The "theory" of backward elimination

$Y =$ response, $V_1,...,V_m$ pool of predictors

at a given stage:

$(V_1,...,V_k) =$ current subset (notational simplicity)

(I) Fit the linear model for $Y$ on $V_1,...,V_k$

$\rightarrow \quad b_j, s_b^2 \quad j=1,...,t$

(II) For each variable $V_j$ in the current subset

compute $\quad F_j = \frac{b_j^2}{s_b^2}\left(\frac{b_j}{s_b}\right)^2$

we know this equals

$\frac{SSR(V_j \mid V_1,\ldots,V_{j-1})}{MSE(V_1,\ldots,V_{j-1})}$

$\rightarrow$ contribution of $V_j$ after all others $V_1,...,V_{j-1}$

i.e. less one would have dropping $V_j$ from the current subset

$\downarrow$ estimate of the error variance obtained with the current subset
take the smallest $F_j$, say $F_{j*}$ and compare it with a cut-off level (e.g. 4)

- if $F_{j*} < F$ to remove
  
  THEN eliminate $V_{j*}$ from the current subset and restart from (i)

- ELSE ($F_{j*} \geq F$ to remove)
  
  stop, with the current subset

We are using a cut-off level instead of rigorously performing $F$-testing based on normality... but the logic is very similar

**STARTING POINT:**
Assuming all the variables in the pool one of interest, one starts with $(V_1, \ldots, V_m)$

(one can also specify a subcollection of the pool that should never be removed—forced into
the model—see later.)
Next stage

Current subset

Drop

Starting point

Pool of URLs

Stage

\[
PS_1 = \frac{HSE(V, V \cap \text{current})}{SSR(V, V \cap \text{current})}
\]

Backward Elimination
The "theory" of forward selection

(V_{t+1}, ... , V_m) current subset
(V_t, ... , V_{t+1}) the "complement" to the poor

(I) Fit the k-t linear models for

Y on V_{t+1}, V_{t+2}, ... , V_{t+m} \rightarrow SSR(V_{t+1}, V_{t+2}, ... , V_{t+m}), SSE(V_{t+1}, V_{t+2}, ... , V_{t+m})

From the previous stage will have SSR(V_{t+1})

(II) For each variable in the complement compute

F_j = \frac{SSR(V_{t+1}, V_{t+2}, ... , V_t) - SSR(V_{t+1})}{SSE(V_{t+1}, V_{t+2}, ... , V_t) / (t - (t+1))} = \frac{SSR(V_{t+1}, V_{t+2}, ... , V_t)}{SSE(V_{t+1}, V_{t+2}, ... , V_t) / (n - (t+1))}

contribution of V_{t+1} after V_{t+2}, ... , V_m, i.e. gain one would have enlarging the current subset to contain V_{t+1}

estimate of the error variance obtained with the subset enlarged to contain V_{t+1}

(iii) Take the largest $F_j$, say $F_{j*}$, and compare it with a cut-off level.

- If $F_{j*} > F$ to include, then enlarge the current subset to contain $V_{j*}$ and restart from (I).
- Else ($F_{j*} \leq F$ to include), stop, with the current subset.

Again, not rigorous F-testing but similar logic. Notice that here at each stage we don't fit one model, but as many models as there are variables in the complement of the current subset.

**Starting Point:** can be the empty subset...

But if there is a subcollection of the ubs that we want to force into the model, we can start from there — going forward, they will never be eliminated.
The "theory" of stepwise selection

\[ V_1, \ldots, V_m \text{ pool} \]
\[ V_{i1}, \ldots, V_{tf} \text{ subcollection of vbls that we want to force into the model} \]

At a given stage

\[ (V_1, \ldots, V_{tf}, \ldots, V_t) \ (V_{t+1}, \ldots, V_m) \]

current subspace complement

**STEP 1**: backward elimination on \( (V_1, \ldots, V_{tf}, \ldots, V_t) \) without touching the "forced" variables

1. i) fit \( Y \) on \( (V_1, \ldots, V_{tf}, \ldots, V_t) \)
2. ii) compute \( F_j = \frac{SSR(Y, V_{0, \ldots, j})}{MSR(Y, , e=1 \ldots t)} \)
3. iii) if \( \text{min } F_j = F_{j*} < F \) to remove, take 0 out \( V_{j*} \)

else, stay with the subset as it is

outcome \( (V_1, \ldots, V_{tf}, \ldots, V_t) \) indicate with

or, say \( (V_1, \ldots, V_{tf}, \ldots, V_{t+1}) \)
**STEP 2:** Forward selection on \((v_0, v_1, v_2, \ldots, v_k)\) as obtained from step 1.

(2.1) Fit \(y\) on \((v_0, v_1, v_2, \ldots, v_k, v_{k+1})\)

\(y\) on \((v_0, v_1, v_2, \ldots, v_k, v_m)\)

(m-k models in \(k+1\) predictors)

(2.2) Compute \(F_j = \frac{SSR(v_j, \ell = 1 \ldots k)}{MSE(v_0, \ell = 1 \ldots k)}\) \(j = k+1 \ldots m\)

(2.3) If \(max F_j = F_j > F\) to include, include \(v_j\) and restart from step 1, (2.1)

else, stop with the subset obtained from step 1.

Here \(F\) to enter \(> F\) to remove.

As a consequence, if \(v_j\) entered the current subset at a given stage, step 2, it will NOT BE REMOVED IN THE SUCCESSIVE STAGE, step 1.

Nevertheless, it might be removed in any stage after the next... if in the meanwhile, inclusions of other variables have made it redundant.
STARTING POINT:
Here we move both backward and forward. So we can start anywhere (although some "clever" starting points may make our sequential investigation more effective ...)

Of course the starting subset must contain the subcollection of variables that we want to force into the model (if any) -