

Simultaneous Inference Procedures for General Linear Hypotheses

Torsten Hothorn

March 5, 2008

1 Introduction

Consider a parametric model $\mathcal{M}(Y, \beta)$ with observations Y and a p -dimensional vector of parameters β . This model could be some kind of regression model where $Y = (y, x)$ can be split up into a dependent variable y and regressors x . An example is a linear regression model $y = x^\top \beta$ or a generalized linear model (GLM) or a survival regression.

Our primary target is simultaneous inference about *general linear hypotheses* on β . More specifically, the global null hypothesis is formulated in terms of linear functions of the parameter vector $\beta \in \mathbb{R}^p$ [Searle, 1971]:

$$H_0 : \mathbf{K}\beta = \mathbf{m}$$

where \mathbf{K} is a $k \times p$ matrix with each row corresponding to one partial hypothesis. However, we are not only interested in the *global* hypothesis H_0 but in all partial hypotheses defined by the rows $K_j, j = 1, \dots, k$, of \mathbf{K} and the elements of $\mathbf{m} = (m_1, \dots, m_k)$:

$$H_0^j : K_j \beta = m_j \text{ with global hypothesis } H_0 = \bigcap_{j=1}^k H_0^j$$

We only consider simultaneous inference procedures, both tests and confidence intervals, which control the *family-wise error rate* (FWE), that is the probability of incorrectly rejecting at least one hypothesis $H_0^j, j = 1, \dots, k$.

1.1 Parameter Estimates

We assume we are provided with an estimate $\hat{\beta}$ of β based on observations Y_1, \dots, Y_n . The estimate $\hat{\beta}$ follows a joint multivariate normal distribution with mean β and covariance matrix Σ , either exactly or asymptotically. Moreover, we assume that an estimate $\mathbb{V}(\hat{\beta})$ of the covariance matrix Σ is available. It then holds that the linear combination $\mathbf{K}\hat{\beta}$ follows a joint normal distribution $\mathcal{N}(\mathbf{K}\beta, \mathbf{K}\Sigma\mathbf{K}^\top)$, either exactly or asymptotically.

1.2 Simultaneous Tests and Confidence Intervals

Under the conditions of the global hypothesis H_0 it holds that

$$\mathbf{K}\hat{\beta} - \mathbf{m} \sim \mathcal{N}(0, \mathbf{K}\Sigma\mathbf{K}^\top),$$

either exactly or asymptotically. Let $\sigma = \text{diag}(\mathbf{K}\mathbb{V}(\hat{\beta})\mathbf{K}^\top)$ denote the estimated standard deviations for all elements of $\mathbf{K}\hat{\beta}$. Then, all inference procedures are based on the vector of all k standardized test statistics

$$\mathbf{z} = (z_1, \dots, z_k) = \sigma^{-\frac{1}{2}}(\mathbf{K}\hat{\beta} - \mathbf{m}).$$

The correlation matrix of the elements of \mathbf{z} is

$$\mathbb{V}(\mathbf{z}) = \sigma^{-\frac{1}{2}}\mathbf{K}\mathbb{V}(\hat{\beta})\mathbf{K}^\top \left(\sigma^{-\frac{1}{2}}\right)^\top.$$

Under H_0 it holds that $\mathbf{z} \rightarrow \mathcal{N}(0, \mathbb{V}(\mathbf{z}))$. When $\hat{\beta}$ follows a normal distribution exactly, the \mathbf{z} statistics follow a multivariate t distribution with $n - \text{Rank}(\mathbf{K})$ degrees of freedom and correlation matrix $\mathbb{V}(\mathbf{z})$.

A simultaneous inference procedure is based on the maximum of the absolute values of the test statistics: $\max |\mathbf{z}|$. Adjusted p values, controlling the family-wise error rate, for each linear hypothesis H_0^j are $p_j = P_{H_0}(\max(|\mathbf{z}|) \geq |z_j|)$. Efficient algorithms for the evaluation of both multivariate distributions are nowadays available [Genz, 1992, Genz and Bretz, 1999, 2002].

