Level M Regression Models Lecture 11

Now we will focus on hypothesis testing of the regression parameters. In the previous week we considered diagnostics for and assumptions about linear regression models. Now we will consider inference for model parameters, model comparison and selection. We will construct interval estimates and hypothesis tests for various parameters of our models.

R code

Please find some R code used to fit and plot some models. You can find most data sets on Moodle. Please download data and try some of the R code as you read through these notes.

Linear Combinations of Parameters

Now we will focus on hypothesis testing of the regression parameters. Instead of solving several different types of inferential problems, e.g. involving a single parameter, involving two parameters or in some cases involving a linear combination of parameters we will develop a general theory for doing inference on linear combinations of parameters. Each of the cases described in the previous sentence can then be derived as a special case of the general theory.

For example if we want to predict a future value, at $x = 5$ based on a simple linear model

$$y_i = \beta_0 + \beta_1 x + \epsilon_i \implies \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$$

we are interested in the linear combination:

$$\beta_0 + 5\beta_1,$$

which can be written as:

$$(1 \quad 5) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

We start by considering how to form linear functions of the parameters in a linear model.

The least-squares estimate of linear functions of the parameters in a multiple linear model

Data: $(y_i, x_{1i}, x_{2i}, \ldots, x_{ki})$; $\quad i = 1, \ldots, n$
Model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} = \{\epsilon_1, \ldots, \epsilon_n\} \sim N(0, \sigma^2)$ independent.

Suppose we want the least-squares estimate for a linear function of the parameters,

- Say $\mathbf{b}_1^T\boldsymbol{\beta}$ for some given vector $\mathbf{b}_1$. For instance in the previous example $\mathbf{b}_1^T = (1 \quad 5)$, or
- Possibly for a set of s linearly independent linear combinations $\mathbf{b}_1^T\boldsymbol{\beta}, \ldots, \mathbf{b}_s^T\boldsymbol{\beta}$, $s \leq p$ where the $\mathbf{b}_i$'s are given vectors (similar to the previous example). Here, p is the number of regression coefficients, hence $\boldsymbol{\beta}$ is a vecotr of length p.

It is always possible to create a non-singular transformation from $\boldsymbol{\beta} \leftrightarrow \boldsymbol{\phi}$ where

$$\boldsymbol{\phi} = \begin{pmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}_s^T \end{pmatrix} \boldsymbol{\beta} = \mathbf{B}\boldsymbol{\beta},$$

where $\mathbf{B}$ is $s \times p$ nonsingular and so invertible matrix. So

$$\boldsymbol{\beta} = \mathbf{B}^{-1}\boldsymbol{\phi}$$

where B^{-1} is of dimension $p \times s$. It is now possible to rewrite our model in terms of $\boldsymbol{\phi}$, where $\phi_1, \ldots, \phi_s$ are the parameters of interest.

Data: $(y_i, x_{1i}, x_{2i}, \ldots, x_{pi}); \quad i = 1, \ldots, n$

Model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = (\mathbf{X}\mathbf{B}^{-1})\boldsymbol{\phi} + \boldsymbol{\epsilon}$

$(\mathbf{X}\mathbf{B}^{-1})$ is an $n \times s$ matrix which is known (and is just a transformed design matrix) and $\boldsymbol{\phi}$ is a s vector of unknown parameters.

The form of the model is mathematically equivalent to our original form, substituting $(\mathbf{X}\mathbf{B}^{-1})$ for the design matrix and $\boldsymbol{\phi}$ for the parameter vector.

Hence, we can write down the solution for the parameter estimates, based on least-squares, from our earlier results.

$$\begin{aligned}
\hat{\boldsymbol{\phi}} &= \{(\mathbf{X}\mathbf{B}^{-1})^T(\mathbf{X}\mathbf{B}^{-1})\}^{-1}(\mathbf{X}\mathbf{B}^{-1})^T\mathbf{Y} \\
&= \{(\mathbf{B}^{-1})^T(\mathbf{X}^T\mathbf{X})\mathbf{B}^{-1}\}^{-1}(\mathbf{B}^{-1})^T\mathbf{X}^T\mathbf{Y} \\
&= \mathbf{B}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{B}^T(\mathbf{B}^T)^{-1}\mathbf{X}^T\mathbf{Y} \\
&= \mathbf{B}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \\
&= \mathbf{B}\hat{\boldsymbol{\beta}}.
\end{aligned}$$

Hence the least-squares estimates of a set of linear functions of parameters is just the set of linear functions of the least-squares estimates.

