stats545.4 545.4 Multivariate confidence intervals

Gunnar Stefansson

14. október 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 Tests of hypotheses including multiple comparisons in the linear model

1.1 On distributions

If $\underbrace{\mathbf{y}}_{n\times 1} \sim n(\underbrace{\mathbf{X}}_{n\times p}\underbrace{\beta}_{p\times 1}, \sigma^2 \underbrace{\mathbf{I}}_{n\times n})$
and ψ are estimable functions, then
then $\hat{\psi} \sim n\left(\psi, \Sigma_{\hat{\psi}}\right)$, $\frac{||\mathbf{y}-\mathbf{X}\hat{\beta}||^2}{\sigma^2} \sim \chi^2_{n-r}$ and these two quantities are independent.

1.1.1 Details

Let $\underbrace{\mathbf{y}}_{n\times 1} \sim n(\underbrace{\mathbf{X}}_{n\times p}\underbrace{\beta}_{p\times 1}, \sigma^2 \underbrace{\mathbf{I}}_{n\times n})$
and assume $rank(\mathbf{X}) = r \leq p$.
The interest will be in obtaining some joint confidence statement on a vector, $\psi = (\psi_1, \ldots, \psi_q)'$, where each $\psi_i = \mathbf{c}_i'\beta$ is an estimable function. Write $\hat{\psi} = (\hat{\psi}_1, \ldots, \hat{\psi}_q)'$ for the least squares estimates with $\hat{\psi}_i = \mathbf{c}_i'\hat{\beta}$ where $\hat{\beta}$ is any LS estimate and one can therefore also write $\hat{\psi}_i = \mathbf{a}_i'\mathbf{y}$ for unique $a_i \in sp(\mathbf{X})$.
The above can be written more concisely as $\psi = \mathbf{C}\beta$ using obvious definitions. It follows that

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

and the variance-covariance matrix of the estimates can be written in several equivalent ways:

$$V[\hat{\psi}] = \Sigma_{\hat{\psi}} = \sigma^2\mathbf{A}\mathbf{A}' = \sigma^2\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}',$$

where the last equality only holds if $\mathbf{X}$ is of full rank ($r = p$). The formulations are equivalent but vary quite a bit in usefulness on a case-by-case basis.
This leads to the following theorem.

Theorem 1.1. $\hat{\psi} \sim n\left(\psi, \Sigma_{\hat{\psi}}\right)$, $\frac{||\mathbf{y}-\mathbf{X}\hat{\beta}||^2}{\sigma^2} \sim \chi^2_{n-r}$ and these two quantities are independent.

1.1.2 Handout

Proof: See stats545.3

1.2 Confidence ellipsoids

$$P_\beta\left[(\hat{\psi}-\psi)'\mathbf{B}^{-1}(\hat{\psi}-\psi) \leq qs^2F_{q,n-r,1-\alpha}\right] = 1-\alpha$$

This is an example of **simultaneous inference**: a single statement on a multivariate estimable function using a single α-level.

1.2.1 Details

Theorem 1.2. Under the above assumptions and definitions,

$$\frac{(\hat{\psi}-\psi)'\mathbf{B}^{-1}(\hat{\psi}-\psi)/q}{||\mathbf{y}-\mathbf{X}\hat{\beta}||^2/(n-r)} \sim F_{q,n-r}$$

Noting that the denominator is the usual estimator, s^2 of σ^2, it follows that the following probability statement holds and can be used to obtain a confidence ellipsoid for ψ.

$$P_\beta\left[(\hat{\psi}-\psi)'\mathbf{B}^{-1}(\hat{\psi}-\psi) \leq qs^2F_{q,n-r,1-\alpha}\right] = 1-\alpha$$

These intervals are very general and lead to several important special cases.

1.2.2 Handout

It is of interest to derive confidence regions, $R(\mathbf{y}) \subseteq \mathbb{R}^n$ such that

$$P_\beta\left[\psi \in R(\mathbf{y})\right] = 1-\alpha \quad \forall \beta \in \mathbb{R}^p.$$

