

stats545.3 545.3 Hypothesis tests in the linear model, model building and predictions

Gunnar Stefansson

5. september 2022

Copyright This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Efnisyfirlit

1	Linear hypotheses in multiple regression	4
1.1	Null hypotheses, matrices and geometry	4
1.1.1	Details	4
1.1.2	Examples	4
1.2	Null hypothesis as matrices	4
1.2.1	Details	5
1.2.2	Examples	5
1.3	Geometric comparisons of models	5
1.3.1	Details	5
1.3.2	Examples	5
1.4	Bases for the span of X	6
1.4.1	Details	6
1.5	Expected values of coefficients	7
1.5.1	Details	7
1.6	Sums of squares and norms	8
1.6.1	Details	8
1.7	Some probability distributions	8
1.7.1	Details	8
1.8	General F-tests in linear models	9
1.8.1	Details	9
2	Building a multiple regression model	10
2.1	Introduction	10
2.1.1	Details	10
2.2	Variable selection: Measuring quality	10
2.2.1	Details	10
2.2.2	Examples	10
2.3	Variable selection: Forward or backward	11
2.3.1	Details	11
2.3.2	Examples	11
2.4	Deleted residuals	12
2.4.1	Handout	12
3	Prediction in the linear model	13
3.1	Prediction and prediction uncertainty	13
3.1.1	Details	13
3.1.2	Examples	13
4	Estimable functions	14
4.1	Estimable functions: The problem	14
4.1.1	Details	14
4.1.2	Examples	14
4.1.3	Handout	15
4.2	Classification of estimable functions	16
4.2.1	Details	16
4.2.2	Examples	16
4.2.3	Handout	16
4.3	Gauss-Markov theorem	17
4.3.1	Details	17

4.3.2	Examples	17
4.3.3	Handout	17
4.4	Testing hypotheses in the linear model	18
4.4.1	Details	18
4.4.2	Handout	19
5	Ranks, constraints and correlations in multivariate regression	21
5.1	Problem statement	21
5.1.1	Details	21
5.2	Constraints	21
5.2.1	Details	21
5.2.2	Examples	21
5.2.3	Handout	22

1 Linear hypotheses in multiple regression

1.1 Null hypotheses, matrices and geometry

The null hypothesis, $H_i : \beta = 0$ in simple linear regression is a question of whether we can drop the variable x in $E[y_i] = \alpha + \beta x_i$, i.e. whether we can drop a column simplify \mathbf{X} to

$$\mathbf{Z} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

or is the projection of y onto $\text{span}(\mathbf{Z})$ is “too much” farther away from y than the projection onto $\text{span}(\mathbf{X})$.

General null hypotheses are almost always concerned with how one can “reduce” or simplify the model, in this case usually whether one can reduce the number of columns in \mathbf{X} or by some other means reduce the number of coefficients in the model.

1.1.1 Details

Tests of hypotheses in linear models can be considered geometrically. The hypothesis $H_i : \beta = 0$ in simple linear regression is the question of whether the matrix

$$\mathbf{Z} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

can be used in place of \mathbf{X} , i.e. whether the projection of y onto $\text{span}(\mathbf{Z})$ is too much farther away from y than the projection onto $\text{span}(\mathbf{X})$.

1.1.2 Examples

The null hypothesis, $H_i : \beta = 0$ in simple linear regression is a question of whether we can

use $\mathbf{Z} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ instead of \mathbf{X} .

In terms of testing, we want to see whether the projection of y onto $\text{span}(\mathbf{Z})$ is “too much” farther away from y than the projection onto $\text{span}(\mathbf{X})$.

1.2 Null hypothesis as matrices

Have $\underbrace{\mathbf{X}}_{n \times p}$ and $\underbrace{\mathbf{Z}}_{n \times q}$ s.t. $\text{span}(\mathbf{Z}) \subseteq \text{span}(\mathbf{X})$.

Can estimate models

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}_1$$

$$\mathbf{y} = \mathbf{Z}\gamma + \mathbf{e}_2$$

Will derive test for

$$H_0 : \mathbf{X}\beta = \mathbf{Z}\gamma$$

1.2.1 Details

Assume that $\underbrace{\mathbf{X}}_{n \times p}$ and $\underbrace{\mathbf{Z}}_{n \times q}$ are matrices such that $\text{span}(\mathbf{Z}) \subseteq \text{span}(\mathbf{X})$.

We can estimate coefficients in the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

and in the reduced model

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$$

We will derive tests for the general null hypothesis

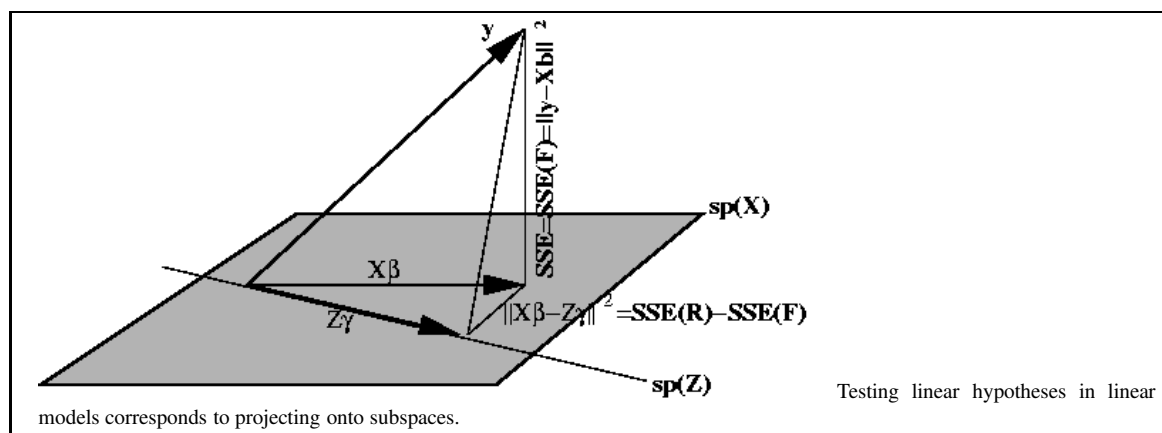
$$H_0 : \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma}$$

which is typically some hypothesis stating that some of the coefficients in the $\boldsymbol{\beta}$ -vector are zero or otherwise restricted.

