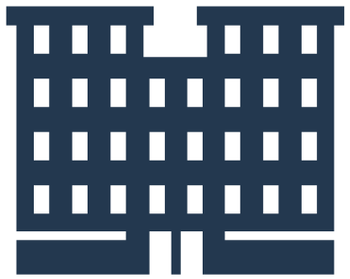
Agenda

1. Visualizing predictions in R (lab from Tuesday)
2. Parsimony and Occam's razor
3. Illustrating overfitting with test and training data
4. Information criteria as formal measures of (over)fit
5. Review of linear model building
6. Comparing criteria in R

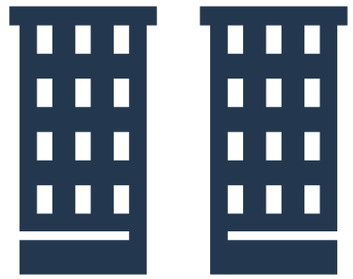
Occam's razor

How many buildings?

Occam's razor



M_1: Four buildings



M_2: Five buildings

$$\frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(M_1)}{\Pr(M_2)} \frac{\Pr(D|M_1)}{\Pr(D|M_2)}$$

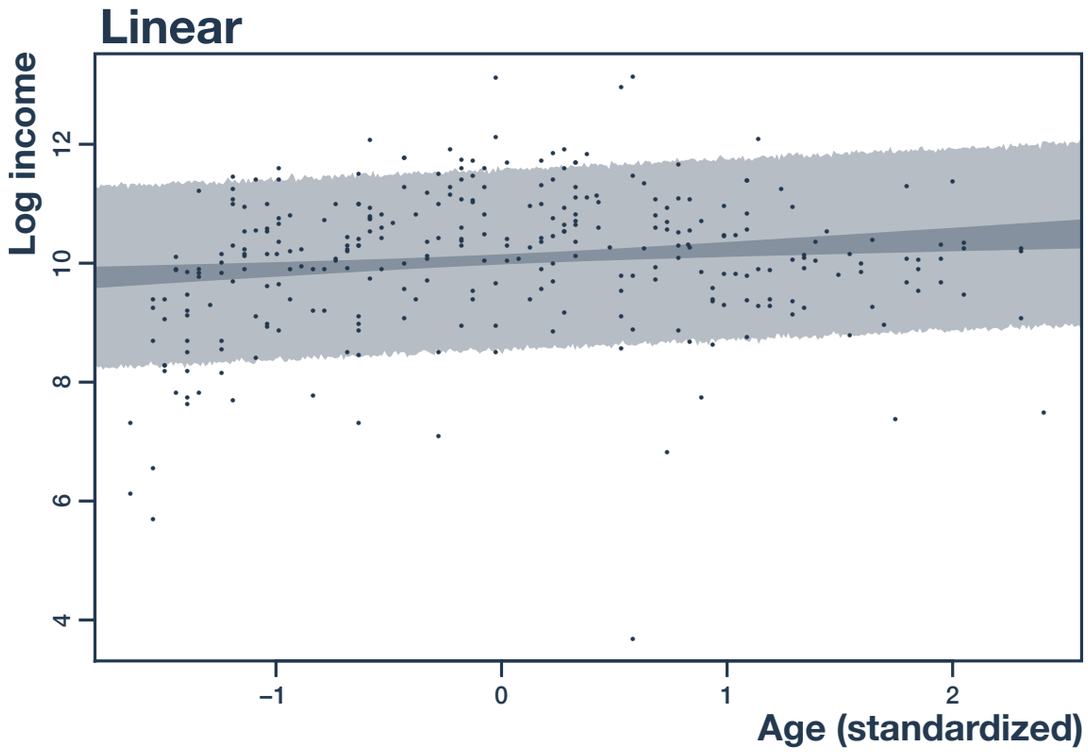
A priori | Simpler models are easier to interpret or more compelling