Application of linear transformation of parameters for two parameter case

A useful application of this result may sometimes simplify the calculation of least-squares estimates. The basic idea is that it may be possible to rewrite a model in terms of parameters whose estimates are "easier" to calculate and then we can transform back to the original parameters. This approach is often referred to as 'centering'. Centering in often useful to produce orthogonal columns which turn gives us diagonal matrices to invert.

Example

Data: $(y_i, x_i); \quad i = 1, \dots, n$

Model: $y_i = \alpha + \beta x_i + \epsilon_i$, or in vector matrix notation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = (\mathbf{X}\mathbf{B}^{-1})\boldsymbol{\phi} + \boldsymbol{\epsilon}$

Suppose we transform
$y_i = \alpha + \beta x_i + \epsilon_i$ (Model 1) to
$y_i = \alpha' + \beta(x_i - \bar{x}) + \epsilon_i$ (Model 2)

Firstly, consider Model 2

$$\begin{aligned} y_i &= \alpha' + \beta(x_i - \bar{x}) + \epsilon_i \\ &= \alpha' + \beta x_i - \beta\bar{x} + \epsilon_i \\ &= \alpha' - \beta\bar{x} + \beta x_i + \epsilon_i \end{aligned}$$

which implies $\alpha = \alpha' - \beta\bar{x}$

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \leftrightarrow \boldsymbol{\phi} = \begin{pmatrix} \alpha' \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha + \beta\bar{x} \\ \beta \end{pmatrix}, \quad \bar{x} = (\sum_{i=1}^{n} x_i)/n.$$

Writing Model 2 in vector matrix notation,

$$E(\mathbf{Y}) = \begin{pmatrix} 1 & (x_1 - \bar{x}) \\ . & . \\ . & . \\ . & . \\ . & . \\ 1 & (x_n - \bar{x}) \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta \end{pmatrix} = \mathbf{X}\mathbf{B}^{-1}\boldsymbol{\phi}$$

where $\boldsymbol{\phi} = \mathbf{B}\boldsymbol{\beta}$ and $\mathbf{B} = \begin{pmatrix} 1 & \bar{x} \\ 0 & 1 \end{pmatrix}$

$$\hat{\boldsymbol{\phi}} = \{(\mathbf{X}\mathbf{B}^{-1})^T(\mathbf{X}\mathbf{B}^{-1})\}^{-1}(\mathbf{X}\mathbf{B}^{-1})^T\mathbf{Y}$$

$$\begin{aligned} (\mathbf{X}\mathbf{B}^{-1})^T(\mathbf{X}\mathbf{B}^{-1}) &= \begin{pmatrix} n & \sum_{i=1}^n (x_i - \bar{x}) \\ \sum_{i=1}^n (x_i - \bar{x}) & \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix} \\ &= \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix} \end{aligned}$$

i.e. $(\mathbf{X}\mathbf{B}^{-1})^T(\mathbf{X}\mathbf{B}^{-1})$ is diagonal.

$$(\mathbf{X}\mathbf{B}^{-1})^T\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i(x_i - \bar{x}) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \end{pmatrix}$$

$$\{(\mathbf{X}\mathbf{B}^{-1})^T(\mathbf{X}\mathbf{B}^{-1})\}^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix}$$

i.e.

$$\hat{\boldsymbol{\phi}} = \begin{pmatrix} \sum_{i=1}^n y_i/n \\ \frac{\sum_{i=1}^n (y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^n (x_i-\bar{x})^2} \end{pmatrix} = \begin{pmatrix} \hat{\alpha}' \\ \hat{\beta} \end{pmatrix}$$

Because of our choice of $\boldsymbol{\phi}$, $(\mathbf{XB}^{-1})^T(\mathbf{XB}^{-1})$ is easier to invert and hence the calculations are simpler.

