


# Stat 401 B – Lecture 25



## 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.

1

---

---

---


---

---

---

---

---



## How does JMP do this?

- Analysis – Fit Model
- Enter the response variable in the Pick Role Variables box as Y.
- Add all the explanatory variables to the Construct Model Effects box.

2

---

---

---


---

---

---

---

---



## JMP – Fit Model

- Make the Personality - Stepwise.
- Click on Run Model.

3

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 25

## Forward – Set up

Stepwise Fit  
Response: MDBH

Stepwise Regression Control  
Prob to Enter 0.250  
Prob to Leave 0.100  
Direction Forward

Current Estimates

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
10.3855	19	0.5466053	0.0000	0.0000	102.4885	-11.1064

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Intercept	6.265	1	0	0.000	1.0000
<input type="checkbox"/>	<input type="checkbox"/>	X1	0	1	6.207045	26.739	0.0001
<input type="checkbox"/>	<input type="checkbox"/>	X2	0	1	0.616027	1.135	0.3008
<input type="checkbox"/>	<input type="checkbox"/>	X3	0	1	7.335255	43.287	0.0000

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
------	-----------	--------	------------	--------	---------	----	---

4

---

---

---

---

---

---

---

---

## Stepwise Regression Control

- Direction – Forward
- Prob to Enter – the P-value for a variable must be less than or equal to the Prob to Enter in order for the variable to be added to the model.

5

---

---

---

---

---

---

---

---

## Current Estimates

- Only the intercept is included at this point.
- The value of the intercept, 6.265 is the mean response (MDBH).

6

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 25



## Current Estimates

- The SSE, at this point, is actually the C. Total sum of squares.
  - 10.3855
- Under the SS column are the sum of squares that will be explained if that variable is added to the model.

7

---

---

---


---

---

---

---

---



## Current Estimates

- Note the  $X_3$  will add the largest sum of squares if it is added to the model.
  - 7.335
- If  $X_3$  is added the SLR of MDBH on  $X_3$  will have
  - $R^2 = 7.335 / 10.3855 = 0.7063$

8

---

---

---


---

---

---

---

---



## Current Estimates

- Note that adding  $X_3$  will be a statistically significant addition to the model
  - "F-Ratio" = 43.287
  - "Prob>F" = 0.0000 (P-value)
  - The P-value is small.
- Click on Step

9

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 25

**Stepwise Fit**  
Response: MDBH

**Stepwise Regression Control**  
Prob to Enter 0.250  
Prob to Leave 0.100  
Direction Forward

**Current Estimates**

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
3.0502449	18	0.1694581	0.7063	0.6900	19.387747	-33.6102

Lock Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	Intercept	3.8956688	1	0	0.000	1.0000
<input type="checkbox"/>	X1	0	1	1.000159	8.294	0.0104
<input type="checkbox"/>	X2	0	1	0.403296	2.590	0.1259
<input checked="" type="checkbox"/>	X3	32.9371533	1	7.335255	43.287	0.0000

**Step History**

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	X3	Entered	0.0000	7.335255	0.7063	19.388	2

10

---

---

---

---

---

---

---

---

---

---

**Current Estimates – Step 1**

- $X_3$  is added to the model
- Predicted MDBH =  $3.896 + 32.937 \cdot X_3$ 
  - $R^2 = 0.7063$
  - $RMSE = \sqrt{MSE} = \sqrt{0.1694581} = 0.4117$

11

---

---

---

---

---

---

---

---

---

---

**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

12

---

---

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 25

## JMP Forward – 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  $X_3$ .

13

---

---

---

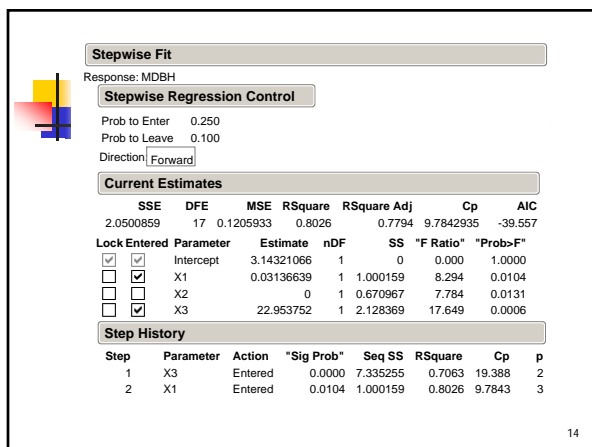
---

---

---

---

---



The screenshot shows the 'Stepwise Fit' dialog box in JMP. The response variable is 'MDBH'. Under 'Stepwise Regression Control', 'Prob to Enter' is 0.250, 'Prob to Leave' is 0.100, and 'Direction' is 'Forward'. The 'Current Estimates' table shows the following values:

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
2.0500859	17	0.1205933	0.8026	0.7794	9.7842935	-39.557

The 'Lock Entered' table shows the following parameters:

Lock Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
<input checked="" type="checkbox"/>	Intercept	3.14321066	1	0	0.000	1.0000
<input checked="" type="checkbox"/>	X1	0.03136639	1	1.000159	8.294	0.0104
<input type="checkbox"/>	X2	0	1	0.670967	7.784	0.0131
<input checked="" type="checkbox"/>	X3	22.953752	1	2.128369	17.649	0.0006

The 'Step History' table shows the following steps:

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	X3	Entered	0.0000	7.335255	0.7063	19.388	2
2	X1	Entered	0.0104	1.000159	0.8026	9.7843	3

14

---

---

---

---

---

---

---

---

## Current Estimates – Step 2

- $X_1$  is added to the model
- Predicted  $MDBH = 3.143 + 0.0314 \cdot X_1 + 22.954 \cdot X_3$ 
  - $R^2 = 0.8026$
  - $RMSE = \sqrt{MSE} = \sqrt{0.1205933} = 0.3473$

15

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 25

## Current Estimates – Step 2

- 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

16

---

---

---

---

---

---

---

---

## JMP Forward – 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  $X_1$ .

17

---

---

---

---

---

---

---

---

**Stepwise Fit**  
Response: MDBH

**Stepwise Regression Control**  
 Prob to Enter 0.250  
 Prob to Leave 0.100  
 Direction Forward

**Current Estimates**

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1.3791191	16	0.0861949	0.8672	0.8423	4	-45.4857

Lock Entered

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	3.23573225	1	0	0.000	1.0000
X1	0.09740562	1	1.26783	14.709	0.0015
X2	-0.0001689	1	0.670967	7.784	0.0131
X3	3.46681347	1	0.014774	0.171	0.6844

**Step History**

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	X3	Entered	0.0000	7.335255	0.7063	19.388	2
2	X1	Entered	0.0104	1.000159	0.8026	9.7843	3
3	X2	Entered	0.0131	0.670967	0.8672	4	4

18

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 25

 **Current Estimates – Step 3**

- $X_2$  is added to the model
- Predicted MDBH =  $3.236 + 0.0974 * X_1 - 0.000169 * X_2 + 3.467 * X_3$ 
  - $R^2 = 0.8672$
  - $RMSE = \sqrt{MSE} = \sqrt{0.0861949} = 0.2936$

19

---

---

---


---

---

---

---

---

 **Current Estimates – Step 3**

- There are no variables remaining and so the forward selection procedure stops.
- Note that variable  $X_3$  is no longer statistically significant.
- The combination of  $X_1$  and  $X_2$  has made  $X_3$  redundant.

20

---

---

---


---

---

---

---

---

 **Backward Selection**

- Start with a full model (a model that contains all of the available explanatory variables).
- Remove variables, one at a time, if they do not add significantly to the model.

21

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 25



## Full Model - MDBH

- Predicted MDBH =  $3.236 + 0.0974 \cdot X_1 - 0.000169 \cdot X_2 + 3.467 \cdot X_3$ 
  - $R^2 = 0.8672$
  - $RMSE = \sqrt{MSE} = \sqrt{0.0861949} = 0.2936$

22

---

---

---


---

---

---

---

---



## Statistical Significance

- $X_1$ :  $F = 14.709$ , P-value = 0.0015
- $X_2$ :  $F = 7.784$ , P-value = 0.0131
- $X_3$ :  $F = 0.171$ , P-value = 0.6844

23

---

---

---


---

---

---

---

---



## Backward Selection – Step 1

- Because the P-value for  $X_3$  is not small, it should be removed from the model.
- Removing  $X_3$  will subtract 0.0148 from the sum of squares model.

24

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 25



**Response MDBH**

**Summary of Fit**

RSquare	0.865785
RSquare Adj	0.849995
Root Mean Square Error	0.286345
Mean of Response	6.265
Observations (or Sum Wgts)	20

**Analysis of Variance**

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	8.991607	4.49580	54.8311
Error	17	1.393893	0.08199	<b>Prob &gt; F</b>
C. Total	19	10.385500		<.0001*

**Parameter Estimates**

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	3.2605137	0.333024	9.79	<.0001*
X1	0.1069135	0.010578	10.11	<.0001*
X2	-0.00019	3.256e-5	-5.83	<.0001*

**Effect Tests**

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
X1	1	1	8.3755800	102.1490	<.0001*
X2	1	1	2.7845615	33.9607	<.0001*

25

---

---

---

---

---


---

---

---

---

---



## Backward Selection

- Because all of the remaining variables in the model are statistically significant, the backward selection procedure stops.