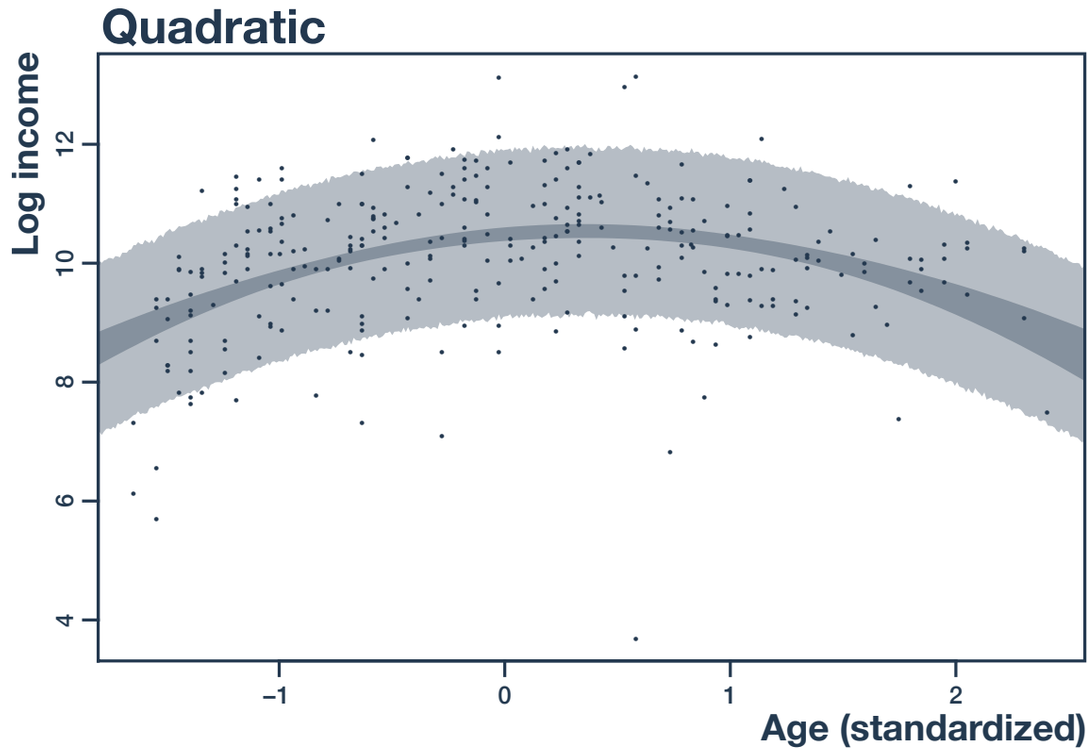
$$\frac{\Pr(M_1)}{\Pr(M_2)} > 1$$

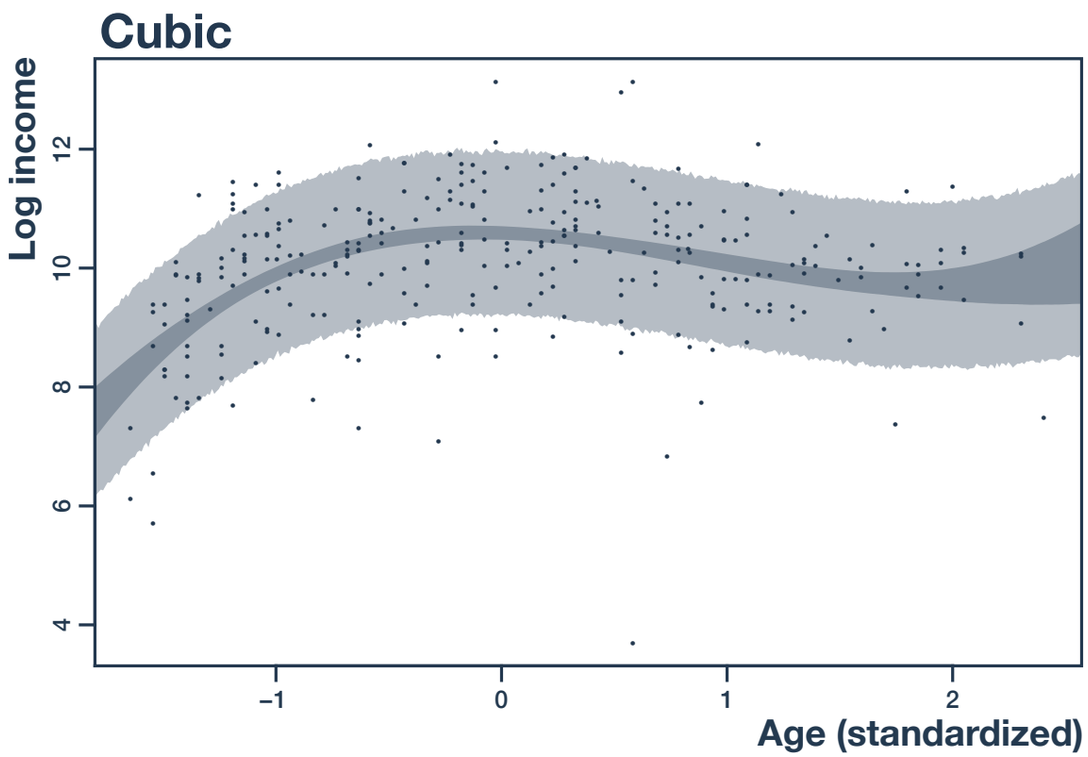
Model likelihood | Simpler models rely less on coincidence