Example: Simple Linear Model. Consider a simple univariate linear model regressing the distance to stop on speed for 50 cars:

```
R> lm.cars <- lm(dist ~ speed, data = cars)
R> summary(lm.cars)
```

Call:

```
lm(formula = dist ~ speed, data = cars)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-29.07	-9.53	-2.27	9.21	43.20

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-17.579	6.758	-2.60	0.012 *
speed	3.932	0.416	9.46	1.5e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.4 on 48 degrees of freedom

Multiple R-squared: 0.651, Adjusted R-squared: 0.644

F-statistic: 89.6 on 1 and 48 DF, p-value: 1.49e-12

The estimates of the regression coefficients β and their covariance matrix can be extracted from the fitted model via:

```
R> betahat <- coef(lm.cars)
R> Vbetahat <- vcov(lm.cars)
```

At first, we are interested in the hypothesis $\beta_1 = 0$ and $\beta_2 = 0$. This is equivalent to the linear hypothesis $\mathbf{K}\beta = 0$ where $\mathbf{K} = \text{diag}(2)$, i.e.,

```
R> K <- diag(2)
R> Sigma <- diag(1/sqrt(diag(K %%% Vbetahat %%% t(K))))
R> z <- Sigma %%% K %%% betahat
R> Cor <- Sigma %%% (K %%% Vbetahat %%% t(K)) %%% t(Sigma)
```

Note that $\mathbf{z} = (-2.6011, 9.464)$ is equal to the t statistics. The multiplicity-adjusted p values can now be computed by means of the multivariate t distribution utilizing the `pmvt` function available in package **mvtnorm**:

```
R> library("mvtnorm")
R> df.cars <- nrow(cars) - length(betahat)
R> sapply(abs(z), function(x) 1 - pmvt(-rep(x, 2), rep(x,
+      2), corr = Cor, df = df.cars))
```

```
[1] 1.661e-02 2.458e-12
```

Note that the p value of the global test is the minimum p value of the partial tests.

The computations above can be performed much more conveniently using the functionality implemented in package **multcomp**. The function `glht` just takes a fitted model and a matrix defining the linear functions, and thus hypotheses, to be tested:

```
R> library("multcomp")
R> cars.ht <- glht(lm.cars, linfct = K)
R> summary(cars.ht)
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lm(formula = dist ~ speed, data = cars)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
(Intercept) == 0	-17.579	6.758	-2.60	0.017 *
speed == 0	3.932	0.416	9.46	<1e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported)

Simultaneous confidence intervals corresponding to this multiple testing procedure are available via

```
R> confint(cars.ht)
```

Simultaneous Confidence Intervals for General Linear Hypotheses

```
Fit: lm(formula = dist ~ speed, data = cars)
```

```
Estimated Quantile = 2.13
```

```
Linear Hypotheses:
```

	Estimate	lwr	upr
(Intercept) == 0	-17.579	-31.977	-3.181
speed == 0	3.932	3.047	4.818

```
95% family-wise confidence level
```

The application of the framework isn't limited to linear models, nonlinear least-squares estimates can be tested as well. Consider constructing simultaneous confidence intervals for the model parameters (example from the manual page of nls):

```
R> DNase1 <- subset(DNase, Run == 1)
R> fm1DNase1 <- nls(density ~ SSlogis(log(conc), Asym,
+   xmid, scal), DNase1)
R> K <- diag(3)
R> rownames(K) <- names(coef(fm1DNase1))
R> confint(glht(fm1DNase1, linfct = K))
```

Simultaneous Confidence Intervals for General Linear Hypotheses

```
Fit: nls(formula = density ~ SSlogis(log(conc), Asym, xmid, scal),
  data = DNase1, algorithm = "default", control = list(maxiter = 50,
  tol = 1e-05, minFactor = 0.0009765625, printEval = FALSE,
  warnOnly = FALSE), trace = FALSE)
```

```
Estimated Quantile = 2.136
```

```
Linear Hypotheses:
```

	Estimate	lwr	upr
Asym == 0	2.345	2.178	2.512
xmid == 0	1.483	1.309	1.657
scal == 0	1.041	0.973	1.110

```
95% family-wise confidence level
```

which is not totally different from univariate confidence intervals

```
R> confint(fm1DNase1)
```

