A Framework for Hypothesis Tests in Statistical Models With Linear Predictors

Georges Monette[1] John Fox[2]

[1]York University
Toronto, Ontario, Canada

[2]McMaster University
Hamilton, Ontario, Canada

useR 2009 Rennes

Basic Results

General Setting

- We have an estimator $\mathbf{b}$ of the $p \times 1$ parameter vector $\boldsymbol{\beta}$.
- $\mathbf{b}$ is asymptotically multivariate-normal, with asymptotic expectation $\boldsymbol{\beta}$ and estimated asymptotic positive-definite covariance matrix $\mathbf{V}$.
- In the applications that we have in mind, $\boldsymbol{\beta}$ appears in a linear predictor $\eta = \mathbf{x}'\boldsymbol{\beta}$, where $\mathbf{x}'$ is a "design" vector of regressors.

Basic Results

Linear Hypotheses

- We address linear hypotheses of the form H_1: $\boldsymbol{\psi}_1 = \mathbf{L}_1 \boldsymbol{\beta} = \mathbf{0}$, where the $k_1 \times p$ hypothesis matrix $\mathbf{L}_1$ of rank $k_1 \leq p$ contains pre-specified constants and $\mathbf{0}$ is the $k_1 \times 1$ zero vector.
- As is well known, the hypothesis H_1 can be tested by the Wald statistic

$$Z_1 = (\mathbf{L}_1 \mathbf{b})' (\mathbf{L}_1 \mathbf{V} \mathbf{L}_1')^{-1} \mathbf{L}_1 \mathbf{b},$$

which is asymptotically distributed as chi-square with k_1 degrees of freedom.

Basic Results

Nested Linear Hypotheses

- Consider another hypothesis H_2: $\boldsymbol{\psi}_2 = \mathbf{L}_2\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{L}_2$ has $k_2 < k_1$ rows and is of rank k_2, and $\mathbf{0}$ is the $k_2 \times 1$ zero vector.
- Hypothesis H_2 is *nested within* the hypothesis H_1 if and only if the rows of $\mathbf{L}_2$ lie in the space spanned by the rows of $\mathbf{L}_1$.
    - Then the truth of H_1 (which is more restrictive than H_2) implies the truth of H_2, but not vice-versa.
    - Typically the rows of $\mathbf{L}_2$ will be a proper subset of the rows of $\mathbf{L}_1$.
- The *conditional hypothesis* $H_{1|2}$ is that $\mathbf{L}_1\boldsymbol{\beta} = \mathbf{0} \mid \mathbf{L}_2\boldsymbol{\beta} = \mathbf{0}$.

Basic Results

Testing Nested Hypotheses: Wald Test

- $H_{1|2}$ can be tested by the Wald statistic

$$Z_{1|2} = (\mathbf{L}_{1|2}\mathbf{b})'(\mathbf{L}_{1|2}\mathbf{V}\mathbf{L}_{1|2}')^{-1}\mathbf{L}_{1|2}\mathbf{b},$$

  - $\mathbf{L}_{1|2}$ is the conjugate complement of the projection of the rows of $\mathbf{L}_2$ into the row space of $\mathbf{L}_1$ with respect to the inner product $\mathbf{V}$.
  - The conditional Wald statistic $Z_{1|2}$ is asymptotically distributed as chi-square with $k_1 - k_2$ degrees of freedom.

Basic Results

Testing Nested Hypotheses: F Test

- In some models, such as a generalized linear model with a dispersion parameter estimated from the data, we can alternatively compute an F-test of $H_{1|2}$ as

$$F_{1|2} = \frac{1}{k_1 - k_2} (\mathbf{L}_{1|2}\mathbf{b})' (\mathbf{L}_{1|2}\mathbf{V}\mathbf{L}'_{1|2})^{-1} \mathbf{L}_{1|2}\mathbf{b}.$$

- If tests for all terms of a linear model are formulated in conformity with the principle of marginality, the conditional F-test produces so-called "Type-II" hypothesis tests.

Basic Results

Sketch of Justification

- Let $\mathbf{L}_*$ be any $r \times p$ matrix whose rows extend the row space of $\mathbf{L}_2$ to the row space of $\mathbf{L}_1$ (i.e., $r = k_1 - k_2$),
- The hypothesis

$$H_*\colon\ \boldsymbol{\psi}_* = \mathbf{L}_*\boldsymbol{\beta} = \mathbf{0} \,|\, H_2\colon\ \boldsymbol{\psi}_2 = \mathbf{L}_2\boldsymbol{\beta} = \mathbf{0}$$

is equivalent to the hypothesis

$$H_1\colon\ \mathbf{L}_1\boldsymbol{\beta} = \mathbf{0} \,|\, H_2\colon\ \mathbf{L}_2\boldsymbol{\beta} = \mathbf{0}$$

and independent of the particular choice of $\mathbf{L}_*$.

