Goodness of Fit and Model Selection

Isaac Sasson
Population Research Center
University of Texas at Austin
Definitions
Types of fit indices
Linear models ($R^2$, F-test for lack of fit)
Logistic regression (pseudo-$R^2$, $\chi^2$ and Deviance goodness of fit tests)
Other complex models (information criteria)
Recommendations
Model selection: choosing between competing models for the same data

Goodness of fit: “how well the model fits the data” (including fixed and random effects, assumptions about covariance structure, etc.)
Types of Fit Indices: Distinction #1

- **Absolute fit**: based on correlation between observed and predicted values ("proportion of explained variation")

- **Incremental/comparative/relative fit**: only useful in comparison to other models

- **Predictive fit**: measure of accuracy in predicting future data
Types of Fit Indices: Distinction #2

- Based on statistics = allows formal tests
- Not based on statistics = rules of thumb
Linear Models
Coefficient of Multiple Determination ($R^2$)

Proportion of variance in $Y$ explained by $X$s in linear regression ("absolute fit")

\[ R^2 = 1 - \frac{SSE}{SSTO} \]

However, high values of $R^2$ do not necessarily imply accurate prediction of new data.
Adjusted-$R^2$

“Theoretical statisticians have recognized for some time however that the multiple correlation coefficient [...] is apt to be deceptively large due to chance factors.”

- R. J. Wherry (1931:440)

Solution:  

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-p}\right) \frac{SSE}{SSTO} \leq R^2$$

More formally, *nested* linear models can be tested for better fit, assuming independent observations

\[ H_0 : \beta_1 = \beta_2 = \beta_3 \]

\[ H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \]

(Recall 2\textsuperscript{nd} semester statistics)
Generalized Linear Models (logistic regression)
Pseudo-$R^2$ (Mcfadden’s form)

Likelihood-based (assumes independence)

$$Psuedo - R^2 = 1 - \frac{\log L(Full)}{\log L(Intercept)}$$

“We do not intend that our pseudo $R^2$ should be reported in formal write-ups of results.”

- Stata website

Source: http://www.rasmusen.org/x/2005/10/17/r2-and-pseudo-r2/
Goodness of Fit Tests (relative)

1. Peasron’s $\chi^2$ test:  
   $$\chi^2 = \sum \frac{(O - E)^2}{E}$$

2. Deviance:  
   $$D = -2[\log L(R) - \log L(F)]$$

3. Hosmer-Lemeshow test (unlike previous tests is valid with continuous covariates)
Complex Models
Other Complex Models

What to do with complex models (hierarchical, random-effects, autocorrelation, etc.)?

Answer: likelihood ratio tests and information criteria
General Likelihood Ratio Test

- Similar to Deviance test
- Follows $\chi^2$ distribution
- Appropriate only for nested models

$$D = -2[\log L(R) - \log L(F)]$$

But what if models are not-nested?
Information Criteria

- Likelihood-based, but not a statistic!
- “Smaller is better”
- Models need not be nested, but based on the same distribution family and same sample size
- AIC emphasizes predictive accuracy, while BIC penalizes additional parameters more strongly

\[
AIC = -2 \log(L) + 2p
\]

\[
BIC = -2 \log(L) + p \log(n)
\]
Recommendations

- Report absolute fit indices when available
- Report relative fit indices only when comparing multiple models
- Be wary of software output (recall underlying assumptions)
- Distinguish between statistics (i.e., formal tests) and non-statistics
- “Break any of these rules sooner than say anything outright barbarous” (Orwell)