1.2.2 Examples

Example 1.1. In simple linear regression, $y_i = \alpha + \beta x_i + e_i$, the most common test is for $\beta = 0$.

1.3 Geometric comparisons of models



1.3.1 Details

Relationships between sums of squares in two linear models is best viewed geometrically.

Starting with a base model as before, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, there is a need to investigate whether this model can be simplified in some manner. A simpler model can be denoted by $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$ where \mathbf{Z} is a matrix, typically with fewer columns than \mathbf{X} , and the column vectors of \mathbf{Z} span a subspace of that spanned by \mathbf{X} .

1.3.2 Examples

Example 1.2. A typical hypothesis test would start with a basic (full) model of the form $y_i = \alpha + \beta x_i + e_i$, wanting to test the null hypothesis $H_0 : \beta = 0$.

Define the matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}, \quad (1)$$

so the model in matrix notation becomes $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.

The null hypothesis can be written as $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$, where

$$\mathbf{Z} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}. \quad (2)$$

1.4 Bases for the span of \mathbf{X}

Orthonormal basis, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for \mathbf{R}^n :

Using Gram-Schmidt, first generate $\mathbf{u}_1, \dots, \mathbf{u}_q$ which span $sp\{\mathbf{Z}\}$, the next vectors, $\mathbf{u}_{q+1}, \dots, \mathbf{u}_r$ are chosen so that $\mathbf{u}_1, \dots, \mathbf{u}_r$ span $sp\{\mathbf{X}\}$, with $rank\{\mathbf{X}\} = r$, and the rest, $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ are chosen so that the entire set, $\mathbf{u}_1, \dots, \mathbf{u}_n$ spans \mathbf{R}^n .

$$\begin{aligned} \mathbf{Z}\hat{\boldsymbol{\gamma}} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q \\ \mathbf{X}\hat{\boldsymbol{\beta}} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q + \hat{\zeta}_{q+1} \mathbf{u}_{q+1} + \dots + \hat{\zeta}_r \mathbf{u}_r \\ \mathbf{y} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q + \hat{\zeta}_{q+1} \mathbf{u}_{q+1} + \dots + \hat{\zeta}_r \mathbf{u}_r + \hat{\zeta}_{r+1} \mathbf{u}_{r+1} + \dots + \hat{\zeta}_n \mathbf{u}_n \end{aligned}$$

1.4.1 Details

The probability distributions can best be viewed by defining a new orthonormal basis, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for \mathbf{R}^n . This basis is defined by first generating a set of r vectors $\mathbf{u}_1, \dots, \mathbf{u}_q$ which span the space defined by the null hypothesis, $sp\{\mathbf{Z}\}$, where $rank\{\mathbf{Z}\} = q$, subsequently the next vectors, $\mathbf{u}_{q+1}, \dots, \mathbf{u}_r$ are chosen so as to span the remainder of $sp\{\mathbf{X}\}$, where $rank\{\mathbf{X}\} = r$, and therefore $sp\{\mathbf{X}\} = sp\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, and the rest, $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ are chosen so that the entire set, $\mathbf{u}_1, \dots, \mathbf{u}_n$ spans \mathbf{R}^n . This is obviously always possible using the method of Gram-Schmidt.

This gives the following sequence of spaces and spans:

$$\begin{aligned} sp\{\mathbf{Z}\} &= sp\{\mathbf{u}_1, \dots, \mathbf{u}_q\} \\ sp\{\mathbf{X}\} &= sp\{\mathbf{u}_1, \dots, \mathbf{u}_q, \mathbf{u}_{q+1}, \dots, \mathbf{u}_r\} \\ \mathbf{R}^n &= sp\{\mathbf{u}_1, \dots, \mathbf{u}_q, \mathbf{u}_{q+1}, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_n\} \end{aligned}$$

One can then write each of $\mathbf{Z}\hat{\boldsymbol{\gamma}}$, $\mathbf{X}\hat{\boldsymbol{\beta}}$, \mathbf{y} in terms of the new basis as follows:

$$\begin{aligned} \mathbf{Z}\hat{\boldsymbol{\gamma}} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q \\ \mathbf{X}\hat{\boldsymbol{\beta}} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q + \hat{\zeta}_{q+1} \mathbf{u}_{q+1} + \dots + \hat{\zeta}_r \mathbf{u}_r \\ \mathbf{y} &= \hat{\zeta}_1 \mathbf{u}_1 + \dots + \hat{\zeta}_q \mathbf{u}_q + \hat{\zeta}_{q+1} \mathbf{u}_{q+1} + \dots + \hat{\zeta}_r \mathbf{u}_r + \hat{\zeta}_{r+1} \mathbf{u}_{r+1} + \dots + \hat{\zeta}_n \mathbf{u}_n \end{aligned}$$

where it is left to the reader to see that the $\hat{\zeta}_i$ -coefficients are indeed the same.