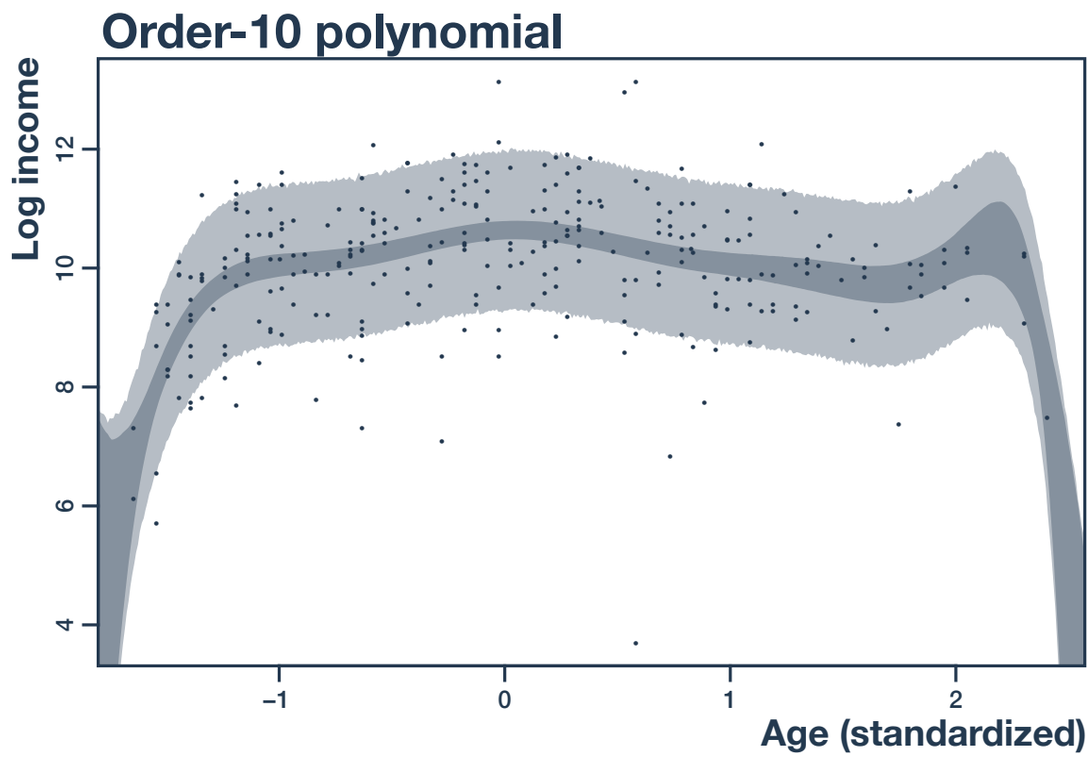
$$\frac{\Pr(D|M_1)}{\Pr(D|M_2)} > 1$$

Overfitting



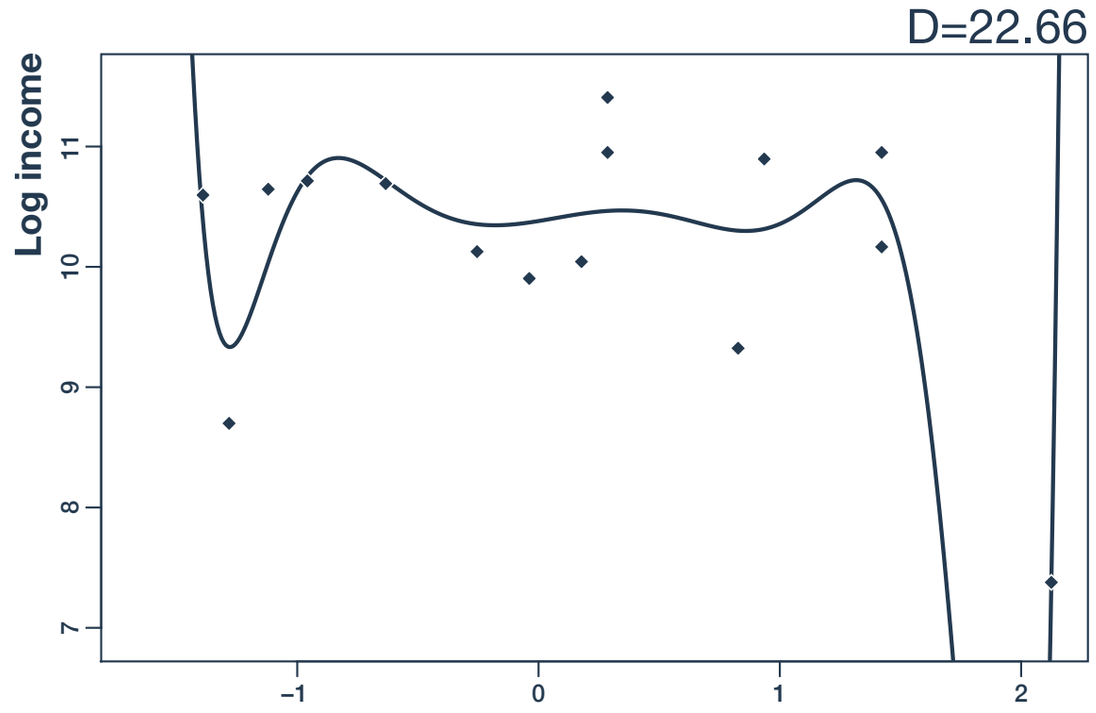




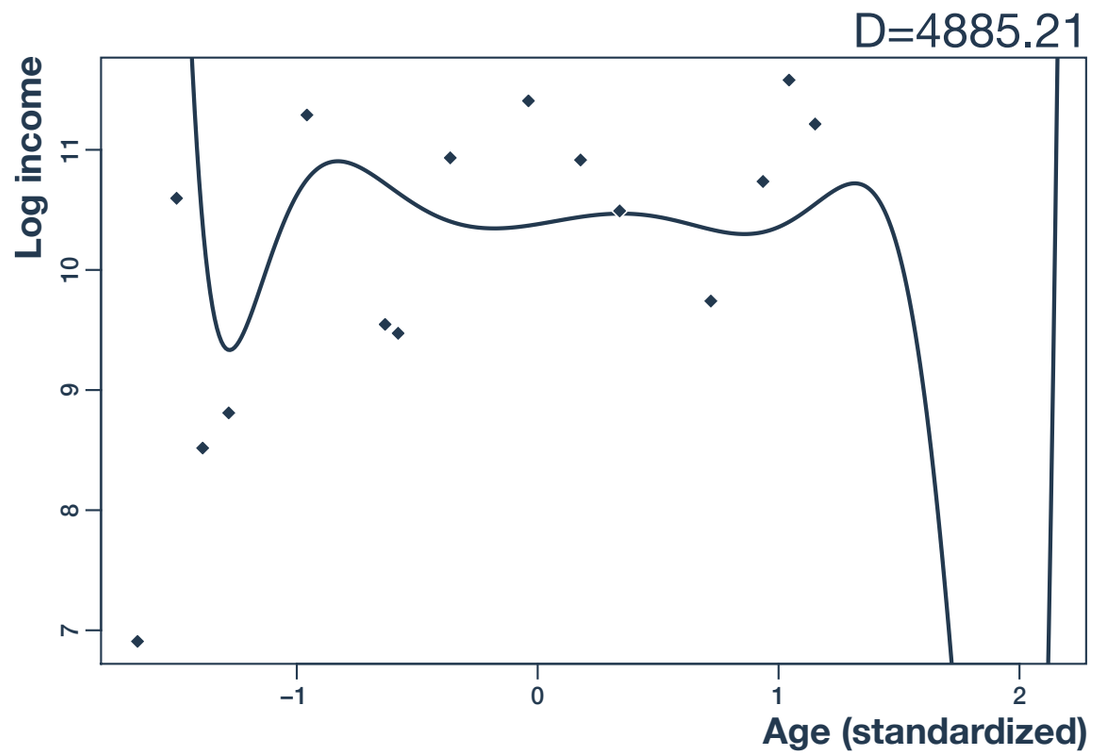


Test and training data

Training data | Fit model on half of the data.



Test data | Assess fit on the other half of the data.



Akaike information criterion

$$\mathrm{D} = -2\log\left(\Pr(\theta|D)\Pr(\theta)\right)$$

$$\begin{aligned}\mathrm{AIC} &= -2\log\left(\Pr(\theta|D)\Pr(\theta)\right) + 2k \\ &= D + 2k\end{aligned}$$

Interpretation 1 | Penalize deviance score for added parameters by some reasonable value.

Interpretation 2 | Model the average difference in deviance between training and test data.

Sample size » number of parameters (k)
Priors have minimal influence (flat or lots of data)