Basic Results

Sketch of Justification

- The minimum-variance asymptotically unbiased estimator of $\boldsymbol{\psi}_*$ under the conditional null hypothesis is

$$\widehat{\boldsymbol{\psi}}_*^C = \mathbf{L}_*\mathbf{b} - \mathbf{L}_*\mathbf{V}\mathbf{L}_2' \left(\mathbf{L}_2\mathbf{V}\mathbf{L}_2'\right)^{-1}\mathbf{L}_2\mathbf{b} = \mathbf{L}_{*|2}\mathbf{b}$$

where

$$\mathbf{L}_{*|2} = \mathbf{L}_* - \mathbf{L}_*\mathbf{V}\mathbf{L}_2' \left(\mathbf{L}_2\mathbf{V}\mathbf{L}_2'\right)^{-1}\mathbf{L}_2$$

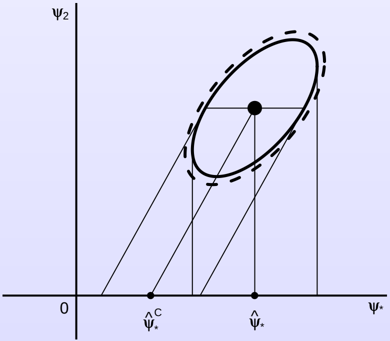
- Thus the test of $H_{1|2}$ is based on the statistic

$$Z_{1|2} = \widehat{\boldsymbol{\psi}}_*^{C\prime} \left(\mathbf{L}_{*|2}\mathbf{V}\mathbf{L}_{*|2}'\right)^{-1} \widehat{\boldsymbol{\psi}}_*^C$$

which is asymptotically distributed as chi-square with r degrees of freedom under H_1 given H_2.

Basic Results

Geometric Interpretion



- If $\mathbf{L}_*$ and $\mathbf{L}_2$ are $1 \times p$, then the 2D confidence ellipse for $\boldsymbol{\psi} = [\psi_*, \psi_2]' = \mathbf{L}_1 \boldsymbol{\beta}$ is based on the estimated asymptotic variance $\widehat{\text{AsyVar}}(\widehat{\boldsymbol{\psi}}) = \mathbf{L}_1 \mathbf{V} \mathbf{L}_1'$.
- The unrestricted estimator $\widehat{\psi}_*$ is the perpendicular projection of $\widehat{\boldsymbol{\psi}} = \left[\widehat{\psi}_*, \widehat{\psi}_2\right]' = \mathbf{L}_1 \mathbf{b}$ onto the ψ_* axis.
- $\widehat{\psi}_*^C$ is the oblique projection of $\widehat{\boldsymbol{\psi}}$ onto the ψ_* axis along the direction conjugate to the ψ_* axis with respect to the inner product $\left(\mathbf{L}_1 \mathbf{V} \mathbf{L}_1'\right)^{-1}$.

Basic Results

Geometric Interpretion

- The dashed ellipse is the asymptotic 2D confidence ellipse,

$$\mathcal{E}_2 = \widehat{\boldsymbol{\psi}} + \sqrt{\chi^2_{.95;2}}\left(\mathbf{L}_1\mathbf{V}\mathbf{L}_1'\right)^{1/2}\mathcal{U}$$

where $\mathcal{U}$ is the unit-circle and $\chi^2_{.95;2}$ is the .95 quantile of the chi-square distribution with two degrees of freedom.

Basic Results

Geometric Interpretion

- The solid ellipse

$$\mathcal{E}_1 = \widehat{\boldsymbol{\psi}} + \sqrt{\chi^2_{.95;1}} \left(\mathbf{L}_1 \mathbf{V} \mathbf{L}_1' \right)^{1/2} \mathcal{U}$$

is generated by changing the degrees of freedom to one.

  - one-dimensional projections of $\mathcal{E}_1$ are ordinary confidence intervals for linear combinations of $\boldsymbol{\psi} = [\psi_*, \psi_2]'$.
  - Under H_2, all projections onto the ψ_* axis are unbiased estimators of ψ_* with 95% confidence intervals given by the corresponding projection of the solid ellipse.
  - The projection in the direction conjugate to the ψ_* axis — that is, along the line through the center of the confidence ellipse and through the points on the ellipse with horizontal tangents — yields the confidence interval with the smallest width.

