Inference

Based on a book by Julian J. Faraway

University of Iceland

Inference in MLR

- If we want to make any confidence intervals or perform any hypothesis tests we will need to assume some distributional form for the errors ε.
- The usual assumption is that the errors are independent and identically normally distributed with mean 0 and variance σ^2, i.e. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma^2 I})$
- It is possible to handle non-identity variance matrices provided we know the form - see Chapter 5.

Hypothesis tests to compare models

- Given several predictors for a response we might wonder whether all are needed.
- Consider a large model, Ω, and a smaller model, ω, which consists of a subset of the predictors that are in Ω.
- We will take ω to represent the null hypothesis and Ω to represent the alternative.
- If RSS_ω - RSS_Ω is small, then ω is an adequate model relative to Ω. This suggests that something like

$$\frac{RSS_\omega - RSS_\Omega}{RSS_\Omega}$$

  would be a potentially good test statistic. You will see in later courses that the same statistic arises from the likelihood ratio testing approach.

Hypothesis tests to compare models

Now suppose that the dimension (no. of parameters) of Ω is q and the dimension of ω is p.

If the null hypothesis is true the following holds (Cochran's theorem):

$$\frac{RSS_\omega - RSS_\Omega}{q-p} \sim \sigma^2 \chi^2_{q-p} \quad \text{and} \quad \frac{RSS_\Omega}{n-q} \sim \sigma^2 \chi^2_{n-q}$$

and

$$\frac{(RSS_\omega - RSS_\Omega)/(q-p)}{RSS_\Omega/(n-q)} \sim F_{q-p,n-q}$$

The overall F-test

Are any of the predictors useful in predicting the response?

- Full model (Ω): $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{X}$ is a full rank $n \times p$ matrix.
- Null model (ω): $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, predict y by the mean.

If the null hypothesis is true the following holds:

$$\frac{(SSY - RSS)/(p-1)}{RSS/(n-p)} \sim F_{p-1,n-p}$$

Thus we would reject the null hypothesis if $F > F_{(1-\alpha),p-1,n-p}$

Hypothesis tests to compare models

- Rejecting the null does not imply that the alternative model is the best model.
  - We don't know whether all the predictors are required to predict the response or just some of them.
  - Other predictors, not included in the model, might also be added (for example quadratic terms in the existing predictors).
- Either way, the overall F-test is just the beginning of an analysis and not the end.

Performing the F-test in R

- Fit the larger model using `lm()` and store it in an lm-object (`fit.1`)
- Fit the smaller model using `lm()` and store it in an lm-object (`fit.2`)
- Use the `anova()` function to perform the test (`anova(fit.1,fit.2)`)

Testing just one predictor

The null hypothesis for dropping one particular predictor, β_i from the model would be $H_0 : \beta_i = 0$

The F-test can be set up in the following manner:

- RSS_Ω is the RSS for the model with all the predictors of interest (p parameters).
- RSS_ω is the RSS for the model with all the above predictors except predictor i.

The F-statistic may be computed using the formula from above.

Testing just one predictor

An alternative approach to test the hypothesis

$$H_0 : \beta_i = 0$$

is to use a t-statistic:

$$t_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$$

The null hypothesis should be rejected if $t > t_{1-\alpha/2,n-p}$.

The two approaches, using the F- and the t-test, are identical.

Testing a pair of predictors

- Except in special circumstances, dropping one variable from a regression model causes the estimates of the other parameters to change.
- This means that we might find that after dropping some variable X_j that a test of the significance of another variable X_k shows that it should be included in the model even though it was not significant when X_j was in the model.
- Therefore: remove one variable at a time from the model.

Testing the predictors

- Variables where the p-value is less than α are said to be *significant*.
- Variables where the p-value is greater than α are said to be *nonsignificant*.
- It is not clear cut when to remove a variable from the model:
  - if p-value > 0.25 we *usually* remove the variable.
  - if $\alpha <$ p-value < 0.25 we *usually* keep the variable.
  - if p-value $< \alpha$ we keep the variable.
- When performing more than one test we need to think about multiplicity issues...

We will look more closely into model selection in Chapter 10.

Concerns about Hypothesis Testing

- Sampling
- Power/lack of power
- Inference depends on the correctness of the model
- Statistical significance is not equivalent to practical significance

Confidence intervals

- Confidence intervals provide an alternative way of expressing the uncertainty in our estimates.
- Closely linked to the tests that we have already constructed.
- The confidence region provides a lot more information than a single hypothesis test in that it tells us the outcome of a whole range of hypotheses about the parameter values.
- The confidence region tells us about plausible values for the parameters in a way that the hypothesis test cannot. This makes it more valuable.

Confidence intervals

Confidence intervals constructed for each parameter individually take the general form of

$$\text{estimate} \pm \text{critical value} \times \text{s.e. of esitmate}$$

or specifically in this case

$$\hat{\beta}_i \pm t_{(1-\alpha/2),n-p}\hat{\sigma}\sqrt{(X^TX)^{-1}_{ii}}$$

These can be found using the `confint()` function in R.

Joint confidence intervals are sometimes used when the $\hat{\beta}$'s are heavily correlated.

Confidence intervals for predictions - prediction intervals

- Given a new set of predictors, $\boldsymbol{x_0}$ the predicted response can be fond with $\hat{y}_0 = \boldsymbol{x_0^T}\hat{\boldsymbol{\beta}}$.
- We need to distinguish between between predictions of the future mean response and predictions of future observations.
- Most times, we will want the latter case, the *prediction interval*, while the first case, the *confidence interval*, is less common.

Confidence intervals for predictions - prediction intervals

A $100 \cdot (1 - \alpha)$ % prediction interval is

$$\hat{y}_0 \pm t_{(1-\alpha/2),n-p}\hat{\sigma}\sqrt{1 + \boldsymbol{x_0^T}(\boldsymbol{X^T X})^{-1}\boldsymbol{x_0}}$$

and a $100 \cdot (1 - \alpha)$ % confidence interval is

$$\hat{y}_0 \pm t_{(1-\alpha/2),n-p}\hat{\sigma}\sqrt{\boldsymbol{x_0^T}(\boldsymbol{X^T X})^{-1}\boldsymbol{x_0}}$$

Identifiability

The least squares estimate is the solution to the normal equations:

$$\boldsymbol{X^T X \hat{\beta} = X^T y}$$

where X is an $n \times p$ matrix.

- If $X^T X$ singular and cannot be inverted there will be infinitely many solutions to the normal equations and $\hat{\boldsymbol{\beta}}$ is at least partially unidentifiable.
- Unidentifiability will occur when X is not of full rank, that is when its columns are linearly dependent.
- With observational data, unidentifiability is usually caused by some oversight.
- We will look more into unidentifiability when we talk about ANOVAs.

What can go wrong in MLR?

- Source and quality of the data
- Error component
- Structural Component

Interpreting Parameter Estimates

We have a multiple linear regression model:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_{(p-1)} x_{(p-1)}$$

what does $\hat{\beta}_1$ mean?

- $\hat{\beta}_1$ is the effect of x_1 when all other variables included in the model are held constant
- In stead of focusing on the values of the parameters an alternative approach is to recognize that the parameters and their estimates are fictional quantities in most regression situations.
- The "true" values may never be known (if they even exist in the first place).
- Instead, concentrate on predicting future values - these may actually be observed and success can then be measured in terms of how good the predictions were.