Forward Selection
- The Forward selection procedure looks to add variables to the model.
- Once added, those variables stay in the model even if they become insignificant at a later step.

Backward Selection
- The Backward selection procedure looks to remove variables from the model.
- Once removed, those variables cannot reenter the model even if they would add significantly at a later step.

Mixed Selection
- A combination of the Forward and Backward selection procedures.
- Starts out like Forward selection but looks to see if an added variable can be removed at a later step.
Mixed - Set up

Stepwise Fit for MDBH
Stepwise Regression Control
Stopping Rule
Prob to Enter: 0.25
Prob to Leave: 0.25
Direction: Mixed

SSE: 19
DFE: 0.7393276
RMSE: 0.0000
RSquare: 0.0000
RSquare Adj: 0.0000
Cp: 48.35699
AICc: 49.64257
BIC: 49.64257

Current Estimates
Step Parameter Action "Sig Prob" Seq SS Rsquare Cp p AICc BIC

Step History
Step Parameter Action "Sig Prob" Seq SS Rsquare Cp p AICc BIC

Stepwise Regression Control
- Direction - Mixed
  - Prob to Enter - controls what variables are added.
  - Prob to Leave - controls what variables are removed.
  - Prob to Enter = Prob to Leave

Current Estimates
- The current estimates are exactly the same as with the Forward selection procedure.
- Clicking on Step will initiate the Mixed procedure that starts like the Forward procedure.
Current Estimates – Step 1

- $X_3$ is added to the model
- Predicted MDBH = 3.896 + 32.937$X_3$
- $R^2=0.7063$
- $RMSE = \sqrt{\text{MSE}} = \sqrt{3.0502499/18} = 0.4117$

Current Estimates – Step 1

- By clicking on Step you will invoke the Backward part of the Mixed procedure.
- Because $X_3$ is statistically significant and is the only variable in the model, clicking on Step will not do anything.
**Stat 301– Lecture 27**

**Current Estimates – Step 1**
- Of the remaining variables not in the model, $X_1$ will add the largest sum of squares if added to the model.
  - $SS = 1.000$
  - "F Ratio" = 8.294
  - "Prob>F" = 0.0104

**JMP Mixed – Step 2**
- Because $X_1$ will add the largest sum of squares and that addition is statistically significant, by clicking on Step, JMP will add $X_1$ to the model with $X_3$.

**Stepwise Fit for MDBH**

**Stepwise Regression Control**
- Stepping Rule: "Enter/Remove Threshold"
- Prob to Enter: 0.25
- Prob to Leave: 0.25

**Current Estimates**

<table>
<thead>
<tr>
<th>Step</th>
<th>Parameter</th>
<th>Action</th>
<th>&quot;Sig Prod&quot;</th>
<th>Beta</th>
<th>95% CI</th>
<th>p</th>
<th>AIC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>Entered</td>
<td>0.000</td>
<td>7.156</td>
<td>6.916 - 7.396</td>
<td>0.000</td>
<td>30.383</td>
<td>30.145</td>
</tr>
<tr>
<td>2</td>
<td>$X_1$</td>
<td>Entered</td>
<td>0.000</td>
<td>0.013</td>
<td>0.0000 - 0.026</td>
<td>0.500</td>
<td>30.383</td>
<td>30.145</td>
</tr>
</tbody>
</table>

**Step History**

- Step 1: Entered $X_1$
- Step 2: Entered $X_2$

- SSE: 2.000000
- DFE: 17
- RMSE: 0.8030
- R-Square: 0.779
- R-Square Adj: 0.778
- Cp: 3
- AIC: 21.8472
- BIC: 23.1834

- SS: 0.000
- "F Ratio": 8.294
- "Prob>F": 0.0104
Current Estimates – Step 2

- $X_1$ is added to the model
- Predicted MDBH = 3.143 + 0.0314*$X_1$ + 22.954*$X_3$
  - $R^2=0.8026$
  - $\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{2.0500859/17} = 0.3473$

By clicking on Step you will invoke the Backward part of the Mixed procedure.

Because $X_3$ and $X_1$ are statistically significant, clicking on Step will not do anything.

Of the remaining variables not in the model $X_2$ will add the largest sum of squares if added to the model.

- $SS = 0.671$
- "F Ratio" = 7.784
- "Prob>F" = 0.0131
JMP Mixed – Step 3

Because $X_2$ will add the largest sum of squares and that addition is statistically significant, by clicking on Step, JMP will add $X_2$ to the model with $X_3$ and $X_1$.