Examples

Dummy Regression

- Suppose, for example, that we are interested in a dummy-regression model with linear predictor

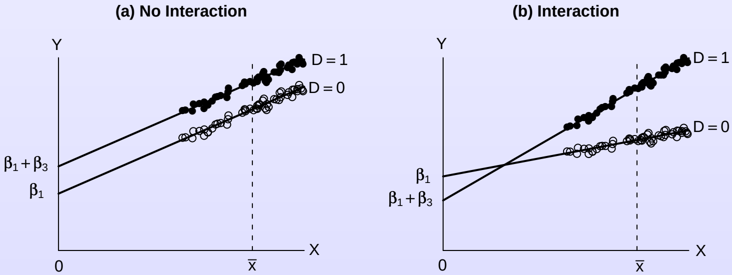
$$\eta = \beta_1 + \beta_2 x + \beta_3 d + \beta_4 xd$$

where x is a covariate and d is a dummy regressor, taking on the values 0 and 1.

- Then the hypotheses H_2: $\beta_4 = 0$ (that there is no interaction between x and d) is nested within the hypothesis H_1: $\beta_3 = \beta_4 = 0$ (that there is neither interaction between x and d nor a "main effect" of d).

Examples

Dummy Regression



Examples

Dummy Regression

- In this case we have

$$\mathbf{L}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{L}_2 = [0, 0, 0, 1]$$

- The conditional hypothesis $H_{1|2}$: $\beta_3 = \beta_4 = 0 \,|\, \beta_4 = 0$ can be restated as $H_{1|2}$: $\beta_3 = 0 \,|\, \beta_4 = 0$ — that is, the hypothesis of no main effect of d assuming no interaction between x and d.
  - Here $\boldsymbol{\psi}_1 = [\beta_3, \beta_4]'$, $\psi_2 = \beta_4$, and $\psi^* = \beta_3$.

Example

Dummy Regression

- This example also illustrates why conditional ("Type II") hypotheses are potentially of interest in models where some terms are marginal to others:
  - The *unconditional* ("Type-III") hypothesis H_0: $\beta_3 = 0$ pertains to the partial effect of d above the origin (i.e., where $x = 0$).
  - If $\beta_4 \neq 0$, then this is not reasonably interpretable as a hypothesis about the main effect of d, and may, indeed, be of no interest at all (when, for example, the values of x are all far from 0).
  - If $\beta_4 = 0$ and the centre of the data is far from $x = 0$, then the unconditional test will have low power.
  - The interpretability and performance of the unconditional test can be improved by centering the x at $\overline{x}$.

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- Data on measured and reported weight from the Davis dataset in the **car** package.

```
> library(car)
> mod.davis <- lm(repwt ~ weight*sex, data=Davis)
> summary(mod.davis)

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.34116    1.87515   1.782   0.0765 .
weight       0.93314    0.03253  28.682   <2e-16 ***
sexM        -1.98252    2.45028  -0.809   0.4195
weight:sexM  0.05668    0.03845   1.474   0.1422
```

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- "Type-II" tests with White-Huber coefficient covariances:

```
> Anova(mod.davis, white=TRUE)

Anova Table (Type II tests)

Response: repwt
           Df         F    Pr(>F)
weight      1 2165.7754 < 2.2e-16 ***
sex         1   15.1678 0.0001388 ***
weight:sex  1    1.8684 0.1733720
Residuals 179
```

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- "Type-III" tests with White-Huber coefficient covariances:

```
> Anova(mod.davis, white=TRUE, type=3)

Anova Table (Type III tests)

Response: repwt
             Df         F  Pr(>F)
(Intercept)   1    4.4271 0.03677 *
weight        1 1148.9590 < 2e-16 ***
sex           1    0.5196 0.47197
weight:sex    1    1.8684 0.17337
Residuals   179
```

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- Refitting with a centered covariate and sigma-contrained contrast for `sex`

```
> Davis$cweight <- with(Davis, weight - mean(weight))
> mod.davis.2 <- lm(repwt ~ cweight*sex, data=Davis,
+      contrasts=list(sex=contr.sum))
> summary(mod.davis.2)
```

```
(Intercept)  65.09131    0.23858 272.823  < 2e-16 ***
cweight       0.96148    0.01923  50.006  < 2e-16 ***
sex1         -0.85817    0.23858  -3.597 0.000416 ***
cweight:sex1 -0.02834    0.01923  -1.474 0.142233
```

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- "Type-II" tests with centered model:

```
> Anova(mod.davis.2, white=TRUE)

Anova Table (Type II tests)

Response: repwt
             Df         F    Pr(>F)
cweight       1 2165.7754 < 2.2e-16 ***
sex           1   15.1678 0.0001388 ***
cweight:sex   1    1.8684 0.1733720
Residuals   179
```

Example

Dummy Regression with White-Huber Coefficient Covariances: Davis Data

- "Type-III" tests with centered model:

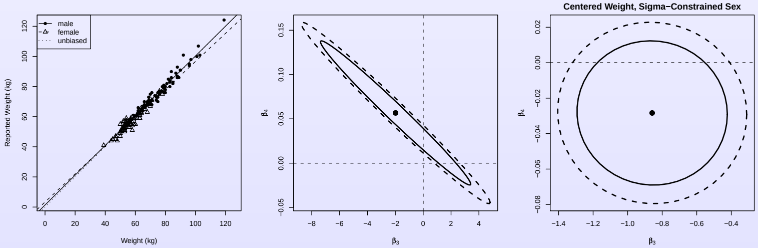
```
> Anova(mod.davis.2, white=TRUE, type=3)

Anova Table (Type III tests)

Response: repwt
             Df          F    Pr(>F)
(Intercept)   1 86075.1218 < 2.2e-16 ***
cweight       1  2150.3272 < 2.2e-16 ***
sex           1    14.9616 0.0001535 ***
cweight:sex   1     1.8684 0.1733720
Residuals   179
```

Examples

Geometry of the Davis Regression Example



- The "Type-III" tests are given by the perpendicular shadows of the solid ellipses on the parameter axes, while the "Type-II" tests are given by the oblique projections producing the narrowest shadows.
- For the centered data, the "Type-II" and "Type-III' tests are very similar.

Examples

Two-Way ANOVA (Briefly!)

- The traditional two-way analysis-of-variance (ANOVA) model:

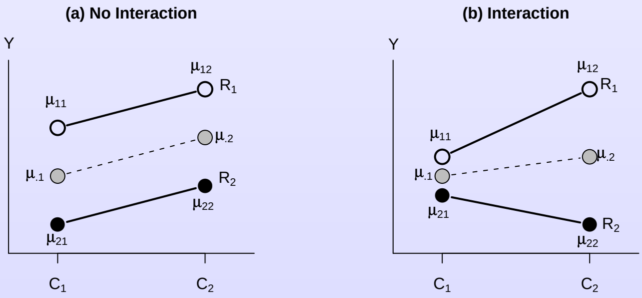
$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

  - Y_{ijk} is the ith of n_{jk} observations in cell $\{R_j, C_k\}$
  - μ is the general mean of Y
  - the α_j and β_k are main-effect parameters
  - the γ_{jk} are interaction parameters
  - the $\varepsilon_{ijk} \sim \text{NID}(0, \sigma^2)$
- Thus $\mu_{jk} = E(Y_{ijk}) = \mu + \alpha_j + \beta_k + \gamma_{jk}$.

Examples

Two-Way ANOVA

- For a 2 × 3 classification:



Examples

Two-Way ANOVA

- **R** uses a full-rank parametrization of the ANOVA model.
- Using sigma contraints to reduce the model to full-rank (i.e., `contr.sum` in **R**), unconditional (i.e., "Type-III") tests of main effects is a test of equality of marginal means, and is interpretable whether or not there is interaction—analogous to centering at $\overline{x}$ in dummy regression.
- The conditional ("Type-II") tests of main effects assumes no interaction and is more powerful under that circumstance.

Implementation

Computation

- Consider the complete QR decomposition of

$$\mathbf{L}_1\mathbf{V}\mathbf{L}_2' = \mathbf{Q}\mathbf{R} = [\mathbf{Q}_1, \mathbf{Q}_2]\begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

  with $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$.
  - Recall that hypothesis matrix L_2 is nested within L_1.
- Let $\mathbf{L}_{1|2} = \mathbf{Q}_2'\mathbf{L}_1$.
  - Then $\mathbf{L}_{1|2}$ has rank r; $\mathbf{L}_{1|2}\mathbf{V}\mathbf{L}_2' = \mathbf{Q}_2'\mathbf{L}_1\mathbf{V}\mathbf{L}_2' = \mathbf{Q}_2'\mathbf{Q}_1 = \mathbf{0}$; and the rows of $\mathbf{L}_{1|2}$ provide a basis for the conjugate complement of the row space of $\mathbf{L}_2$ with respect to the inner product $\mathbf{V}$.
  - Thus, the complete QR decomposition of $\mathbf{L}_1\mathbf{V}\mathbf{L}_2'$ can be used to generate a hypothesis matrix $\mathbf{L}_{1|2}$ from which $Z_{1|2}$ can be obtained.

Implementation

In the **car** Package

- The `Anova` function in the **car** package implements this approach.
  - For `lm` objects, this produces traditional "Type-II" incremental F-tests.
  - For `glm` objects, analogous "Type-II" Wald tests can be computed without refitting the model, as is required for likelihood-ratio tests.
  - A `default` method can be used in other settings, such as linear models with sandwich coefficient covariance matrix estimators, where alternative methods for computing "Type-II" tests are unavailable.
- Additional applications are possible, such as "Type-II" Wald tests of fixed effects in mixed-effect models.