STATS 2107
Statistical Modelling and Inference II
Lecture notes
Chapter 4: Linear models part 2

Jono Tuke

School of Mathematical Sciences, University of Adelaide

Semester 2 2017

Model selection

Model selection

How do we decide on the best model for our data.

Two parts:

- Choice of algorithm.
- Choice of heuristic.

The forward selection algorithm with P-values

1. Begin with the null model.
2. For every term not currently included in the model, calculate a P-value for the inclusion of that term.
3. If the smallest P-value is less than the threshold p_{in} (usually chosen to be 0.05), add that term to the model.
4. Iterate (2), (3) until no further terms are significant.

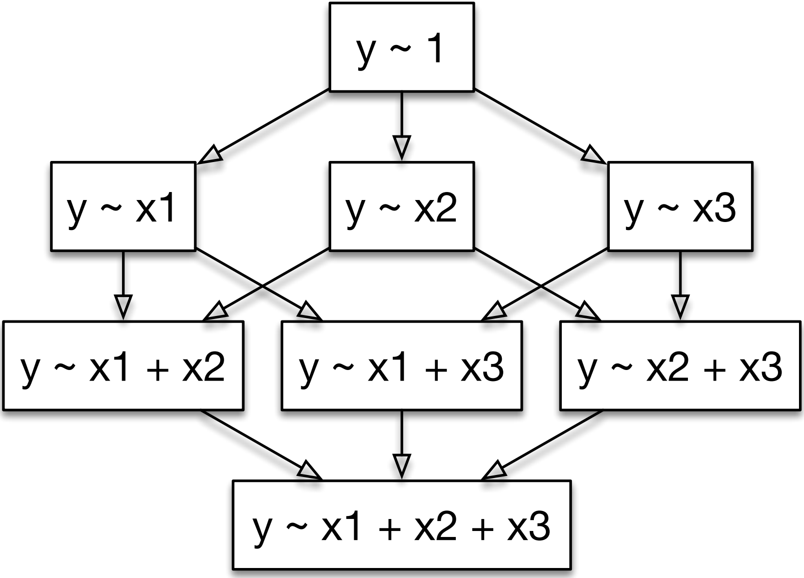
The backwards elimination algorithm with P-values

1. Begin with the most complicated model to be considered.
2. For each term included in the model, calculate the P-value for the removal of that term.
3. If the largest P-value is greater than the threshold p_{out} (usually chosen to be 0.05), remove that term from the model.
4. Iterate (2), (3) until the model contains only significant terms.

The stepwise selection procedure with P-values

1. Begin with the null model.
2. Perform one step of forward selection using a liberal value of p_{in} such as 0.2 or 0.15.
3. Perform one step of backward elimination with a value of P_{out} such as 0.05.
4. Iterate (2), (3) until no further changes occur or the algorithm cycles.

Comparison of methods



Principle of marginality

Whenever an interaction term is included in the model, all implied lower order interactions and main effects must also be included.

For example if we find that we have an interaction term $x_{i1}x_{i2}$ in the model, the we must keep the main effects x_{i1} and x_{i2} in the model.

Akaike information criterion (AIC)

The Akaike information criterion (AIC) is defined as

$$AIC = 2k - 2\ln(\hat{L}),$$

where k is the number of parameters in the model, and $\ln(\hat{L})$ is the log likelihood evaluated at the maximum likelihood estimates. We choose the model with the lowest AIC.

Akaike information criterion corrected (AICc)

To adjust for small sample sizes, the AICc is used. It is defined as

$$AICc = AIC + \frac{2k(k+1)}{n-k-1},$$

where n is the sample size.

Bayesian information criterion (BIC)

A more stingent criterion with respect to the number of parameters is the Bayesian information criterion (BIC). It is defined as

$$BIC = \ln(n)k - 2\ln(\hat{L}).$$

Cross-validation

A useful method to access how good a model is for prediction is cross-validation.

For k-fold cross-validation, you split the data into k parts.

In each step, train the model on $k-1$ parts and test for the kth part.

