1 Automated Model Selection Procedures

- If there are $\boldsymbol{P-1}$ total predictors available for use in a model, then there will be a total of $\mathbf{2}^{\boldsymbol{P-1}}$ possible models (*including* the model with just an intercept).

  Thus the number of models to compare can be *very* large. ***Automated model selection procedures*** are tools for navigating the vast array of models to zero in on a few good ones.

1.1 Best Subset Procedures

- If P isn't too large, we can fit **all** $\mathbf{2}^{\boldsymbol{P-1}}$ models and select a good one based on their values of R_a^2, AIC, BIC, or PRESS (Class Notes 19).

  A model selection procedure that examines **all possible models** (i.e. models corresponding to *all possible subsets* of the predictors) is called a ***best subset procedure***.

  Software output displays the *best* one-predictor model, the *best* two-predictor model, the *best* three-predictor model, etc.

1.2 Stepwise Procedures

- When P is large, the best subset procedure isn't feasible – there are just too many models to be able to fit them all.

- ***Stepwise procedures*** fit a **sequence of models**, at each step either adding or dropping a predictor to improve the model, and terminating when neither adding nor dropping a predictor gives a better model. We'll look at a few **stepwise procedures**:

  1. Backward elimination
  2. Forward selection
  3. Backward stepwise regression
  4. Forward stepwise regression

- As described below, the procedures use the **p-values** for the $\boldsymbol{t}$ **tests** of

$$\begin{aligned} H_0 : \beta_k &= 0 \\ H_a : \beta_k &\neq 0 \end{aligned}$$

  to decide which predictor to add or drop from the model at each step.

They can also be performed using the **p-values** for the **partial $\boldsymbol{F}$ tests** or the changes in **AIC**, **SSE**, or $\boldsymbol{R_a^2}$ to decide which predictor to add or drop.

1.2.1 Backward Elimination and Forward Selection

- To carry out the ***backward elimination procedure***:

  1. Choose a significance level $\boldsymbol{\alpha}$.
  2. Fit the model with **all** $\boldsymbol{P-1}$ predictors. The predictor that has the **largest p-value** is the candidate for dropping from the model.

     If this p-value is greater than α, this predictor is selected to be dropped. Otherwise conclude that none of the predictors should be dropped and terminate the procedure.
  3. Assume X_7 is the predictor selected to be dropped in the previous step. Now **refit** the model with X_7 **excluded**.

     The predictor in this new model that has the **largest p-value** is the next candidate for dropping from the model.

     If this p-value is greater than α, this predictor is selected to be dropped. Otherwise conclude that none of the remaining predictors should be dropped and terminate the procedure.
  4. The elimination procedure is repeated until no more predictors can be dropped from the model.

     At termination, **all** of the predictors in the **model** will have **p-values less** than $\boldsymbol{\alpha}$.

- The **forward selection procedure** is like *backward elimination*, but predictors are **added** to the model at each step instead of being dropped.

  To carry out the ***forward selection procedure***:

  1. Choose a significance level $\boldsymbol{\alpha}$.
  2. For each of the $\boldsymbol{P-1}$ predictors, fit separate **simple** (one-predictor) linear regression models. The predictor that has the **smallest p-value** is the candidate for adding to the model (that just has an intercept).

     If this p-value is less than α, this predictor is added. Otherwise conclude that none of the predictors should be added and terminate the procedure.
  3. Assume X_7 is the predictor added in the previous step, so the current model **includes** only X_7.

Now fit all models with **two predictors**, X_7 being one of the two. For each fit, determine the p-value for the predictor X_k that was added to the model that already included X_7. The X_k whose **p-value** is **smallest** is the candidate for adding.

If the p-value for this X_k is less than α, add it to the model. Otherwise terminate the procedure with the model that just includes X_7.

4. The selection procedure is repeated until no more predictors can be added to the model.

   At termination, **all** of the predictors **not** in the **model** would have **p-values greater** than $\boldsymbol{\alpha}$ if they were added to the model.

1.2.2 Forward Stepwise Regression

- The **forward stepwise procedure** combines *forward selection* with *backward elimination*.

To carry out the ***forward stepwise procedure***:

1. Choose a value, call it $\boldsymbol{\alpha_{\text{in}}}$, for which a predictor will only be **added** to the model if its **p-value** is **less** than $\boldsymbol{\alpha_{\text{in}}}$.

   Choose another value, call it $\boldsymbol{\alpha_{\text{out}}}$, for which a predictor will only be **dropped** from the model if its **p-value** is **greater** than $\boldsymbol{\alpha_{\text{out}}}$.
2. For each of the $\boldsymbol{P-1}$ predictors, fit separate **simple** (one-predictor) linear regression models. The predictor that has the **smallest p-value** is the candidate for adding to the model (that just has an intercept).

   If this p-value is less than $\boldsymbol{\alpha_{\text{in}}}$, this predictor is added. Otherwise conclude that none of the predictors should be added and terminate the procedure.
3. Assume X_7 is the predictor added in the previous step, so the current model **includes** only X_7.

   Now fit all models with **two predictors**, X_7 being one of the two. For each fit, determine the p-value for the predictor X_k that was added to the model that already included X_7. The X_k whose **p-value** is **smallest** is the candidate for adding.

   If the p-value for this X_k is less than $\boldsymbol{\alpha_{\text{in}}}$, add it to the model. Otherwise terminate the procedure with the model that just includes X_7.
4. Suppose X_3 is added to the model in the previous step. Now decide whether any of the *other* predictors already in the model (only X_7 at this point) can be

dropped. Of these other predictors, the one with the **largest p-value** is the candidate for being dropped.

If this p-value is greater than $\boldsymbol{\alpha}_{\mathbf{out}}$, this predictor is dropped from the model, otherwise it is retained.

5. Suppose that X_7 is retained in the previous step, so that both X_3 and X_7 are included in the model.

   Now fit all models with **three predictors** such that X_3 and X_7 are two of the three. For each fit, determine the p-value for the predictor X_k that was just added to the model. The X_k whose **p-value** is **smallest** is the candidate for adding.

   If the p-value for this X_k is less than $\boldsymbol{\alpha}_{\mathbf{in}}$, add it to the model. Otherwise terminate the procedure with the model that just includes X_3 and X_7.

   If X_k is added to the model, determine whether any of the *other* predictors already in the model (X_3 and X_7) can now be dropped by comparing the largest of their p-values to $\boldsymbol{\alpha}_{\mathbf{out}}$.

6. The stepwise procedure is repeated until no more predictors can be added to or dropped from the model.

   At termination, **all** of the predictors **in** the model will have **p-values less** than $\boldsymbol{\alpha}_{\mathbf{out}}$ and **all** of the predictors **not** in the **model** would have **p-values greater** than $\boldsymbol{\alpha}_{\mathbf{in}}$ if they were added to the model.

1.2.3 Backward Stepwise Regression

- The ***backward stepwise procedure*** is similar to *forward stepwise*, except that the **starting model** is the one with **all** $\boldsymbol{P-1}$ predictors. At each step you either drop one of the predictors that's already in the model or add one that's not in it. This is repeated until no more predictors can be added or dropped.

1.2.4 Some Comments on Stepwise Regression

- There's no guarantee that the forward and backward procedures will terminate on the same model, or that they'll identify the same model as the best subset procedure.

- **Recommendation**: Use a stepwise procedure to decide **how many** predictors to include, then compare **all** models with that many predictors using R_a^2, AIC, BIC, or PRESS.