26

---

---

---

---

---


---

---

---

---

---



## Stepwise Regression Control

- Direction – Backward
- Prob to Leave – the P-value for a variable must be greater than the Prob to Leave in order for the variable to be removed from the model.
- Click on Enter All

27

---

---

---

---

---

---

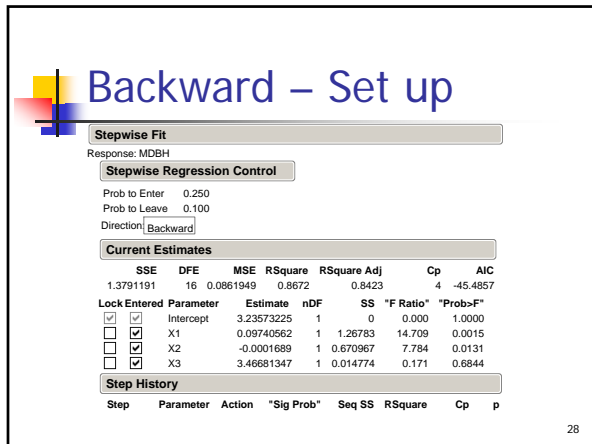
---

---

---

---

# Stat 401 B – Lecture 25



**Stepwise Fit**  
Response: MDBH

**Stepwise Regression Control**  
 Prob to Enter 0.250  
 Prob to Leave 0.100  
 Direction **Backward**

**Current Estimates**

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1.3791191	16	0.0861949	0.8672	0.8423	4	-45.4857

**Lock Entered**

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	3.23573225	1	0	0.000	1.0000
X1	0.09740562	1	1.26783	14.709	0.0015
X2	-0.0001689	1	0.670967	7.784	0.0131
X3	3.46681347	1	0.014774	0.171	0.6844

**Step History**

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p

28

---

---

---

---

---

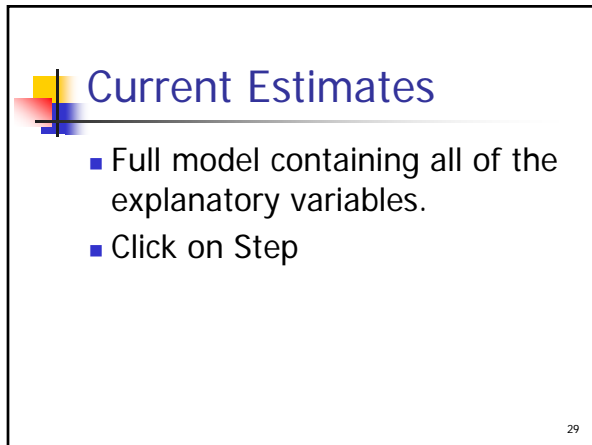
---

---

---

---

---



**Stepwise Fit**  
Response: MDBH

**Stepwise Regression Control**  
 Prob to Enter 0.250  
 Prob to Leave 0.100  
 Direction **Backward**

**Current Estimates**

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1.3938931	17	0.0819937	0.8658	0.8500	2.1714023	-47.2726

**Lock Entered**

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	3.26051366	1	0	0.000	1.0000
X1	0.10691347	1	8.37558	102.149	0.0000
X2	-0.0001898	1	2.784561	33.961	0.0000
X3	0	1	0.014774	0.171	0.6844

**Step History**

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	X3	Removed	0.6844	0.014774	0.8658	2.1714	3

29

---

---

---

---

---

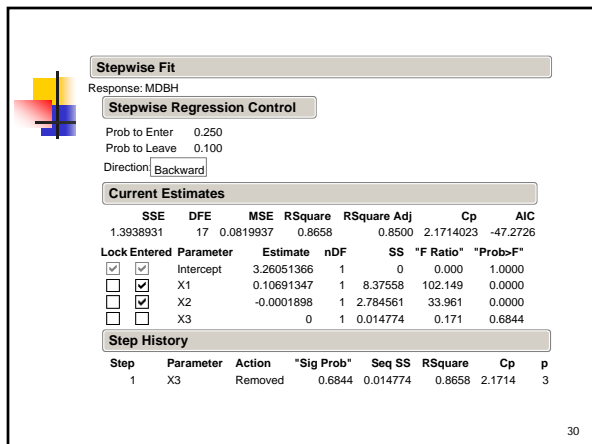
---

---

---

---

---



**Stepwise Fit**  
Response: MDBH

**Stepwise Regression Control**  
 Prob to Enter 0.250  
 Prob to Leave 0.100  
 Direction **Backward**

**Current Estimates**

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1.3938931	17	0.0819937	0.8658	0.8500	2.1714023	-47.2726

**Lock Entered**

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	3.26051366	1	0	0.000	1.0000
X1	0.10691347	1	8.37558	102.149	0.0000
X2	-0.0001898	1	2.784561	33.961	0.0000
X3	0	1	0.014774	0.171	0.6844

**Step History**

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	X3	Removed	0.6844	0.014774	0.8658	2.1714	3

30

---

---

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 25



## Current Estimates – Step 1

- $X_3$  is removed from the model
- Predicted MDBH =  $3.261 + 0.1069 * X_1 - 0.0001898 * X_2$ 
  - $R^2 = 0.8658$
  - $RMSE = \sqrt{MSE} = \sqrt{0.0819937} = 0.2863$

31

---

---

---


---

---

---

---

---



## Current Estimates – Step 1

- All of the remaining variables are statistically significant.
- Clicking on Step will not change anything because no variable can be removed.

32

---

---

---


---

---

---

---

---



## Backward Selection

- Once a variable is removed it can never be entered again, even if it would add significantly to a model later on.

33

---

---

---

---

---

---

---

---