1.5 Expected values of coefficients

For $i = r + 1, \dots, n$ we obtain

$$E[\hat{\zeta}_i] = 0$$

If $H_0 : \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma}$ is true then for $i = q + 1, \dots, r$ we obtain

$$E[\hat{\zeta}_i] = \mathbf{u}_i \cdot (\mathbf{Z}\boldsymbol{\gamma}) = 0$$

1.5.1 Details

The expected values of the coefficients, $\hat{\zeta}_i$ depend on which space they correspond to.

Define

$$\zeta_i = E[\hat{\zeta}_i]$$

and by linearity we obtain

$$\zeta_i = E[\mathbf{u}_i \cdot \mathbf{y}] = \mathbf{u}_i \cdot (\mathbf{X}\boldsymbol{\beta}).$$

Now note that we have defined the basis vectors in three sets. The first is such that they span the same space as the columns of \mathbf{Z} . The second set complements the first to span the \mathbf{X} and the last set complements the set to span all of \mathbf{R}^n . The basis vectors are of course all orthogonal and each basis vector is orthogonal to all vectors in spaces spanned by preceding vectors.

For $i = r + 1, \dots, n$ we obtain

$$E[\hat{\zeta}_i] = \mathbf{u}_i \cdot (\mathbf{X}\boldsymbol{\beta}) = 0$$

since $\mathbf{X}\boldsymbol{\beta}$ is trivially in the space spanned by the column vectors of \mathbf{X} and is therefore a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_r$ and \mathbf{u}_i is orthogonal to all of these.

If the null hypothesis that $E[\mathbf{Y}]$ can be written as $H_0 : \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma}$ is true then for $i = q + 1, \dots, r$ we obtain

$$E[\hat{\zeta}_i] = \mathbf{u}_i \cdot (\mathbf{Z}\boldsymbol{\gamma}) = 0$$

but this only holds under the null hypothesis.

1.6 Sums of squares and norms

$$\begin{aligned}
 SSE(F) &= \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 &= \sum_{i=r+1}^n \hat{\zeta}_i^2 \\
 SSE(F) - SSE(R) &= \|\mathbf{Z}\hat{\boldsymbol{\gamma}} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 &= \sum_{i=q+1}^r \hat{\zeta}_i^2 \\
 SSE(R) &= \|\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2 &= \sum_{i=q+1}^n \hat{\zeta}_i^2
 \end{aligned}$$

1.6.1 Details

It is now quite easy to see how to form sums of squared deviations based on the new orthonormal basis, since each set of deviations corresponds to a specific portion of the space.

$$\begin{aligned}
 SSE(F) &= \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 &= \sum_{i=r+1}^n \hat{\zeta}_i^2 \\
 SSE(F) - SSE(R) &= \|\mathbf{Z}\hat{\boldsymbol{\gamma}} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 &= \sum_{i=q+1}^r \hat{\zeta}_i^2 \\
 SSE(R) &= \|\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2 &= \sum_{i=q+1}^n \hat{\zeta}_i^2
 \end{aligned}$$

Since each $\hat{\zeta}_i$ is a coordinate in an orthonormal basis, this is formed as an inner product with the corresponding basis vector, i.e. $\hat{\zeta}_i = \mathbf{y} \cdot \mathbf{u}_i$.

1.7 Some probability distributions

1.7.1 Details

Suppose we have two matrices, \mathbf{X} and \mathbf{Z} which satisfy $rank(\mathbf{Z}) = q < p = rank(\mathbf{X})$ and $sp(\mathbf{Z}) \subseteq sp(\mathbf{X})$ (usually \mathbf{Z} is $n \times q$ and \mathbf{X} is $n \times p$).

Then $H_0 : E[\mathbf{Y}] = \mathbf{Z}\boldsymbol{\gamma}$ is a reduction from the model $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$.

Write \mathbf{F} = full model and \mathbf{R} = for the reduced model.

Then we have

- 1) $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \perp \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}$
- 2) $\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$ and $\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2$ are independent
- 3) $\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\sigma^2} \sim \chi_{n-p}^2$ if the model is correct
- 4) $\frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2}{\sigma^2} \sim \chi_{p-q}^2$ if H_0 is correct.
- 5) $SSE(F) = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$

and

$$SSE(R) - SSE(F) = \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2$$

are independent.

- 6) $\frac{(SSE(R) - SSE(F))/(p-q)}{SSE(F)/(n-p)} \sim F_{p-q, n-p}$

Here F_{v_1, v_2} is the distribution of a ratio

$$F = \frac{U/v_1}{V/v_2}$$

of independent χ^2 -random variables, $U \sim \chi_{v_1}^2$, and $V \sim \chi_{v_2}^2$.

1.8 General F-tests in linear models

1.8.1 Details

In general one can compute the sum of squares from the full model, $SSE(F)$ as above and then compute the sum of squared deviations from the reduced model, $SSE(R) = \|\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\|^2$. Denote the corresponding degrees of freedom by $df(F)$ and $df(R)$, and assume that both matrices \mathbf{Z} and \mathbf{X} have full ranks, i.e. $rank(\mathbf{X}) = p$ and $rank(\mathbf{Z}) = r$. Then $df(F) = n - p$ and $df(R) = n - r$.

The null hypothesis can then be tested by noting that

$$F = \frac{(SSE(R) - SSE(F))/(p - r)}{SSE(F)/(n - p)} \quad (3)$$

is a realisation of a random variable from an F-distribution with $p - r$ and $n - p$ degrees of freedom under H_0 .

References Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill, Boston. 1408pp. **Copyright** 2021, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

2 Building a multiple regression model

2.1 Introduction

Have several independent variables
Want to select some into regression
Want to evaluate quality of resulting model
Want to improve into a final model

2.1.1 Details

Building multiple regression models includes several steps. It is, firstly, rarely pre-defined what independent variables should be included in the model, so a method for selecting these is needed. Having obtained an initial model one needs to evaluate not only the assumptions of the model but also identify possible influential observations and possibly undertake other diagnostics. Having obtained regression diagnostic, the model needs to be improved by taking these into account.

2.2 Variable selection: Measuring quality

- R^2
- AIC
- BIC
- SSE
- MSE
- P -values

2.2.1 Details

- R^2
- AIC
- BIC
- SSE
- MSE
- P -values

2.2.2 Examples

Example 2.1. Use the ecosystem data set and select a single variable in a simple linear regression to predict the growth of cod. Compare the various criteria.

2.3 Variable selection: Forward or backward

Model selection:

- All subset regression
- Forward stepwise regression
- Backwards stepwise regression

2.3.1 Details

Several methods exist to select a regression model.

All subset regression simply considers every possible combination of independent variables. Although this will indicate all possible “good” models and will certainly find the “best” model (using any given criterion), this is often not feasible.

Backwards stepwise regression starts by taking all independent variables into a single model and then dropping variables one at a time. The variable to be dropped is the one giving the least increase in SSE. This approach is often preferred, but is not feasible if the total number of variables are very large.

Forward stepwise regression selects a sequence of variables, at each stage deciding what variable to add next. The addition is based on including the variable giving the largest amount of (marginal) explained variation.

Forward stepwise regression is often augmented by allowing a variable to be dropped after a variable has been added. Thus a sequence of insertions may make an earlier variable redundant and thus dropped. Either version of forward regression is quite feasible but may lead to an incorrect or bad model since important combinations of variables may not be found.

Each approach thus has good and bad points.

2.3.2 Examples

Example 2.2. Use the ecosystem data set and conduct a forward stepwise regression to predict the growth of cod. Compare the various criteria for model selection.
R commands: `add1` repeatedly - followed by `anova(fm.final, fm.full)`

Example 2.3. Use the ecosystem data set and conduct a backwards stepwise regression to predict the growth of cod. Compare the various criteria for model selection.
R commands: `drop1` or `summary` - followed by `anova(fm.final, fm.full)`

2.4 Deleted residuals

Deleted residuals are based on the quantity

$$t_i = \frac{y_i - \hat{y}_{i(i)}}{s_{y_i - \hat{y}_{i(i)}}}$$

2.4.1 Handout

some formulas...

$$\sigma_{y_i - \hat{y}_{i(i)}}^2 = V[y_i - \hat{y}_{i(i)}]$$

$$\hat{y}_{i(i)} = \mathbf{x}_i \hat{\boldsymbol{\beta}}_{(i)}$$

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{X}'_{(i)} \mathbf{y}_{(i)}$$

$$T_i = \frac{y_i - \hat{y}_{i(i)}}{\hat{\sigma}_{y_i - \hat{y}_{i(i)}}}$$

$$\mathbf{y} - \hat{\mathbf{y}}_{(i)} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{(i)}$$

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{X}'_{(i)} \mathbf{y}_{(i)}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \mathbf{y}$$

$$\|\mathbf{y}_{(i)} - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)}\|^2 / \sigma^2 \sim \chi_{n-p-1}^2$$

$$\frac{e_i}{(\sigma \sqrt{1 - h_{ii}})} \sim n(0, 1)$$

References Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill, Boston. 1408pp.

Belsey, D. A., Kuh, E. and Welsh, R. E. 1980. Regression diagnostics: Identifying influential data and sources of collinearity. John. Wiley and Sons, New York. 292pp.

Copyright 2021, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

3 Prediction in the linear model

3.1 Prediction and prediction uncertainty

A new observation: \mathbf{x}_h
The prediction: $\hat{y}_h = E[y_h] = \mathbf{x}_h \hat{\beta}$
The variance: $\sigma^2 \mathbf{x}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_h$
Prediction uncertainty: $V[y_h - \hat{y}_h]$

3.1.1 Details

A new observation: \mathbf{x}_h
The prediction: $\hat{y}_h = E[\hat{y}_h] = \mathbf{x}_h \hat{\beta}$
The variance: $\sigma^2 \mathbf{x}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_h$
Prediction uncertainty: $V[y_h - \hat{y}_h]$

3.1.2 Examples

Example 3.1. Age and live weight of lambs. Project: Predict the weight (with uncertainty) at a given day.

```
days weight
135 39
125 35
120 33
126 38
125 37
137 38
133 36
140 41
130 38
129 36
123 34
132 40
129 38
121 34
126 35
137 44
121 34
137 41
130 39
137 43
```

Copyright 2021, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

4 Estimable functions

4.1 Estimable functions: The problem

If \mathbf{X} is not of full rank, then the LS problem does not have a unique solution for $\hat{\beta}$.

In general not all combinations of the form $\mathbf{c}'\hat{\beta}$ may have unique solutions.

A linear combination $\mathbf{c}'\beta$ is an **estimable function** if there is a vector of numbers, \mathbf{a} , such that

$$E[\mathbf{a}'\mathbf{y}] = \mathbf{c}'\beta$$

for all β .

NB: Viewed as a function of the unknown parameter vector, β .