Cross-validation

Test	Train	Train	Train	Train
Train	Test	Train	Train	Train
Train	Train	Test	Train	Train
Train	Train	Train	Test	Train
Train	Train	Train	Train	Test

Notation

Label each part

$$C_1, C_2, \ldots, C_K$$

Let the number of observations in C_k be n_k. So that

$$n = \sum_{k=1}^{K} n_k,$$

i.e. n is the total number of observations.

Prediction error

The cross validation estimate of the prediction error is

$$CV_{(K)} = \sum_{k=1}^{K} \frac{n_k}{n} MSE_k,$$

where

$$MSE_k = \sum_{i \in C_k} \frac{(y_i - \hat{y}_i)^2}{n_k},$$

where $\hat{y}_i$ is the fitted value for observation i for the model with part k removed.

Polynomial regression and two-sample pooled T-test

Polynomial regression

Consider a regression model of the form

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_r x_i^r + \varepsilon_i,$$

where

$$\varepsilon_i \sim i.i.d. N(0, \sigma^2), i = 1, 2, \ldots, n.$$

Notes

This model is not linear in x_is but it is linear in the coefficients $\beta_0, \beta_1, \ldots, \beta_r$

Formulation

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^r \\ 1 & x_2 & x_2^2 & \dots & x_2^r \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^r \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix}$$

Remarks

- The reason this formulation is valid is that the multiple regression model makes no assumptions about the X matrix other than that its columns must be linearly independent. (It can also be proved that the X matrix will have linearly independent columns provided that r is less than the number of different x-values.)
- Prediction for polynomial regression models may be done in the same way as for general multiple regression.
- Unless there are special reasons for not doing so, whenever x^r is included in the model, we should also include $x, x^2, \ldots, x^{r-1}$.

Interaction terms

Consider a model with two predictors x_{i1} and x_{i2}. We can also consider an interaction between these preditors:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i,$$

where

$$\varepsilon_i \sim i.i.d. N(0, \sigma^2), i = 1, 2, \ldots, n.$$

Design matrix

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} \end{pmatrix}$$

Two-sample pooled T-test

Consider independent observations

$$\begin{array}{cccc} y_{11}, & y_{12}, & \dots & y_{1n_1} \\ y_{21}, & y_{22}, & \dots & y_{2n_2} \end{array}$$

with

$$Y_{ij} \sim N(\mu_i, \sigma^2) \text{ for } j = 1, 2, \dots, n_i; i = 1, 2,$$

Set as a MLR model

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{pmatrix}.$$

Remarks

$$\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{Y} = \begin{pmatrix} \bar{Y}_{1\bullet} \\ \bar{Y}_{2\bullet} - \bar{Y}_{1\bullet} \end{pmatrix}$$

Proof

Remarks

$$S_e^2 = \frac{1}{n-p}\|\boldsymbol{Y} - X\hat{\beta}\|^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}.$$

Proof

Remarks

$$H_0 : \mu_1 - \mu_2 = 0,$$
$$H_a : \mu_1 - \mu_2 \neq 0,$$

The appropriate test statistic is

$$T = \frac{\bar{Y}_{2\bullet} - \bar{Y}_{1\bullet}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Proof

One-way ANOVA

How to represent categorical variables in R

```
stats  <- tribble(
  ~name, ~pet, ~happiness,
  "Jono", "Rabbit", 10,
  "Ben", "Cat", 7,
  "SJ", "Cat", 9,
  "Ty", "Dog", 6,
  "Gary", "Dog", 2
)
stats$pet  <- factor(stats$pet)
```

How to represent categorical variables in R

name	pet	happiness
Jono	Rabbit	10
Ben	Cat	7
SJ	Cat	9
Ty	Dog	6
Gary	Dog	2

How to represent categorical variables in R

```
library(modelr)
model_matrix(stats, happiness~pet)
```

```
## # A tibble: 5 x 3
##   `(Intercept)` petDog petRabbit
##           <dbl>  <dbl>     <dbl>
## 1             1      0         1
## 2             1      0         0
## 3             1      0         0
## 4             1      1         0
## 5             1      1         0
```

Representation in R

Consider a predictor X with k levels. R represents this as $k-1$ binary variables:

$$x_{ij} = \begin{cases} 1, & \text{if } X \text{ for the } i\text{th individual is level } k \\ 0, & \text{otherwise.} \end{cases}$$

Where is the final level?