### JMP Mixed – Step 3

- $X_2$ is added to the model
- Predicted $\text{MDBH} = 3.236 + 0.0974 \times X_1 - 0.000169 \times X_2 + 3.467 \times X_3$
- $R^2 = 0.8672$
- $\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{1.3791191/16} = 0.2936
Current Estimates – Step 3

- By clicking on Step you will invoke the Backward part of the Mixed procedure.
- Note that variable X₃ is no longer statistically significant and so it will be removed from the model when you click on Step.

Stopping Rule:

<table>
<thead>
<tr>
<th>P-value Threshold</th>
<th>Prob to Enter</th>
<th>Prob to Leave</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

Direction: Mixed

Stepwise Regression Control

<table>
<thead>
<tr>
<th>SS</th>
<th>DFE</th>
<th>RMSE</th>
<th>R²</th>
<th>R² Adjusted</th>
<th>Cp</th>
<th>s</th>
<th>AICc</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3938931</td>
<td>17</td>
<td>0.8658</td>
<td>0.8500</td>
<td>14.15158</td>
<td>15.46784</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Current Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Entered</th>
<th>Estimate</th>
<th>t-stat</th>
<th>p-value</th>
<th>SS</th>
<th>Adj R²</th>
<th>Cp</th>
<th>s</th>
<th>AICc</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Entered</td>
<td>3.26051366</td>
<td>1</td>
<td>1</td>
<td>0.000</td>
<td>19.388</td>
<td>2</td>
<td>3</td>
<td>28.1345</td>
<td>26.6473</td>
</tr>
<tr>
<td>X₁</td>
<td>Entered</td>
<td>0.10691347</td>
<td>1.32e-8</td>
<td>0.0104</td>
<td>0.8026</td>
<td>9.7843</td>
<td>3</td>
<td>4</td>
<td>23.1835</td>
<td>21.8672</td>
</tr>
<tr>
<td>X₂</td>
<td>Entered</td>
<td>-0.0001898</td>
<td>0.6844</td>
<td>0.0131</td>
<td>0.8672</td>
<td>4</td>
<td>0.67597</td>
<td>4</td>
<td>18.2505</td>
<td>17.5575</td>
</tr>
<tr>
<td>X₃</td>
<td>Removed</td>
<td>0.00015884</td>
<td>0.000</td>
<td>0.6844</td>
<td>0.8658</td>
<td>2.1714023</td>
<td>3</td>
<td>5</td>
<td>15.46784</td>
<td>14.15158</td>
</tr>
</tbody>
</table>

Step History

<table>
<thead>
<tr>
<th>Step</th>
<th>Parameter</th>
<th>Action</th>
<th>Estimate</th>
<th>t-stat</th>
<th>p-value</th>
<th>SS</th>
<th>Adj R²</th>
<th>Cp</th>
<th>s</th>
<th>AICc</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>Entered</td>
<td>3.26051366</td>
<td>1</td>
<td>1</td>
<td>0.000</td>
<td>19.388</td>
<td>2</td>
<td>3</td>
<td>28.1345</td>
<td>26.6473</td>
</tr>
<tr>
<td>2</td>
<td>X₁</td>
<td>Entered</td>
<td>0.10691347</td>
<td>1.32e-8</td>
<td>0.0104</td>
<td>0.8026</td>
<td>9.7843</td>
<td>3</td>
<td>4</td>
<td>23.1835</td>
<td>21.8672</td>
</tr>
<tr>
<td>3</td>
<td>X₂</td>
<td>Entered</td>
<td>-0.0001898</td>
<td>0.6844</td>
<td>0.0131</td>
<td>0.8672</td>
<td>4</td>
<td>0.67597</td>
<td>4</td>
<td>18.2505</td>
<td>17.5575</td>
</tr>
<tr>
<td>4</td>
<td>X₃</td>
<td>Removed</td>
<td>0.00015884</td>
<td>0.000</td>
<td>0.6844</td>
<td>0.8658</td>
<td>2.1714023</td>
<td>3</td>
<td>5</td>
<td>15.46784</td>
<td>14.15158</td>
</tr>
</tbody>
</table>

Current Estimates – Step 4

- X₃ is removed from the model
- Predicted MDBH = 3.2605 + 0.1069*X₁ – 0.0001898*X₂
- R²=0.8658
- RMSE = \sqrt{MSE} = \sqrt{1.3938931/17} = 0.2863
Current Estimates - Step 4

- Because $X_1$ and $X_2$ add significantly to the model they cannot be removed.
- Because $X_3$ will not add significantly to the model it cannot be added.
- The Mixed procedure stops.