NB: The E -operator depends on β , could define $g(\beta) := \mathbf{c}'\beta$ and $h(\beta) := E_{\beta}[\mathbf{a}'\mathbf{y}]$ and require $g(\beta) = h(\beta) \quad \forall \beta$ for some \mathbf{a} .

4.1.1 Details

If \mathbf{X} is not of full rank, then the LS problem does not have a unique solution for $\hat{\beta}$. In general not all combinations of the form $\mathbf{c}'\hat{\beta}$ may have unique solutions.

A linear combination $\mathbf{c}'\beta$ is an **estimable function** if there is a vector of numbers, \mathbf{a} , such that

$$E[\mathbf{a}'\mathbf{y}] = \mathbf{c}'\beta$$

for all β .

The terminology is not accidental as the linear combination of parameters is viewed as a function of the unknown parameter vector, β . In words the requirement is simply that it is possible to obtain a linear unbiased estimator.

4.1.2 Examples

Example 4.1. The one-way layout is the simplest example giving \mathbf{X} -matrices which are not of full rank when writing the model in the form

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}.$$

Example 4.2. A common issue in regression is whether the same line can be fit to two data sets or e.g. whether different slopes should be used. This can be modelled by writing

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij} \tag{4}$$

for the simple model with the same slopes and

$$y_{ij} = \alpha + \beta_i x_{ij} + \varepsilon_{ij} \tag{5}$$

for a model with different slopes in the the groups.

Alternatively one may be interested in how the slopes in the groups differ and/or in a simple evaluation of whether a single slope can be used. In this case it is reasonable to rewrite the complex model as

$$y_{ij} = \alpha + \beta x_{ij} + \beta_i x_{ij} + \varepsilon_{ij} \quad (6)$$

and the test of whether the reduced model is enough is a test of whether the β_i -values are all zero (and can be dropped).

Naturally, equation 6 is not completely determined. On the other hand, the model can easily be fit to data - most statistical packages will simply select an arbitrary LS estimate of the parameter set unless told explicitly to select a specific representation. All such solutions will lead to the same tests. The tests are really just based on comparing whether $SSE(R)$ is too much smaller than $SSE(F)$ and these sums are based on the LS projections onto subspaces. The projections are uniquely defined since they are based on the span, V , of the column vectors in the X -matrix. This space V does not change when columns are added, as long as these columns are linear combinations of existing ones - or when such columns are dropped.

Packages such as R will easily compare 6 and 4 with the `drop1`-command since 4 corresponds to deleting a term from 6. The better-determined model 5 can be compared to 4 using an `anova`-command in R since 4 is indeed a reduced model from 5 through a restriction of the form $\beta_1 = \dots = \beta_I$.

4.1.3 Handout

Some further clarifications to the above definitions may be useful.

It must be emphasized that the E -operator depends on the vectors of unknown parameters, β , since the underlying model is

$$E[\mathbf{y}] = \mathbf{X}\beta.$$

One could therefore define

$$f(\beta) := E_{\beta}[\mathbf{y}]$$

and

$$g(\beta) := c'\beta$$

so the criterion of estimability would be that these two functions are uniformly the same:

$$f(\beta) = g(\beta) \quad \forall \beta \in \mathbf{R}^p.$$

This formal approach has the merit that the meaning is clear, but the notation becomes quite cumbersome.

Estimable functions are commonly denoted by the symbol ψ , e.g. $\psi = \beta_1 - \beta_2$ etc.

Note 4.1. Recall that if $\underbrace{\mathbf{X}}_{n \times p}$ with $n > p$ is of full rank if $\text{rank}(\mathbf{X}) = \text{dim}(sp(\mathbf{X})) = p$ and also $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X})$ so \mathbf{X} is of full rank iff $\mathbf{X}'\mathbf{X}$ has an inverse. Hence, if \mathbf{X} is of full rank we can write $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ which satisfies $E[\hat{\beta}] = \beta$.

From this we also see that

$$\begin{aligned} \mathbf{c}'\beta &= \mathbf{c}'E[\hat{\beta}] \\ &= E[\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= E[\mathbf{a}'\mathbf{y}] \end{aligned}$$

where $\mathbf{a}' = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Hence any linear combination $\mathbf{c}'\beta$ is estimable if \mathbf{X} is of full rank.

Conversely, if \mathbf{X} is not of full rank then we can find vectors β and γ with $\beta \neq \gamma$ such that $\mathbf{X}(\beta - \gamma) = 0$ and therefore $E[\mathbf{y}] = \mathbf{X}\beta = \mathbf{X}\gamma$ can be expressed in more than one way.

Existence of non-estimable functions are therefore an expression of the matrix not being of full rank.

4.2 Classification of estimable functions

Theorem: A parametric function $\psi = \mathbf{c}'\beta$ is estimable if and only if $\mathbf{c}' = \mathbf{a}'\mathbf{X}$ for some $\mathbf{a} \in \mathbf{R}^n$.

4.2.1 Details

Theorem 4.1. A parametric function $\psi = \mathbf{c}'\beta$ is estimable if and only if $\mathbf{c}' = \mathbf{a}'\mathbf{X}$ for some $\mathbf{a} \in \mathbf{R}^n$.

4.2.2 Examples

Example 4.3. In the linear model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, the coefficients are not all estimable.

