

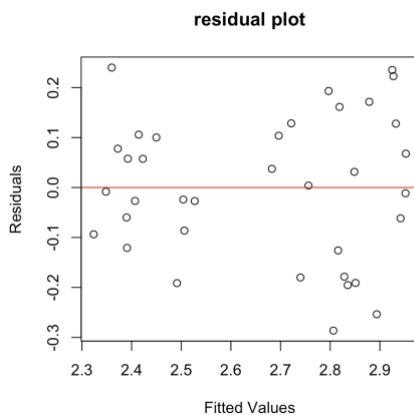
Yunlu Li  
STAT 5120  
Homework 7

Question 1

(a) The predictors giving the best value for all 5 criterions is

<b>price</b>	<b>nielsen</b>	<b>discount</b>	<b>promo</b>	<b>time</b>
TRUE	FALSE	TRUE	TRUE	FALSE

(b) The residuals generally fall in a horizontal band around 0 with no obvious pattern. The variance of the residuals appear to be constant. Thus, assumptions of the regression model are met.



(c) The estimated regression equation is

$$\text{Share} = 3.18527 - 0.35269 * \text{price} + 0.39914 * \text{discount} + 0.11803 * \text{promo}$$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.18527	0.36505	8.726	5.7e-10	***
price	-0.35269	0.15738	-2.241	0.0321	*
discount	0.39914	0.05125	7.787	7.0e-09	***
promo	0.11803	0.05149	2.292	0.0286	*

$\beta_1 = -0.35269$  means that the estimated share decreases by 0.35269 percentage points for every one dollar increases in *price*, for given class of *discount* and *promo*.

$\beta_2 = 0.39914$  means that the estimated share is 0.35269 percentage points higher for discount price in effect than no discount, for given value of *price* and class of *promo*.

$\beta_3 = 0.11803$  means that the estimated share is 0.11803 percentage points higher for promotion in effect than no promotion, for given value of *price* and class of *discount*.

(d) The model giving best smallest AIC<sub>p</sub> is

```
> models[13,]
      price nielsen discount      promo      time
      FALSE   FALSE      TRUE      TRUE      FALSE
```

The model giving best smallest PRESS<sub>p</sub> is

```
> models[6,]
      price nielsen discount      promo      time
      TRUE   FALSE      TRUE      FALSE      FALSE
```

(e) For model 13, the PRESS<sub>p</sub> is 0.9797197 and SSE<sub>p</sub> is 0.8306269.

For model 6, the PRESS<sub>p</sub> is 0.9766704 and SSE<sub>p</sub> is 0.8358599.

The difference between SSE<sub>p</sub> with the PRESS<sub>p</sub> for model 6 with *price*, *discount* is smaller so model 6 is a better fit.

(f) The model based on forward selection consists of *price*, *discount*, and *promo*.

Coefficients:

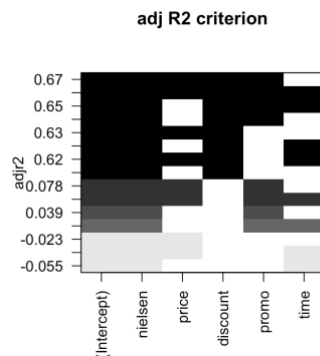
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.18527	0.36505	8.726	5.7e-10	***
discount	0.39914	0.05125	7.787	7.0e-09	***
promo	0.11803	0.05149	2.292	0.0286	*
price	-0.35269	0.15738	-2.241	0.0321	*

(g) The model based on backward selection consists of *price*, *discount*, and *promo*.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.18527	0.36505	8.726	5.7e-10	***
price	-0.35269	0.15738	-2.241	0.0321	*
discount	0.39914	0.05125	7.787	7.0e-09	***
promo	0.11803	0.05149	2.292	0.0286	*

(h) The best model consists of *nielsen*, *price*, *discount*, and *promo*.



## Question 2

(a)

Step 1:  $x_4$

Step 2:  $x_4, x_1$

Step 3:  $x_4, x_1, x_2$

Step 4:  $x_4, x_1, x_3, x_2$

(b)

Step 1:  $x_4, x_1, x_3, x_2$

Step 2:  $x_1, x_2, x_3$

Step 3:  $x_2, x_3$

Step 4:  $x_3$