The one-way layout

Consider independent observations in k groups

$$\begin{array}{cccc}
y_{11}, & y_{12}, & \cdots & y_{1n_1} \\
y_{21}, & y_{22}, & \cdots & y_{2n_2} \\
\vdots, & & & \vdots \\
y_{k1}, & y_{k2}, & \cdots & y_{kn_k}
\end{array}$$

with

$$Y_{ij} \sim i.i.d. N(\mu_i, \sigma^2)$$

Testing for the same mean

Suppose we want to test

$$H_0: \ \mu_1 = \mu_2 = \ldots = \mu_k.$$

How can we perform this using a linear model?

Write as a multiple regression model

This model may also be formulated as a multiple regression model by considering the model formulation

$$M : \eta_{ij} = \mu_i = \mu + \alpha_i,$$

where μ denotes the overall mean and α_i is a parameter specific to group i.

We need to set $\alpha_1 = 0$. Why?

MLR model

$$
\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ y_{k2} \\ \vdots \\ y_{kn_k} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_k \end{pmatrix}
$$

New hypothesis

We can rewrite

$$H_0: \ \mu_1 = \mu_2 = \ldots = \mu_k$$

as

$$H_0: \ \alpha_2 = \alpha_3 = \ldots = \alpha_k = 0$$

ANOVA table

Source	SS	df
Between Groups	$\sum_i n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	$k-1$
Within Groups	$\sum_{ij}(y_{ij} - \bar{y}_{i\bullet})^2$	$n-k$
Total	$\sum_{ij}(y_{ij} - \bar{y}_{\bullet\bullet})^2$	$n-1$

where

$$n = \sum_{i=1}^{k} n_i, \;\; \bar{y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad \bar{y}_{\bullet\bullet} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}.$$

The F-statistic

$$F = \frac{\sum_i n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2/(k-1)}{\sum_{ij}(y_{ij} - \bar{y}_{i\bullet})^2/(n-k)},$$

and H_0 is rejected when $F \geq F_{k-1,n-k}(\alpha)$.

Two-way ANOVA

Setup (no replication)

Consider the two-way layout with one observation per cell

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

where

$$\varepsilon_{ij} \sim i.i.d. N(0, \sigma^2), \;\; i = 1, 2, \ldots, I, \;\; j = 1, 2, \ldots J$$

Contraints

- **Zero Sum Constraints:**

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0.$$

- **Reference Category Constraints:**

$$\alpha_1 = \beta_1 = 0.$$

Hypotheses

Within the present context, there are two hypotheses of interest:

$$H_1 : \ \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0 \text{ and } H_2 : \ \beta_1 = \beta_2 = \ldots = \beta_J = 0$$

MLR (Zero sum)

The additive model may also be specified as a multiple regression model. The formulation for the zero-sum constraints is shown below.

$$
\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1J} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2J} \\ \vdots \\ y_{I1} \\ y_{I2} \\ \vdots \\ y_{IJ} \end{pmatrix}, \quad X = \begin{pmatrix}
1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & \ddots & \vdots \\
1 & 1 & 0 & \dots & 0 & -1 & -1 & \dots & -1 \\
1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & \ddots & \vdots \\
1 & 0 & 1 & \dots & 0 & -1 & -1 & \dots & -1 \\
\vdots & & & & & & & & \vdots \\
1 & -1 & -1 & \dots & -1 & 1 & 0 & \dots & 0 \\
1 & -1 & -1 & \dots & -1 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & \ddots & \vdots \\
1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1
\end{pmatrix}
$$

MLR (reference)