4.2.3 Handout

Proof of theorem: By definition, $\psi = \mathbf{c}'\beta$ is estimable if and only if there is a vector $\mathbf{a} \in \mathbf{R}^n$ such that $E[\mathbf{a}'\mathbf{y}] = \mathbf{c}'\beta$ for all β . This is equivalent to requiring

$$\mathbf{a}'\mathbf{X}\beta = \mathbf{c}'\beta$$

for all β which is equivalent to

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}.$$

Example 4.4. The reader should take the simple example $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $1 \leq j \leq J_i$, $1 \leq i \leq I$, set up the \mathbf{X} -matrix and consider the form of the vectors $\mathbf{a}'\mathbf{X}$ for the case $I = 2$, $J_1 = n$, $J_2 = m$.

Writing $\mathbf{a}'\mathbf{X} = (u, v, u - v)$, it is seen that the resulting estimable functions are precisely $\mu + \alpha_1$, $\mu + \alpha_2$ and $\alpha_1 - \alpha_2$.

4.3 Gauss-Markov theorem

Theorem: (Gauss-Markov theorem): Let $EY = X\beta$, $VY = \sigma^2 I$. Then every estimable function $c'\beta$ has a unique unbiased linear estimate which has minimum variance in the class of all unbiased linear estimates. This estimate can be written the form $c'\hat{\beta}$ where $\hat{\beta}$ is any LS estimator.

4.3.1 Details

A fundamental result in the theory of linear models is that estimable functions have unique unbiased linear estimates.

Lemma: If $\psi = c'\beta$ is estimable and $V = sp(\mathbf{X})$ then there is a unique linear unbiased estimator of $c'\beta$ of the form $\mathbf{a}'\mathbf{y}$ with $\mathbf{a} \in V$. If $\mathbf{a}_0\mathbf{y}$ is unbiased for $c'\beta$ then \mathbf{a} is the projection of \mathbf{a}_0 onto V .

Theorem 4.2 (Gauss-Markov theorem). Let $E[\mathbf{y}] = \mathbf{X}\beta$, $V[\mathbf{y}] = \sigma^2 I$. Then every estimable function $c'\beta$ has a unique unbiased linear estimate which has minimum variance in the class of all unbiased linear estimates. This estimate can be written the form $c'\hat{\beta}$ where $\hat{\beta}$ is any LS estimator.

Note 4.2. For estimable functions this is defined as the LS estimator.

4.3.2 Examples

Example 4.5. In the model $y_{ik} = \mu + \alpha_i + e_{ik}$, it is clear that parameters are not estimable but it is easy to see that $\alpha_i - \alpha_j$ are estimable.

4.3.3 Handout

Proof of lemma: Suppose $\psi = c'\beta$ is estimable so we can find $\mathbf{a} \in \mathbf{R}^n$ such that $E[\mathbf{a}'\mathbf{y}] = \psi$. Now write $\mathbf{a} = \mathbf{a}^* + \mathbf{b}^*$ with $\mathbf{a}^* \in V$ and $\mathbf{b}^* \perp V$, i.e. we define \mathbf{a}^* as the projection of \mathbf{a} onto V . Then it is easy to see that $E\mathbf{b}^{*\prime}\mathbf{y} = 0$ since \mathbf{b}^* is perpendicular to the columns of \mathbf{X} , all of which are in V . Hence $\psi = E\mathbf{a}'\mathbf{y} = E\mathbf{a}^{*\prime}\mathbf{y}$ and hence $\mathbf{a}^{*\prime}\mathbf{y}$ is unbiased for ψ and $\mathbf{a}^* \in V$.

For uniqueness of \mathbf{a}^* , suppose $E\alpha'\mathbf{y} = \psi$ for some $\alpha \in V$. Then $0 = E\mathbf{a}^{*\prime}\mathbf{y} - E\alpha'\mathbf{y} = (\mathbf{a}^* - \alpha)'\mathbf{X}\beta$. This holds for all $\beta \in R^p$ and hence $(\mathbf{a}^* - \alpha)'\mathbf{X} = \mathbf{0}$. Since $(\mathbf{a}^* - \alpha)$ is perpendicular to all columns of the \mathbf{X} -matrix, it follows that $(\mathbf{a}^* - \alpha) \in V^\perp$. But both vectors were in V to begin with, so

$$(\mathbf{a}^* - \alpha) \in V \cap V^\perp = \{0\}$$

i.e. $\mathbf{a}^* = \alpha$ so \mathbf{a}^* is unique. Since \mathbf{a}^* was taken as the projection of \mathbf{a} onto V , the proof is complete.

Proof of Gauss-Markov theorem: Use the lemma to find a unique $\mathbf{a}^* \in V$ with $E\mathbf{a}^{*\prime}\mathbf{y} = \mathbf{c}'\boldsymbol{\beta}$ and let $\mathbf{a}'\mathbf{y}$ be any unbiased linear estimate of $\boldsymbol{\psi}$. Then $\mathbf{a}^* = \text{proj}_V(\mathbf{a})$ and $\|\mathbf{a}\|^2 = \|\mathbf{a}^*\|^2 + \|\mathbf{a} - \mathbf{a}^*\|^2$ so

$$\begin{aligned} V\mathbf{a}'\mathbf{y} &= \mathbf{a}'\Sigma_y\mathbf{a} = \sigma^2\|\mathbf{a}\|^2 \\ &= \sigma^2\|\mathbf{a}^*\|^2 + \sigma^2\|\mathbf{a} - \mathbf{a}^*\|^2 = V\mathbf{a}^*\mathbf{y} + \sigma^2\|\mathbf{a} - \mathbf{a}^*\|^2 \geq V\mathbf{a}^*\mathbf{y} \end{aligned}$$

and equality holds iff $\mathbf{a} = \mathbf{a}^*$ so $\mathbf{a}^*\mathbf{y}$ is best.

Now let $\hat{\boldsymbol{\beta}}$ be any least squares estimate.