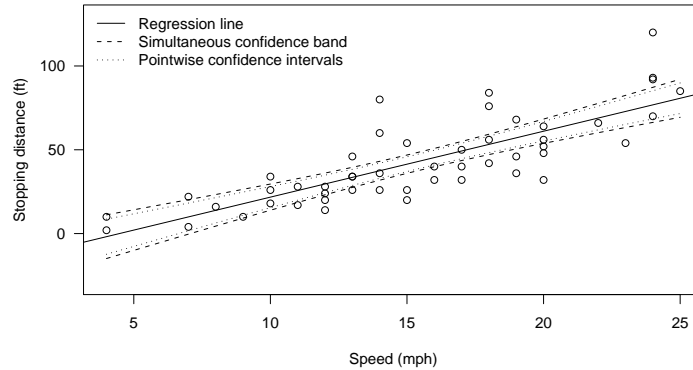


Figure 1: `cars` data: Regression line with confidence bands (dashed) and intervals (dotted).

```

          2.5% 97.5%
Asym 2.1935 2.539
xmid 1.3215 1.679
scal 0.9743 1.115

```

because the parameter estimates are highly correlated

```
R> cov2cor(vcov(fm1DNase1))
```

```

      Asym  xmid  scal
Asym 1.0000 0.9868 0.9008
xmid 0.9868 1.0000 0.9063
scal 0.9008 0.9063 1.0000

```

Example: Confidence Bands for Regression Line. Suppose we want to plot the linear model fit to the `cars` data including an assessment of the variability of the model fit. This can be based on simultaneous confidence intervals for the regression line $x_i^\top \hat{\beta}$:

```

R> K <- model.matrix(lm.cars)[!duplicated(cars$speed),
+   ]
R> ci.cars <- confint(glht(lm.cars, linfct = K), abseps = 0.1)

```

Figure 1 depicts the regression fit together with the confidence band for the regression line and the pointwise confidence intervals as computed by `predict(lm.cars)`.

2 Multiple Comparison Procedures

Multiple comparisons of means, i.e., regression coefficients for groups in AN(C)OVA models, are a special case of the general framework sketched in the previous section. The main difficulty is that the comparisons one is usually interested in, for example all-pairwise differences, can't be directly specified based on model parameters of an AN(C)OVA regression model. We start with a simple one-way ANOVA example and generalize to ANCOVA models in the following.

Consider a one-way ANOVA model, i.e., the only covariate x is a factor at j levels. In the absence of an intercept term only, the elements of the parameter vector $\beta \in \mathbb{R}^j$ correspond to the mean of the response in each of the j groups:

```
R> ex <- data.frame(y = rnorm(12), x = gl(3, 4, labels = LETTERS[1:3]))
R> aov.ex <- aov(y ~ x - 1, data = ex)
R> coef(aov.ex)
```

```
      xA      xB      xC
0.2013 -0.2532 -0.3249
```

Thus, the hypotheses $\beta_2 - \beta_1 = 0$ and $\beta_3 - \beta_1 = 0$ can be written in form of a linear function $\mathbf{K}\beta$ with

```
R> K <- rbind(c(-1, 1, 0), c(-1, 0, 1))
R> rownames(K) <- c("B - A", "C - A")
R> colnames(K) <- names(coef(aov.ex))
R> K
```

```
      xA xB xC
B - A -1  1  0
C - A -1  0  1
```

Using the general linear hypothesis function `glht`, this so-called ‘many-to-one comparison procedure’ [Dunnett, 1955] can be performed via

```
R> summary(glht(aov.ex, linfct = K))
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: aov(formula = y ~ x - 1, data = ex)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
B - A == 0	-0.454	0.559	-0.81	0.64
C - A == 0	-0.526	0.559	-0.94	0.56

(Adjusted p values reported)

Alternatively, a symbolic description of the general linear hypothesis of interest can be supplied to `glht`:

```
R> summary(glht(aov.ex, linfct = c("xB - xA = 0", "xC - xA = 0")))
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: aov(formula = y ~ x - 1, data = ex)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
xB - xA == 0	-0.454	0.559	-0.81	0.64
xC - xA == 0	-0.526	0.559	-0.94	0.56

(Adjusted p values reported)

However, in the presence of an intercept term, the full parameter vector $\beta = c(\mu, \beta_1, \dots, \beta_j)$ can't be estimated due to singularities in the corresponding design matrix. Therefore, a vector of *contrasts* β^* of the original parameter vector β is fitted. More technically, a contrast matrix \mathbf{C} is included into this model such that $\beta = \mathbf{C}\beta^*$ any we only obtain estimates for β^* , but not for β :

```
R> aov.ex2 <- aov(y ~ x, data = ex)
R> coef(aov.ex2)
```