Assume (without loss of generality) that $rank(C) = q$ and note that $q \leq p$.
Now, $\psi = \mathbf{C}\beta \in \mathbb{R}^q$ and the estimates can be written $\hat{\psi} = \mathbf{A}\mathbf{y}$ for an appropriate choice of $\mathbf{A}$ so $E\hat{\psi} = \psi$ and $V\hat{\psi} = \sigma^2\mathbf{B}$ with $\mathbf{B} = \mathbf{A}\mathbf{A}'$. Next note that

$$\mathbf{C}\beta = \psi = E\hat{\psi} = \mathbf{A}\mathbf{X}\beta \quad \forall \beta$$

so that $\mathbf{C} = \mathbf{A}\mathbf{X}$ and hence $q = rank(\mathbf{C}) = rank(\mathbf{A}\mathbf{X}) \leq rank(\mathbf{A}) \leq q$ where the last inequality follows from $\mathbf{A}$ being a $q \times n$ matrix. But this implies that $rank(\mathbf{A}) = q$ and it is a know result from linear algebra that $rank(\mathbf{B}) = rank(\mathbf{A})$. Since $\mathbf{B}$ is a $q \times q$ matrix, it follows that $\mathbf{B}$ is nonsingular.
Hence

$$\hat{\psi} \sim n\left(\psi, \sigma^2\mathbf{B}\right).$$

Now, for any ν-dimensional multivariate normal random vector $\mathbf{Z}$ with positive definite variance-covariance matrix $\Sigma_\mathbf{Z}$ and mean vector $\mu_\mathbf{Z}$, it will be considered known that

$$(\mathbf{Z}-\mu_\mathbf{Z})'\Sigma_\mathbf{Z}^{-1}(\mathbf{Z}-\mu_\mathbf{Z}) \sim \chi^2_\nu.$$

This result easily follows from decomposing $\Sigma_\mathbf{Z}^{-1}$ into $\mathbf{L}\mathbf{L}'$ where $\mathbf{L}$ is a lower triangular matrix and defining $\mathbf{U} = \mathbf{L}(\mathbf{Z}-\mu_\mathbf{Z})$. Then the components of $\mathbf{U}$ will be i.i.d. $n(0,1)$ and therefore

$$(\mathbf{Z}-\mu_\mathbf{Z})'\Sigma_\mathbf{Z}^{-1}(\mathbf{Z}-\mu_\mathbf{Z}) = ||\mathbf{U}||^2 \sim \chi^2_\nu.$$

It is therefore seen that

$$(\hat{\psi}-\psi)'\left(\sigma^2\mathbf{B}\right)^{-1}(\hat{\psi}-\psi) \sim \chi^2_q. \tag{1}$$

From above we know that this is independent of $\frac{||\mathbf{y}-\mathbf{X}\hat{\beta}||^2}{\sigma^2} \sim \chi^2_{n-r}$
from which we obtain the above theorem.

1.3 Confidence interval for a single estimable function

For a single estimable function with estimator $\hat{\psi} = \mathbf{c}'\hat{\beta} = \mathbf{a}'\mathbf{y}$,

$$\hat{\sigma}^2_{\hat{\psi}} = \mathbf{a}'\mathbf{a}s^2$$

and
A confidence interval for ψ: can be based on

$$(\hat{\psi} - \psi)^2 \leq \mathbf{a}'\mathbf{a}s^2 F_{1,n-r,1-\alpha}$$

or on

$$P\left[\psi \in \left[\hat{\psi} - t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s, \hat{\psi} + t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s\right]\right] = 1 - \alpha$$

1.3.1 Details

Consider a single ($q = 1$) confidence interval for a general estimable function. Write $\psi = \mathbf{c}'\beta$ and note that $rank(\mathbf{c}) = 1$ if $\mathbf{c} \neq \mathbf{0}$. Our estimator for ψ is $\hat{\psi} = \mathbf{c}'\hat{\beta}$ and can be written $\psi = \mathbf{a}'\mathbf{y}$ for an appropriate $\mathbf{a}$.
It follows that the variance of $\hat{\psi}$ is

$$\hat{\sigma}^2_{\hat{\psi}} = \mathbf{a}'\mathbf{a}s^2$$

and a confidence interval for ψ can be based on

$$(\hat{\psi} - \psi)^2 \leq \mathbf{a}'\mathbf{a}s^2 F_{1,n-r,1-\alpha}$$

or on the following corresponding probability statement:

$$P\left[\psi \in \left[\hat{\psi} - t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s, \hat{\psi} + t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s\right]\right] = 1 - \alpha$$

1.3.2 Examples

Example 1.1. When it comes to computing a confidence interval for a single estimable function, we have seen that we can simply compute the values using an interval of the form

$$\left[\hat{\psi} - t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s, \hat{\psi} + t_{n-r,1-\alpha/2}\sqrt{\mathbf{a}'\mathbf{a}}s\right].$$