---

Response MDBH

<table>
<thead>
<tr>
<th>Summary of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSquare</td>
</tr>
<tr>
<td>RSquare Adj</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>Mean of Response</td>
</tr>
<tr>
<td>Observations (or Sum Wgts)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>C. Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
</tbody>
</table>

Finding the “best” model

- For this example, the Forward selection procedure did not find the “best” model.
- The Backward and Mixed selection procedures came up with the “best” model.
Finding the “best” model
- None of the automatic selection procedures are guaranteed to find the “best” model.
- The only way to be sure, is to look at all possible models.

All Possible Models
- For $k$ explanatory variables there are $2^k - 1$ possible models.
- There are $k$ 1-variable models.
- There are $\binom{k}{2}$ 2-variable models.
- There are $\binom{k}{3}$ 3-variable models.

All Possible Models
- When confronted with all possible models, we often rely on summary statistics to describe features of the models.
- $R^2$: Larger is better
- adj$R^2$: Larger is better
- RMSE: Smaller is better
All Possible Models

Another summary statistic used to assess the fit of a model is Mallows $C_p$.

$$C_p = \left( \frac{SSE}{MSE_{full}} \right) - (n - 2p); \quad p = k + 1$$

MDBH Example

- Model with $X_1$ and $X_2$ ($p=3$)
- $SSE_p=1.3938931$
- $MSE_{Full}=0.08619$

$$C_p = \left( \frac{1.3938931}{0.0861949} \right) - (20 - 6) = 16.1714 - 14 = 2.1714$$

All Possible Models

- The smaller $C_p$ is the “better” the fit of the model.
- The full model will have $C_p=p$. 
All Possible Models

There are several criteria that incorporate the maximized value of the Likelihood function, \( L \). Which gives the probability of getting the sample data given the best estimates of the parameters in the model.

\[ AIC = 2p - 2\ln(L) \]
\[ AICc = AIC + \frac{2p(p+1)}{n - p - 1} \]

The smaller the AICc the “better” the fit of the model.

Another summary statistic is the Bayesian Information Criterion or BIC.

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

The smaller the BIC the better the fit of the model.
JMP – Fit Model

- Personality – Stepwise – Run
- Red triangle pull down next to Stepwise Fit
  - All Possible Models
  - Maximum number of terms: 3
  - Number of best models: 3

All Possible Models

- Right click on table – Columns
- Check Cp

<table>
<thead>
<tr>
<th>Model</th>
<th>Number</th>
<th>R-squared</th>
<th>RSS</th>
<th>AICc</th>
<th>Cp</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>1</td>
<td>0.7063</td>
<td>0.4117</td>
<td>26.6473</td>
<td>0.1098</td>
<td>0.0</td>
</tr>
<tr>
<td>X1</td>
<td>1</td>
<td>0.5977</td>
<td>0.4818</td>
<td>32.9417</td>
<td>0.2000</td>
<td>0.0</td>
</tr>
<tr>
<td>X3</td>
<td>1</td>
<td>0.0593</td>
<td>0.7367</td>
<td>49.9281</td>
<td>0.3000</td>
<td>0.0</td>
</tr>
<tr>
<td>X1,X2</td>
<td>2</td>
<td>0.8658</td>
<td>0.2863</td>
<td>14.1516</td>
<td>0.4000</td>
<td>0.0</td>
</tr>
<tr>
<td>X1,X3</td>
<td>2</td>
<td>0.8026</td>
<td>0.3473</td>
<td>21.8672</td>
<td>0.5000</td>
<td>0.0</td>
</tr>
<tr>
<td>X2,X3</td>
<td>2</td>
<td>0.7451</td>
<td>0.3946</td>
<td>26.9777</td>
<td>0.6000</td>
<td>0.0</td>
</tr>
<tr>
<td>X1,X2,X3</td>
<td>3</td>
<td>0.8672</td>
<td>0.2936</td>
<td>17.5575</td>
<td>0.7000</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Ordered up to best 3 models up to 3 terms per model.
**All Possible Models**

- Lists all 7 models.
- 1-variable models – listed in order of the $R^2$ value.
- 2-variable models – listed in order of the $R^2$ value.
- 3-variable (full) model.

---

**All Possible Models**

- Model with $X_1$, $X_2$, $X_3$ – Highest $R^2$ value.
- Model with $X_1$, $X_2$ – Lowest RMSE, Lowest AICc, Lowest BIC, and lowest $C_p$.

---

**Best Model**

- Which is “best”?
- According to our definition of “best” we can’t tell until we look at the significance of the individual variables in the model.