The formulation for the reference category constraints is shown below.

$$
\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1J} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2J} \\ \vdots \\ y_{I1} \\ y_{I2} \\ \vdots \\ y_{IJ} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & & \vdots & & \ddots & \vdots \\ 1 & 0 & \dots & 0 & 0 & \dots & 1 \\ 1 & 1 & \dots & 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & & \vdots & & \ddots & \vdots \\ 1 & 1 & \dots & 0 & 0 & \dots & 1 \\ \vdots & & & & & & \vdots \\ 1 & 0 & \dots & 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 1 & 1 & \dots & 0 \\ \vdots & & & \vdots & & \ddots & \vdots \\ 1 & 0 & \dots & 1 & 0 & \dots & 1 \end{pmatrix}.
$$

ANOVA

Source	SS	df	MSE	F
Row Effects	$J\sum_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	I-1	MSA	MSA/MSE
Col Effects	$I\sum_j(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$	J-1	MSB	MSB/MSE
Residual	$\sum_{ij}(y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2$	(I-1)(J-1)	MSE	
Total	$\sum_{ij}(y_{ij} - \bar{y}_{\bullet\bullet})^2$	IJ-1	MST	

Model for two-way ANOVA with interaction (replication)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where

$$\varepsilon_{ij} \sim i.i.d.N(0, \sigma^2),\ \ i = 1, 2, \ldots, I,\ \ j = 1, 2, \ldots J, k = 1, \ldots, K.$$

Constraints

- **Zero Sum Constraints:**

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0.$$

- **Reference Category Constraints:**

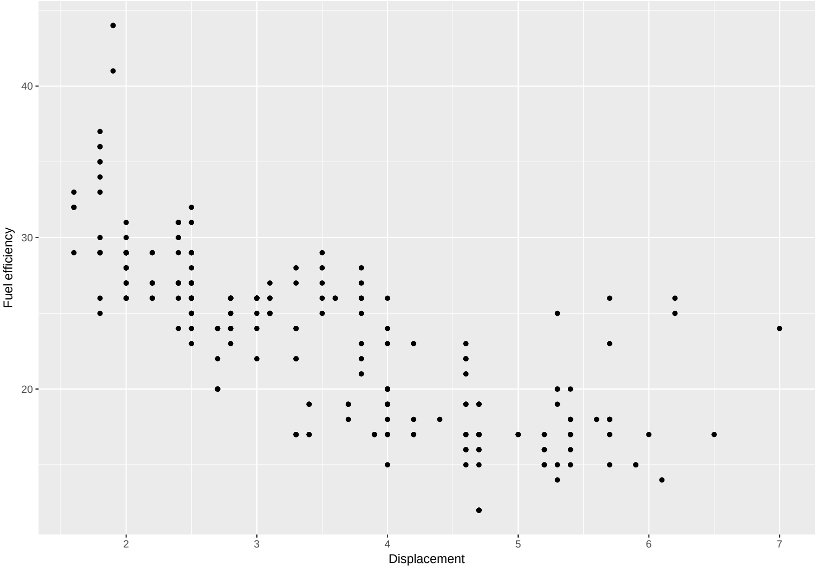
$$\alpha_1 = \beta_1 = \gamma_{1j} = \gamma_{i1} = 0,\ \ i = 1, \ldots, I,\ \ j = 1, \ldots J.$$