From the nature of the transformation clearly

$$\hat{\alpha} = \hat{\alpha}' - \hat{\beta}\bar{x}$$

Application of linear transformation of parameters for three parameter case

Data: $(y_i, x_{1i}, x_{2i}), \quad i = 1, \ldots, n$

Model: $E(Y_i) = \alpha + \beta x_{1i} + \gamma x_{2i}$

Reparameterise to

Model : $E(Y_i) = \alpha' + \beta(x_{1i} - \bar{x}_{1.}) + \gamma(x_{2i} - \bar{x}_{2.})$

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \leftrightarrow \boldsymbol{\phi} = \begin{pmatrix} \alpha' \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha + \beta\bar{x}_{1.} + \gamma\bar{x}_{2.} \\ \beta \\ \gamma \end{pmatrix},$$

where $\bar{x}_{1.} = \sum_{i=1}^n x_{1i}/n, \quad \bar{x}_{2.} = \sum_{i=1}^n x_{2i}/n$. i.e.

$$E(\mathbf{Y}) = \begin{pmatrix} 1 & (x_{11} - \bar{x}_{1.}) & (x_{21} - \bar{x}_{2.}) \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & (x_{1n} - \bar{x}_{1.}) & (x_{2n} - \bar{x}_{2.}) \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta \\ \gamma \end{pmatrix} = \mathbf{XB}^{-1}\boldsymbol{\phi}$$

$$\hat{\boldsymbol{\phi}} = ((\mathbf{XB}^{-1})^T(\mathbf{XB}^{-1}))^{-1}(\mathbf{XB}^{-1})^T\mathbf{Y}$$

$$((\mathbf{XB}^{-1})^T(\mathbf{XB}^{-1})) = \begin{pmatrix} n & 0 & 0 \\ 0 & \sum_{i=1}^n (x_{1i} - \bar{x}_{1.})^2 & \sum_{i=1}^n (x_{1i} - \bar{x}_{1.})(x_{2i} - \bar{x}_{2.}) \\ 0 & \sum_{i=1}^n (x_{1i} - \bar{x}_{1.})(x_{2i} - \bar{x}_{2.}) & \sum_{i=1}^n (x_{2i} - \bar{x}_{2.})^2 \end{pmatrix}$$

$$= \begin{pmatrix} n & \mathbf{0} \\ \mathbf{0}^T & \boldsymbol{\Psi} \end{pmatrix}$$

$$((\mathbf{XB}^{-1})^T(\mathbf{XB}^{-1}))^{-1} = \begin{pmatrix} \frac{1}{n} & \mathbf{0} \\ \mathbf{0}^T & \boldsymbol{\Psi}^{-1} \end{pmatrix}$$

Hence inversion of $((\mathbf{XB}^{-1})^T(\mathbf{XB}^{-1}))$ is reduced to inversion of a (2×2) matrix, a great saving in calculation. In general a similar transformation will reduce the inversion of a $(p \times p)$ matrix to the inversion of a $((p-1) \times (p-1)$ matrix.

After calculation of $\hat{\boldsymbol{\phi}}, \hat{\alpha}$ can be obtained from

$$\hat{\alpha} = \hat{\alpha}' - \hat{\beta}\bar{x}_{1.} - \hat{\gamma}\bar{x}_{2.}$$

Inferences from regression equations

If we are interested in $\mathbf{b}^T\boldsymbol{\beta}$ (a linear function of the parameters), where $\mathbf{b}$ is a given vector of constants, we will use the concept of pivotal functions that you have learnt in Statistical Inference.

Pivotal function for a linear function of the parameters

$$\frac{(\mathbf{b}^T\hat{\boldsymbol{\beta}} - \mathbf{b}^T\boldsymbol{\beta})}{\sqrt{\frac{RSS}{n-p}\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b}}}$$

is a pivotal function since

$$\frac{(\mathbf{b}^T\hat{\boldsymbol{\beta}} - \mathbf{b}^T\boldsymbol{\beta})}{\sqrt{\frac{RSS}{n-p}\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b}}} \sim t(n-p),$$

where p is the number of parameters, n is the sample size and RSS is the residual sum-of-squares in a linear model.

This result is stated without proof. It is helpful to use the notation ***estimated standard error*** for the quantity

$$\sqrt{\frac{RSS}{n-p}\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b}}$$

(The word "estimated" is often omitted from the name).

The above result can be used to construct hypothesis tests and interval estimates for model parameters.