Posterior is approximately (multivariate) normal

Akaike information criterion

$$\mathrm{AIC} = \boxed{-2\log\big(\Pr(\theta|D)\Pr(\theta)\big)} + \boxed{2k}$$

Measure of model fit

Penalty for model complexity

Information criteria

Criterion	Fit	Penalty
Akaike Information Criterion (AIC)	Deviance at MAP estimate	Number of parameters
“Bayesian” Information Criterion (BIC)	Deviance at MAP estimate	#parameters times log(#observations)
Deviance Information Criterion (DIC)	Deviance averaged across posterior	“Effective” #parameters (posterior)
Widely Applicable Information Criterion (WAIC)	Deviance averaged across observations	“Effective” #parameters (posterior and obs.)

Using information criteria

Strategy 1 | Pick the model with the lowest value.

WAIC(M_1) = 209.0; WAIC(M_2) = 208.1
M_2 is the winner

Strategy 2 | Report several models along with values.

Multi-model table showing estimates for different combinations of coefficients, along with WAIC

Strategy 3 | Average predictions across models.

Simultaneous posterior predictions from all models, weighted by WAIC

Building linear models

Considerations when choosing covariates

Theoretical relevance

Independent variables chosen address theoretical concerns

Test theoretical predictions, account for theorized connections

Causal inference

Independent variables chosen to make robust causal claims

Worry about including confounders, omitting colliders, and thinking through role of moderating and mediating variables

Predictive accuracy

Independent variables chosen to maximize predictive power

Accuracy of out-of-sample predictions, interpretation of models with many moving parts