ANOVA table

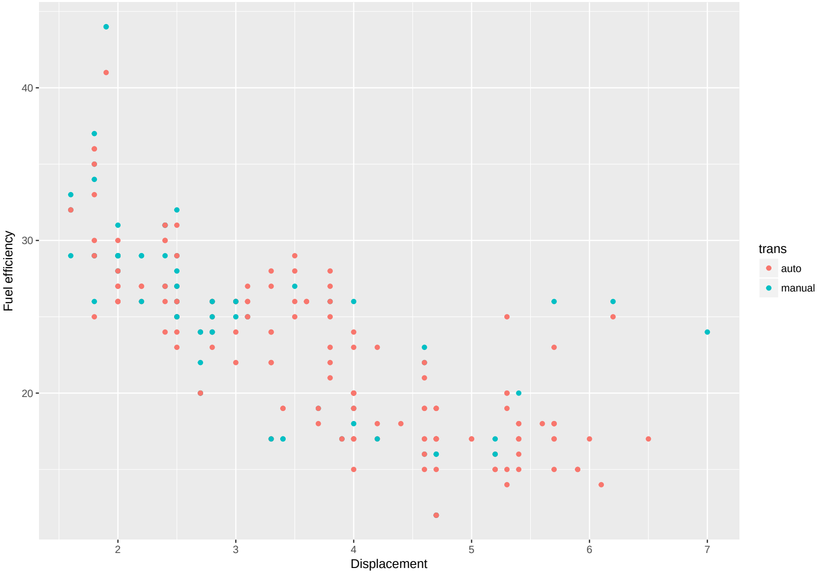
Source	SS	df	MSE	F
Tx A	$JK\sum_i(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$	I-1	MSA	MSA/MSE
Tx B	$IK\sum_j(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$	J-1	MSB	MSB/MSE
Interaction	$K\sum_{ij}(y_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2$	(I-1)(J-1)	MSI	MSI / MSE
Residual	$\sum_{ijk}(y_{ijk} - \bar{y}_{ij\bullet})^2$	IJ(K-1)	MSE	
Total	$\sum_{ij}(y_{ij} - \bar{y}_{\bullet\bullet})^2$	IJ-1	MST	

Analysis of covariance (ANCOVA)

Motivating example



Motivating example



Types of models

- **Identical regression**:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \ \ i = 1, 2, \ldots, I, \ \ j = 1, 2, \ldots n_i.$$

- **Parallel regression**:

$$Y_{ij} = \beta_{i0} + \beta_1 x_{ij} + \varepsilon_{ij}, \ \ i = 1, 2, \ldots, I, \ \ j = 1, 2, \ldots n_i.$$

- **Seperate regression**:

$$Y_{ij} = \beta_{i0} + \beta_{i1} x_{ij} + \varepsilon_{ij}, \ \ i = 1, 2, \ldots, I, \ \ j = 1, 2, \ldots n_i.$$

Identical regression

```
model_matrix(mpg, hwy ~ displ)
```

```
## # A tibble: 234 x 2
##    `(Intercept)` displ
##            <dbl> <dbl>
##  1             1   1.8
##  2             1   1.8
##  3             1   2.0
##  4             1   2.0
##  5             1   2.8
##  6             1   2.8
##  7             1   3.1
##  8             1   1.8
##  9             1   1.8
## 10             1   2.0
## # ... with 224 more rows
```

Identical regression

We can represent this as the linear model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where

$$\varepsilon_i \sim i.i.d. N(0, \sigma^2), \ \ i = 1, 2, \ldots, n.$$

Parallel regression

```
model_matrix(mpg, hwy ~ displ + trans)
```

```
## # A tibble: 234 x 3
##    `(Intercept)` displ transmanual
##            <dbl> <dbl>       <dbl>
##  1             1   1.8           0
##  2             1   1.8           1
##  3             1   2.0           1
##  4             1   2.0           0
##  5             1   2.8           0
##  6             1   2.8           1
##  7             1   3.1           0
##  8             1   1.8           1
##  9             1   1.8           0
## 10             1   2.0           1
## # ... with 224 more rows
```

Parallel regression

We can represent this as the linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

where x_{i1} is the displacement, and x_{i2} is 1 for a manual and 0 for an automatic, and

$$\varepsilon_i \sim i.i.d.N(0, \sigma^2),\ i = 1, 2, \ldots, n.$$

What is the model for a manual; a automatic?