(Intercept)	xB	xC
0.2013	-0.4545	-0.5262

The default contrasts in R are so-called treatment contrasts, nothing but differences in means for one baseline group (compare the Dunnett contrasts and the estimated regression coefficients):

```
R> contr.treatment(table(ex$x))
```

```
  4 4
4 0 0
4 1 0
4 0 1
```

```
R> K %*% contr.treatment(table(ex$x)) %*% coef(aov.ex2)[-1]
```

```
      [,1]
B - A -0.4545
C - A -0.5262
```

so that $\mathbf{KC}\hat{\beta}^* = \mathbf{K}\hat{\beta}$.

When the `mcp` function is used to specify linear hypotheses, the `glht` function takes care of contrasts. Within `mcp`, the matrix of linear hypotheses \mathbf{K} can be written in terms of β , not β^* . Note that the matrix of linear hypotheses only applies to those elements of $\hat{\beta}^*$ attached to factor `x` but not to the intercept term:

```
R> summary(glht(aov.ex2, linfct = mcp(x = K)))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts

```
Fit: aov(formula = y ~ x, data = ex)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
B - A == 0	-0.454	0.559	-0.81	0.64
C - A == 0	-0.526	0.559	-0.94	0.56

(Adjusted p values reported)

or, a little bit more convenient in this simple case:

```
R> summary(glht(aov.ex2, linfct = mcp(x = c("B - A = 0",  
+      "C - A = 0"))))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts

```
Fit: aov(formula = y ~ x, data = ex)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	p value
B - A == 0	-0.454	0.559	-0.81	0.64
C - A == 0	-0.526	0.559	-0.94	0.56

(Adjusted p values reported)

More generally, inference on linear functions of parameters which can be interpreted as ‘means’ are known as *multiple comparison procedures* (MCP). For some of the more prominent special cases, the corresponding linear functions can be computed by convenience functions part of **multcomp**. For example, Tukey all-pair comparisons for the factor *x* can be set up using

```
R> glht(aov.ex2, linfct = mcp(x = "Tukey"))
```

General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:

	Estimate
B - A == 0	-0.4545
C - A == 0	-0.5262
C - B == 0	-0.0717

The initial parameterization of the model is automatically taken into account:

```
R> glht(aov.ex, linfct = mcp(x = "Tukey"))
```

General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:

	Estimate
B - A == 0	-0.4545
C - A == 0	-0.5262
C - B == 0	-0.0717

3 Test Procedures

Several global and multiple test procedures are available from the `summary` method of `glht` objects and can be specified via its `test` argument:

- `test = univariate()` univariate p values based on either the t or normal distribution are reported. Controls the type I error for each partial hypothesis only.
- `test = Ftest()` global F test for H_0 .
- `test = Chisqtest()` global χ^2 test for H_0 .
- `test = adjusted()` multiple test procedures as specified by the `type` argument to `adjusted`: "`free`" denotes adjusted p values as computed from the joint normal or t distribution of the `z` statistics (default), "`Shaffer`" implements Bonferroni-adjustments taking logical constraints into account Shaffer [1986] and "`Westfall`" takes both logical constraints and correlations among the `z` statistics into account Westfall [1997]. In addition, all adjustment methods implemented in `p.adjust` can be specified as well.

4 Quality Assurance

The analyses shown in Westfall et al. [1999] can be reproduced using `multcomp` by running the R transcript file in `inst/MCMT`.

References

Charles W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121, 1955.

- Alan Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–149, 1992.
- Alan Genz and Frank Bretz. Numerical computation of multivariate t -probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63:361–378, 1999.
- Alan Genz and Frank Bretz. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971, 2002.
- Shayle R. Searle. *Linear Models*. John Wiley & Sons, New York, 1971.
- Juliet P. Shaffer. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81:826–831, 1986.
- Peter H. Westfall. Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, 92:299–306, 1997.
- Peter H. Westfall, Randall D. Tobias, Dror Rom, Russell D. Wolfinger, and Yosef Hochberg. *Multiple Comparisons and Multiple Tests Using the SAS System*. SAS Institute, Cary, NC, 1999.