Linear hypotheses in multiple regression

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

Gunnar Stefansson

November 28, 2020

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 $span(\mathbf{Z})$ is "too much" farther away from y than the projection onto $span(\mathbf{X})$.

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 X or by some other means reduce the number of coefficients in the model.

Null hypothesis as matrices

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

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

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

Will derive test for

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

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

Geometric comparisons of models

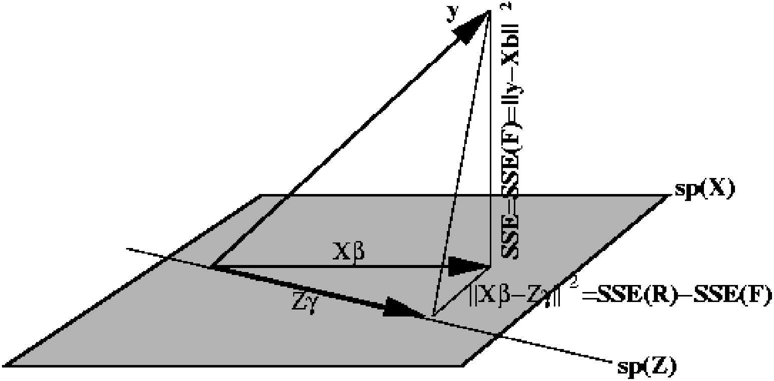

Figure : Testing linear hypotheses in linear models corresponds to projecting onto subspaces.

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 $\mathbf{y} = \mathbf{Z}\gamma + \mathbf{e}$ 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 $\mathbf{X}$.
Example:

Bases for the span of X

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

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

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

Expected values of coefficients

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

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

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

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

Sums of squares and norms

$$SSE(F) = ||\mathbf{y} - \mathbf{X}\hat{\beta}||^2 = \sum_{i=r+1}^{n} \hat{\zeta}_i^2$$

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

$$SSE(R) = ||\mathbf{y} - \mathbf{Z}\hat{\gamma}||^2 = \sum_{i=q+1}^{n} \hat{\zeta}_i^2$$

Each $\hat{\zeta}_i$ is a coordinate in an orthonormal basis, $\hat{\zeta}_i = \mathbf{y} \cdot \mathbf{u}_i$. When Y_i are independent Gaussian random variables, $\hat{\zeta}_i$ also become independent. As a result, the sums of squares are related to $\sigma^2 \cdot \chi^2$-distributions in a natural way.

Some probability distributions

Matrices, $\mathbf{X}$, $\mathbf{Z}$ with $rank(\mathbf{Z}) = q < r = rank(\mathbf{X})$
and $sp(\mathbf{Z}) \subseteq sp(\mathbf{X})$ ($\mathbf{Z}$ may be $n \times q$ and $\mathbf{X}$ $n \times p$ w/$p = r$).
$H_0 : E[\mathbf{Y}] = \mathbf{Z}\gamma$ is a reduction of $E[\mathbf{Y}] = \mathbf{X}\beta$.
If $\mathbf{F}$ = full model and $\mathbf{R}$ = reduced then

1) $y - \mathbf{X}\hat{\beta} \perp \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}$

2) $||y - \mathbf{X}\hat{\beta}||^2$ and $||\mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}||^2$ are independent

3) $\frac{||y - \mathbf{X}\hat{\beta}||^2}{\sigma^2} \sim \chi^2_{n-r}$ if the model is correct

4) $\frac{||\mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}||^2}{\sigma^2} \sim \chi^2_{r-q}$ if H_0 is correct.

3) $\frac{||y - \mathbf{X}\hat{\beta}||^2}{\sigma^2} \sim \chi^2_{n-p}$ if the model is correct

4) $\frac{||\mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}||^2}{\sigma^2} \sim \chi^2_{p-q}$ if H_0 is correct.

5) $SSE(F) = ||y - \mathbf{X}\hat{\beta}||^2$
and
$SSE(R) - SSE(F) = ||\mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}||^2$
are independent.

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

Here F_{ν_1,ν_2} is the distribution of a ratio

$$F = \frac{U/\nu_1}{V/\nu_2}$$

General F-tests in linear models

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, $SSE(R) = ||\mathbf{y} - \mathbf{Z}\hat{\gamma}||^2$. 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. $rank(\mathbf{X}) = r$ og $rank(\mathbf{Z}) = q$.
Then $df(F) = n - r$ and $df(R) = n - q$.
The null hypothesis can then be tested by noting that

$$F = \frac{(SSE(R) - SSE(F))/(r - q)}{SSE(F)/(n - r)}$$

is a realisation of a random variable from an F-distribution with $r - q$ and $n - r$ 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.