Note 4.3. Note that $\mathbf{a}^* \in V$ and $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \in V^\perp$ so that $\mathbf{a}^*(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$ and therefore $\mathbf{a}^*\mathbf{y} = \mathbf{a}^*\mathbf{X}\hat{\boldsymbol{\beta}}$.

Further, since $\boldsymbol{\psi} = \mathbf{c}'\boldsymbol{\beta}$ is estimable and $\mathbf{a}^{*\prime}\mathbf{y}$ is the unbiased linear estimate,

$$\mathbf{c}'\boldsymbol{\beta} = E\mathbf{a}^{*\prime}\mathbf{y} = \mathbf{a}^{*\prime}\mathbf{X}\boldsymbol{\beta}$$

and this holds for any $\boldsymbol{\beta} \in \mathbf{R}^p$ so $\mathbf{c}' = \mathbf{a}^{*\prime}\mathbf{X}$. Combining this with the previous paragraph,

$$\mathbf{a}^{*\prime}\mathbf{y} = \mathbf{a}^{*\prime}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{c}'\hat{\boldsymbol{\beta}}$$

which concludes the proof.

4.4 Testing hypotheses in the linear model

Theorem: If $\underbrace{\mathbf{y}}_{n \times 1} \sim n(\underbrace{\mathbf{X}}_{n \times p}, \underbrace{\boldsymbol{\beta}}_{p \times 1}, \underbrace{\sigma^2 \mathbf{I}}_{n \times n})$ and $\hat{\boldsymbol{\psi}}$ is a vector of estimable functions, then $\hat{\boldsymbol{\psi}} \sim n(\boldsymbol{\psi}, \Sigma_{\hat{\boldsymbol{\psi}}})$, $\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\sigma^2} \sim \chi_{n-r}^2$ and these two quantities are independent.

4.4.1 Details

Let $\underbrace{\mathbf{y}}_{n \times 1} \sim n(\underbrace{\mathbf{X}}_{n \times p}, \underbrace{\boldsymbol{\beta}}_{p \times 1}, \underbrace{\sigma^2 \mathbf{I}}_{n \times n})$
and assume $\text{rank}(\mathbf{X}) = r \leq p$.

The interest will be in obtaining some joint confidence statement on a vector, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)'$, where each $\psi_i = \mathbf{c}'_i\boldsymbol{\beta}$ is an estimable function. Write $\hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \dots, \hat{\psi}_q)'$ for the least squares estimates with $\hat{\psi}_i = \mathbf{c}'_i\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is any LS estimate and one can therefore also write $\hat{\psi}_i = \mathbf{a}_i\mathbf{y}$ for unique $\mathbf{a}_i \in \text{sp}(\mathbf{X})$.

The above can be written more concisely as $\boldsymbol{\psi} = \mathbf{C}\boldsymbol{\beta}$ using obvious definitions. It follows that

$$\hat{\boldsymbol{\psi}} = \mathbf{A}\mathbf{y} = \mathbf{C}\hat{\boldsymbol{\beta}} \sim n(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{A}\mathbf{A}')$$

and the variance-covariance matrix of the estimates will be denoted

$$V[\hat{\boldsymbol{\psi}}] = \Sigma_{\hat{\boldsymbol{\psi}}}$$

which leads to the following theorem.

Theorem 4.3. If $\mathbf{y} \sim n(\underbrace{\mathbf{X}}_{n \times p}, \underbrace{\boldsymbol{\beta}}_{p \times 1}, \sigma^2 \underbrace{\mathbf{I}}_{n \times n})$ and $\hat{\boldsymbol{\psi}}$ is a vector of estimable functions, then $\hat{\boldsymbol{\psi}} \sim n(\boldsymbol{\psi}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}})$, $\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{\sigma^2} \sim \chi_{n-r}^2$ and these two quantities are independent.

It follows that hypothesis tests can be constructed in an obvious manner for individual estimable functions.

4.4.2 Handout

Proof: Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be an orthonormal basis for \mathbb{R}^n such that $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ form a basis for $sp(\mathbf{X})$ and let $\hat{\zeta}_1, \dots, \hat{\zeta}_n$ be the coordinates of \mathbf{y} in this basis, so that $\hat{\zeta}_i = \mathbf{u}_i \cdot \mathbf{y}$. Also define $\zeta_i = E[\hat{\zeta}_i]$. It is established that $\hat{\zeta}_i$ are independent, Gaussian with common variance σ^2 .

Write $\mathbf{z} = (\hat{\zeta}_1, \dots, \hat{\zeta}_n)'$, $\mathbf{Q} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n \end{bmatrix}$ and note that \mathbf{Q}' has the \mathbf{u}_i' as row vectors.

Since $\hat{\zeta}_i = \mathbf{u}_i' \mathbf{y}$ we have

$$\mathbf{z} = \mathbf{Q}' \mathbf{y}.$$

Since the \mathbf{u}_i form a basis and form the columns of \mathbf{Q} , those same columns are independent so \mathbf{Q} is invertible (as is \mathbf{Q}').

Due to the nature of an orthonormal basis, $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ so the two are inverses of each other and $\mathbf{Q}\mathbf{Q}' = \mathbf{I}$. Further, $\mathbf{z} = \mathbf{Q}'\mathbf{y}$ and therefore $\mathbf{y} = \mathbf{Q}\mathbf{z}$.