Seperate regression

```r
model_matrix(mpg, hwy ~ displ + trans + displ:trans)
```

```
## # A tibble: 234 x 4
##    `(Intercept)` displ transmanual `displ:transmanual`
##            <dbl> <dbl>      <dbl>               <dbl>
##  1             1   1.8          0                 0.0
##  2             1   1.8          1                 1.8
##  3             1   2.0          1                 2.0
##  4             1   2.0          0                 0.0
##  5             1   2.8          0                 0.0
##  6             1   2.8          1                 2.8
##  7             1   3.1          0                 0.0
##  8             1   1.8          1                 1.8
##  9             1   1.8          0                 0.0
## 10             1   2.0          1                 2.0
## # ... with 224 more rows
```

Seperate regression

```r
model_matrix(mpg, hwy ~ displ * trans)
```

```
## # A tibble: 234 x 4
##    `(Intercept)` displ transmanual `displ:transmanual`
##            <dbl> <dbl>       <dbl>               <dbl>
##  1             1   1.8           0                 0.0
##  2             1   1.8           1                 1.8
##  3             1   2.0           1                 2.0
##  4             1   2.0           0                 0.0
##  5             1   2.8           0                 0.0
##  6             1   2.8           1                 2.8
##  7             1   3.1           0                 0.0
##  8             1   1.8           1                 1.8
##  9             1   1.8           0                 0.0
## 10             1   2.0           1                 2.0
## # ... with 224 more rows
```

Seperate regression

We can represent this as the linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i,$$

where x_{i1} is the displacement, and x_{i2} is 1 for a manual and 0 for an automatic, and

$$\varepsilon_i \sim i.i.d. N(0, \sigma^2), \; i = 1, 2, \ldots, n.$$

What is the model for a manual; a automatic?

Model selection

Start with largest model - separate regression.

```
model_sep  <- lm(hwy ~ displ * trans, data = mpg)
summary(model_sep)
```

```
##
## Call:
## lm(formula = hwy ~ displ * trans, data = mpg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1441 -2.2946 -0.2436  2.1184 14.7553
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        35.39457    0.94674  37.386   <2e-16 ***
## displ              -3.52217    0.24090 -14.621   <2e-16 ***
## transmanual         0.02559    1.51343   0.017    0.987
## displ:transmanual   0.27194    0.44143   0.616    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.828 on 230 degrees of freedom
## Multiple R-squared:  0.5921, Adjusted R-squared:  0.5868
## F-statistic: 111.3 on 3 and 230 DF,  p-value: < 2.2e-16
```

Parallel model

```
model_parallel  <- update(model_sep, .~. - displ:trans)
summary(model_parallel)
```

```
##
## Call:
## lm(formula = hwy ~ displ + trans, data = mpg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8130 -2.2109 -0.2639  2.0964 14.5517
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.0933     0.8096  43.348   <2e-16 ***
## displ        -3.4412     0.2016 -17.070   <2e-16 ***
## transmanual   0.8933     0.5531   1.615    0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.823 on 231 degrees of freedom
## Multiple R-squared:  0.5914, Adjusted R-squared:  0.5879
## F-statistic: 167.2 on 2 and 231 DF,  p-value: < 2.2e-16
```

Identical model

```r
model_identical  <- update(model_parallel, .~. - trans)
summary(model_identical)
```

```
##
## Call:
## lm(formula = hwy ~ displ, data = mpg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1039 -2.1646 -0.2242  2.0589 15.0105
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.6977     0.7204   49.55   <2e-16 ***
## displ        -3.5306     0.1945  -18.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.836 on 232 degrees of freedom
## Multiple R-squared:  0.5868, Adjusted R-squared:  0.585
## F-statistic: 329.5 on 1 and 232 DF,  p-value: < 2.2e-16
```

AIC

```r
AIC(model_identical, model_parallel, model_sep)
```

```
##                 df      AIC
## model_identical  3 1297.246
## model_parallel   4 1296.619
## model_sep        5 1298.233
```

BIC

```r
BIC(model_identical, model_parallel, model_sep)
```

```
##                 df      BIC
## model_identical  3 1307.612
## model_parallel   4 1310.440
## model_sep        5 1315.510
```

Example