Hypothesis Testing

If we were interested in making inferences about β in a simple linear regression model i.e.

$$y_i = \alpha + \beta x_i + \epsilon_i$$

$\mathbf{b}^T\boldsymbol{\beta} = \beta$ i.e. $\mathbf{b}^T = (0 \quad 1)$ and this gives us:

$$\frac{\hat{\beta} - \beta}{\text{e.s.e}(\hat{\beta})} \sim t(n-p)$$

Under the null hypothesis:

H_0: $\beta = 0$ (where $\text{H}_1 : \beta \neq 0$)

$$\frac{\hat{\beta} - 0}{\text{e.s.e}(\hat{\beta})} = \frac{\hat{\beta}}{\text{e.s.e}(\hat{\beta})} \sim t(n-p)$$

and $\frac{\hat{\beta}}{\text{e.s.e}(\hat{\beta})}$ is typically called the t-statistic. Therefore, the null hypothesis is rejected for large absolute values of the t-statistic, usually values > 2 i.e. for small p-values in `R` (where a p-value is the probability that we obtain a t-statistic value as extreme or more extreme if the null hypothesis is true). In general, we reject H_0 for p-values < 0.05 and this would indicate a significant relationship between a response and an explanatory variable in the model.

Notice we can test any hypothsis with respect to β, for instance

H_0: $\beta = 5$ (where $H_1 : \beta \neq 5$)

$$\frac{\hat{\beta} - 5}{\text{e.s.e}(\hat{\beta})} \sim t(n-p).$$

Example

Hypothesis testing for Pregnancy Data

Data were collected through interest in whether, and if so, in what way the level of protein changes in expectant mothers throughout their pregnancy. Observations were taken on 19 healthy women. Each woman was at a different stage of pregnancy, gestation.

We have seen this example previously for parameter estimation assessing model fit.

Data: $(y_i, x_i) \quad i = 1, \ldots, 19$

Model: $E(Y_i) = \alpha + \beta x_i$

Perform a hypothesis test to test $H_0 : \beta = 0$.

We will use the `R` output to answer this question, however please make sure you are able to estimate regression coefficients by hand.

```
pregnancy<-read.csv("PROTEIN.CSV",header=T)
fit1<-lm(formula = Protein ~ Gestation,data=pregnancy)
summary(fit1)
```

```
##
## Call:
## lm(formula = Protein ~ Gestation, data = pregnancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16853 -0.08720 -0.01009  0.08578  0.20422
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.201738   0.083363   2.420    0.027 *
## Gestation   0.022844   0.003295   6.934 2.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1151 on 17 degrees of freedom
## Multiple R-squared:  0.7388, Adjusted R-squared:  0.7234
## F-statistic: 48.08 on 1 and 17 DF,  p-value: 2.416e-06
```

The hypotheses being tested for the coefficient of β are:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

From the regression output

$$\frac{\hat{\beta}}{\text{e.s.e}(\hat{\beta})} = \frac{0.022844}{0.003295} = 6.932929 \sim t(n-p) \text{ under } H_0.$$

Since the p-value for gestation is < 0.001 (and hence < 0.05) the null hypothesis is rejected and we conclude that there is a statistically significant relationship between protein and gestation. The gestational age is a useful predictor of the protein level.

Analysing the ANOVA table

The F statistic value: $\text{MS}_{\text{model}}/\text{MS}_{\text{residuals}}$ provides a test statistic that allows us to test whether there is any evidence that at least one of the model parameters is not zero.

The null hypothesis is

H_0: all p parameters = 0,

which will be tested against the alternative that at least one of the parameters is not zero. If the null hypothesis is true, the statistic has an $\text{F}(\text{Df}_{\text{model}}, \text{Df}_{\text{residuals}})$ distribution. This implies that

$$F = \frac{MS_{\text{model}}}{MS_{\text{residuals}}} \sim F(Df_{\text{model}}, Df_{\text{residuals}}).$$

If H_0 is false, we would expect $\text{MS}_{\text{residuals}}$ to be smaller than MS_{model} and so large values of F should lead us to reject H_0. i.e. for large values of F the p-value will be small.