As elsewhere, write the LS estimates of the estimable functions in the form $\hat{\boldsymbol{\psi}}_i = \mathbf{a}_i' \mathbf{y}$ where $\mathbf{a}_i \in V = sp\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ so that $\hat{\boldsymbol{\psi}}_i = \mathbf{a}_i' \mathbf{Q}\mathbf{z}$. It follows that $\mathbf{a}_i' \mathbf{Q} = \begin{bmatrix} \mathbf{a}_i' \mathbf{u}_1 & \dots & \mathbf{a}_i' \mathbf{u}_n \end{bmatrix}$ and of these various inner products, $\mathbf{a}_i' \mathbf{u}_j = 0$ if $j > r$ (since $\mathbf{a}_i \in V$) from which it is seen that

$$\mathbf{a}_i' \mathbf{Q}\mathbf{z} = \begin{bmatrix} \mathbf{a}_i' \mathbf{u}_1 & \dots & \mathbf{a}_i' \mathbf{u}_r & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \hat{\zeta}_1 & \dots & \hat{\zeta}_r & \hat{\zeta}_{r+1} & \dots & \hat{\zeta}_n \end{bmatrix}' = \mathbf{a}_i' \mathbf{u}_1 \hat{\zeta}_1 + \dots + \mathbf{a}_i' \mathbf{u}_r \hat{\zeta}_r$$

i.e. the estimable functions are all formed from the first r of the $\hat{\zeta}_i$ and are all of the form

$$\hat{\boldsymbol{\psi}}_i = \sum_1^r k_j \hat{\zeta}_j \quad (7)$$

for some constants k_1, \dots, k_r .

This important result is quite general and basically states that anything that can be estimated can be derived from \mathbf{y} through the column vectors of the \mathbf{X} -matrix.

On the other hand it is also known that $\mathbf{X}\hat{\boldsymbol{\beta}}$ is the projection of \mathbf{y} onto the space spanned by $\mathbf{u}_1, \dots, \mathbf{u}_r$ and therefore the residual, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is in the span of $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ and in fact

$$\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \sum_{j=r+1}^n \hat{\zeta}_j^2 \quad (8)$$

All the results in the theorem follow easily from (7) and (8).

References Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill, Boston. 1408pp.

Scheffe, H. 1959. The analysis of variance. John Wiley and Sons, Inc, New York. 477pp.

Copyright 2021, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

5 Ranks, constraints and correlations in multivariate regression

5.1 Problem statement

When $\text{rank}(\mathbf{X}) < p$, the estimate $\hat{\beta}$ is not unique.

5.1.1 Details

When $r = \underbrace{\text{rank}(\mathbf{X})}_{n \times p} < p$, the estimate $\hat{\beta}$ is not unique. Similarly, β in $E[\mathbf{Y}] = \mathbf{X}\beta$ is not unique.¹

But if the function $\psi = \mathbf{c}'\beta$ is estimable, then the number $\mathbf{c}'\beta$ is unique, i.e. the same for all β in the set $\{b : E\mathbf{Y} = \mathbf{X}b\}$, since $\mathbf{X}\beta$ is unique and $\mathbf{c}'\beta = \mathbf{a}'\mathbf{X}\beta$ for some \mathbf{a} .

5.2 Constraints

To specify β uniquely we can add constraints...

5.2.1 Details

In order to specify the vector β and $\hat{\beta}$ one could simply drop some of these until the \mathbf{X} -matrix becomes of full rank. More generally it is possible to add constraints of the form $\underbrace{\mathbf{H}}_{t \times p} \beta = \mathbf{0}$

This can be formulated in the following manner: Suppose we have β and we want unique $\tilde{\beta}$ through $\mathbf{X}\beta = \mathbf{X}\tilde{\beta}$ and $\mathbf{H}\tilde{\beta} = \mathbf{0}$.

Theorem 5.1. $\tilde{\beta}$ is unique if $\text{rank}\left(\begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix}\right) = p$ and $\tilde{\beta}$ are then estimable.²

The reader is referred to Scheffe (1959) for the proof of the theorem.

5.2.2 Examples

Example 5.1. If $Y_{ik} \sim n(\mu + \alpha_i, \sigma^2)$, independent, with $1 \leq k \leq n_i$ and $1 \leq i \leq I$, then one can use the constraints $\sum \alpha_i = 0$.

It is a useful exercise to write the \mathbf{X} -matrix and \mathbf{H} -matrix for this problem.

¹This is easy to see since if ξ_1, \dots, ξ_p are the **columns** of the \mathbf{X} -matrix then $E[\mathbf{Y}] = \mathbf{X}\beta$ is a linear combination of ξ_1, \dots, ξ_p which only span a r -dimensional space and a subset of ξ_1, \dots, ξ_p can be used to span this space. The vector $E[\mathbf{Y}]$ can be written as a linear combination of vectors in any such subset.

5.2.3 Handout

Write $\mathbf{G} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix}$ for the joint data and constraint matrices.

Note that we obtain

$$\mathbf{G}\tilde{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}$$

and thus

$$\mathbf{G}'\mathbf{G}\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

where $\mathbf{G}'\mathbf{G}$ is invertible and we can write

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1} \mathbf{X}'\mathbf{X}\boldsymbol{\beta} \quad (9)$$

and we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1} \mathbf{X}'\mathbf{y}$$

an unbiased estimate.

Note 5.1. The vector $\tilde{\boldsymbol{\beta}}$ defined in Eq. (9) is a vector of elements, each of which is a linear function of $\boldsymbol{\beta}$ and each of these functions is estimable since each is of the form $\mathbf{a}'\mathbf{X}\boldsymbol{\beta}$.

References ISBN: 0256117365

ISBN: 0471758345 **Copyright** 2021, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.