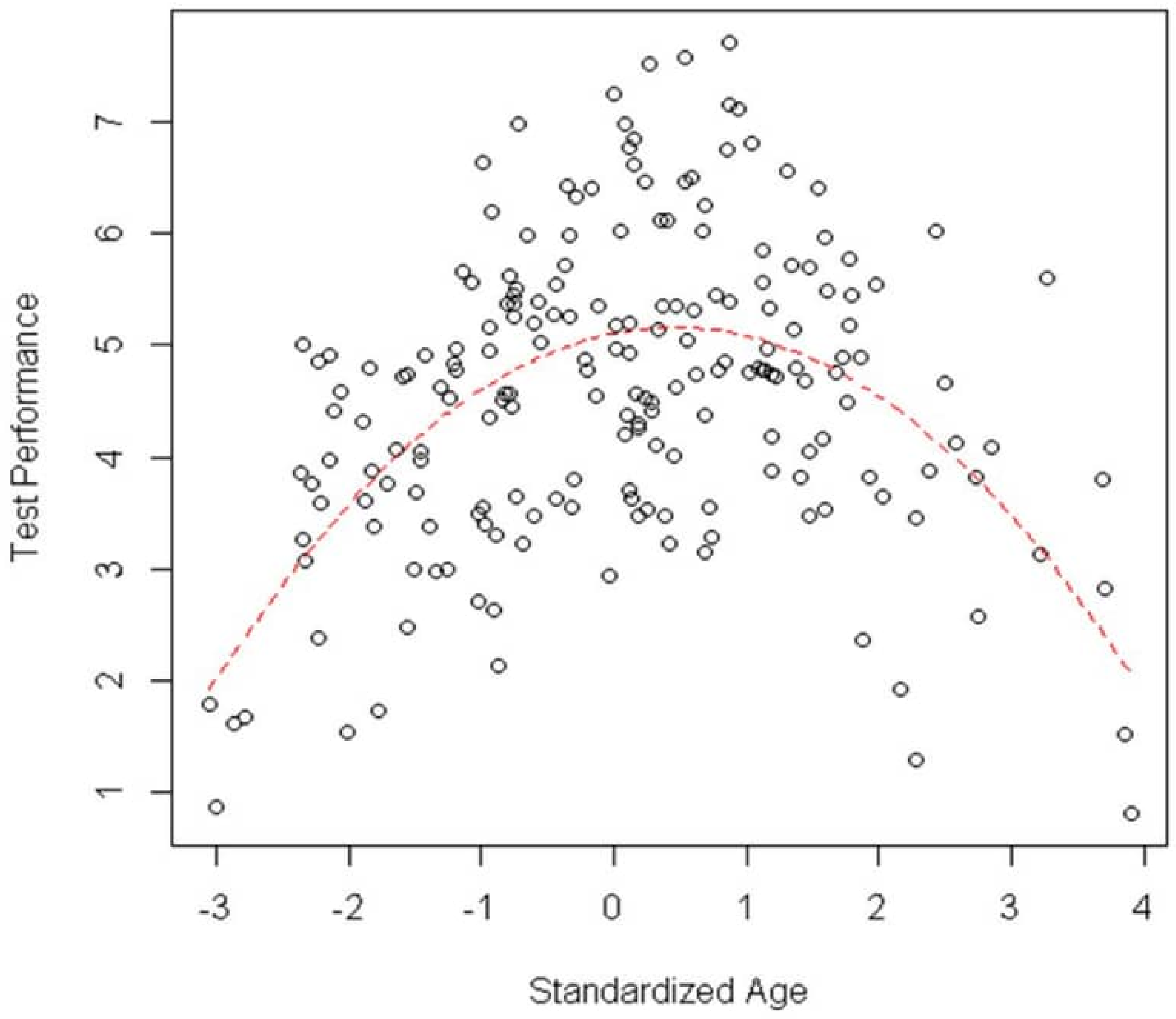


Objective 2: Making Predictions

Use the fitted regression model to predict the performance at a standardized age of +1:

predicted performance
$= 5.11 + 0.24*(1) – 0.26*(1)^2$
$= 5.09$ points

Check it out: does 5.09 points make sense with the plot?



Objective 2: Making Predictions

Remember…

- Using simple model for **__mean__ test performance**
  - predictions represent ***expectations*** of what mean test performance will be for a future observation

Objective 2: Making Predictions

Remember…

- Using simple model for <u>**mean**</u> **test performance**
   predictions represent ***expectations*** of what mean test performance will be for a future observation
- Don't forget about the errors ~ predictions will have **uncertainty!** The poorer the fitted model, the higher the uncertainty! ***Need to account for this.***

Objective 2: Making Predictions

Remember…

- Using simple model for **mean test performance**
   predictions represent ***expectations*** of what mean test performance will be for a future observation
- Don't forget about the errors ~ predictions will have **uncertainty!** The poorer the fitted model, the higher the uncertainty! ***Need to account for this.***

Aside: Some models will allow prediction of other features of distributions (e.g., the 95th percentile), with uncertainty

What's Next?

- How to compute those parameter estimates when fitting models to dependent variables
- How to test hypotheses, form confidence intervals, make inferences, and make predictions.
- *Always* need to assess the quality of model fit!
- Discuss different schools of thought about model-based inference

Frequentist Inference versus Bayesian Inference