Example

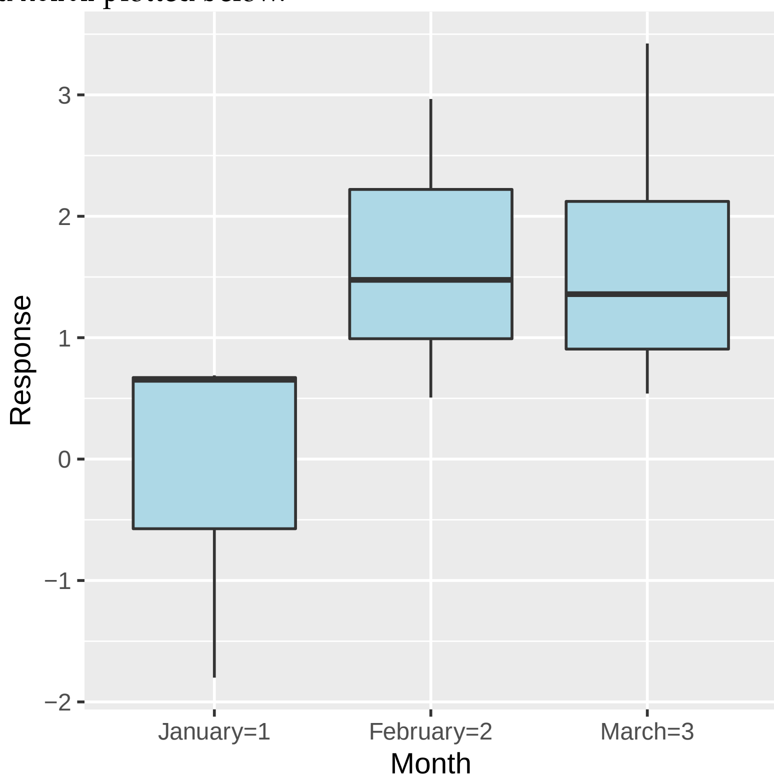
Looking again at the anova table for protein in pregnancy,

```r
fit1<-lm(formula = Protein ~ Gestation, data=pregnancy)
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: Protein
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Gestation  1 0.63667 0.63667  48.076 2.416e-06 ***
## Residuals 17 0.22513 0.01324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the null hypothesis with a p-value of 2.4×10^{-6} suggesting that at least one model parameter is not zero.

Let's now re-vist an example detailing the relationship between a repsonse variables `Response` and `Month` plotted below.



Let's now look at some summaries and regression model output from these data

```r
summary(data$Response)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.8000  0.5683  0.8590  1.1172  1.6358  3.4230
```

```r
tapply(data$Response, data$Month, mean)
```

```
##          1          2          3
## -0.1523333  1.6496667  1.6700000
```

The average value of `Response` was 1.12 with month averages of -0.152, 1.650 and 1.670 for January, February and March respectively. We want to know if there are any significant differences in `Response` across the three months.

```r
model<-lm(Response ~ factor(Month), data=data)
summary(model)
```

```
##
## Call:
## lm(formula = Response ~ factor(Month), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.64767 -1.00800 -0.07733  0.83308  1.75300
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.1523     0.7529  -0.202    0.845
## factor(Month)2   1.8020     1.0648   1.692    0.134
## factor(Month)3   1.8223     0.9960   1.830    0.110
##
## Residual standard error: 1.304 on 7 degrees of freedom
## Multiple R-squared:  0.3672, Adjusted R-squared:  0.1864
## F-statistic: 2.031 on 2 and 7 DF,  p-value: 0.2016
```

The intercept term is in fact the average values of our baseline category January. The estimate 1.802 corresponds to the difference between Month 2, February, and January. Therefore, the average balue in February is $1.802 - 0.152 = 1.65$. Likewise, the average value in March is $1.822 - 0.153 = 1.67$.

We can now interpret p-values corresponding to each estimate. The p-value of 0.85 correspoding the the intercept terms tells us the that average value in January is not significantly different from 0. More interestingly, the p-value of 0.13 corresponsinf to Month2 tells us that this coefficient is not signficantly different from 0 and so Febraruy is not signficantly different from January. The p-value of 0.11 corresponding to Month3 tells us that March is not signficantly different from January

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Response
##               Df  Sum Sq Mean Sq F value Pr(>F)
## factor(Month)  2  6.9081  3.4540  2.0311 0.2016
## Residuals      7 11.9039  1.7006
```

The null hypothesis of this anova is that all regression coefficients in our fitted model are equal to 0. The alternative hypothesis is that at least one is not equal to zero. Given a p-value of 0.2, this suggests we cannot reject the null hypothesis. This means that all regression coefficients are zero. In particular, this implies that there is no signficant difference between the three months. In other words, in this case with one categorical variable, the anova is testing for *a* difference between categories.

The anova reports one p-value that test for at least coefficient to be significantly different from zero. A regression reports one mean (as the intercept intercept) and the differences between that one and all other means, but the p-values evaluate those specific comparisons.

Additional Reading

Please see

- Section 3.2 in **Linear Models with R**.
- Sections 3.10 in **Regression Analysis By Example**.
- Section 3.1.2 in **An Introduction to Statistical Learning**.