There are a few tricks to this.
First of all, since $\mathbf{a}'\mathbf{y} = \mathbf{c}'\beta$, the variance can be obtained either from $V[\hat{\psi}] = \mathbf{a}'\mathbf{a}\sigma$ as is done above, or by using the alternative formulation

$$V[\beta] = \Sigma_{\hat{\beta}} = \sigma^2 \left(X'X\right)^{-1}$$

which gives

$$V[\hat{\psi}] = \sigma^2 \mathbf{c}' \left(X'X\right)^{-1} \mathbf{c}$$

and the corresponding confidence interval for ψ is:

$$\left[\mathbf{c}'\hat{\beta} - t_{n-r,1-\alpha/2}\sqrt{\mathbf{c}'\left(X'X\right)^{-1}\mathbf{c}}s, \mathbf{c}'\hat{\beta} + t_{n-r,1-\alpha/2}\sqrt{\mathbf{c}'\left(X'X\right)^{-1}\mathbf{c}}s\right].$$

Example 1.2. In the case of a linear model involving parameters for two groups (factor levels), μ_i and other parameters, possibly a large number of regression parameters, the above result still holds.
The corresponding confidence interval for $\psi = \mu_1 - \mu_2$ is based on

$$\left[\mathbf{c}'\hat{\beta} - t_{n-r,1-\alpha/2}\sqrt{\mathbf{c}'\left(X'X\right)^{-1}\mathbf{c}}s, \mathbf{c}'\hat{\beta} + t_{n-r,1-\alpha/2}\sqrt{\mathbf{c}'\left(X'X\right)^{-1}\mathbf{c}}s\right].$$

and reduces to

$$\hat{\mu}_1 - \hat{\mu}_2 \pm t_{n-r,1-\alpha/2}s_{\hat{\mu}_1-\hat{\mu}_2}.$$

Notice how the degrees of freedom in the t-cutoff are the same as in the regression, i.e. these are the degrees of freedom in $s^2 = MSE = SSE/(n-r)$ where n and r are, as always, the number of rows in the X-matrix and the rank of the X-matrix.

Example 1.3. In many cases it is trivial to compute $V[\hat{\psi}]$ since the estimates are classical and well known. For example there is no need to complicate the issue when looking at a contrast of the form

$$\bar{y}_{1.} - 2\bar{y}_{2.} + \bar{y}_{3.}$$

in the one-way layout with equal sample sizes J for each i. Here we see trivially that the variance of $\hat{\psi}$ is simply $\sigma^2(4/J)$ and the confidence interval becomes correspondingly trivial to compute.

Example 1.4. In the fixed-replicate two-way layout with with interaction, the model is

$$y_{ijk} = \mu_{ij} + \varepsilon_{ij} = \mu + \alpha_i + \beta_i + \gamma_{ij} + \varepsilon_{ijk}$$

and the variance is estimated with $s^2 = SSE/(n-r)$ where

$$SSE = \sum_{i,j,k}\left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

and the degrees of freedom are given in the usual anova table by $n - IJ = IJ(K-1)$. Hence

$$s^2 = \frac{\sum_{i,j,k}\left(y_{ijk} - \bar{y}_{ij.}\right)^2}{IJ(K-1)}$$

To get a confidence interval for a single mean, this is based on $\hat{\mu}_{ij} = \bar{y}_{ij.}$ and we know the variance of this is $\sigma^2_{\hat{\mu}_{ij}} = \sigma^2/K$ and the veriance is therefore estimated using

$$s_{\bar{y}_{ij.}} = s^2/K$$

and the confidence interval becomes

$$\bar{y}_{ij.} \pm t^* s^2/\sqrt{K}$$

where t^* is based on $IJ(K-1)$ df, NOT $K-1$ df!!

1.4 Testing hypotheses for multiple estimable functions

$$H_0 : \psi_1 = \psi_2 = \ldots = \psi_q = 0 \text{ vs } H_a : \text{ not } H_0$$

Reject H_0 **if**

$$\hat{\psi}'\mathbf{B}^{-1}\hat{\psi} > qs^2 F_{q,n-r,1-\alpha}$$

1.4.1 Details

As another example, consider testing the hypothesis that several (linearly independent) estimable functions are zero, i.e. test

$$H_0 : \psi_1 = \psi_2 = \ldots = \psi_q = 0 \text{ vs } H_a : \text{ not } H_0$$

The simplest method to test this hypothesis is to reject H_0 if ψ is not in the confidence set, i.e.: **Reject** H_0 **if**

$$\hat{\psi}'\mathbf{B}^{-1}\hat{\psi} > qs^2 F_{q,n-r,1-\alpha}$$

1.5 Multiple comparisons

$$P\left[\hat{\psi}_i - \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}_i} < \psi_i < \hat{\psi}_i + \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}_i} \quad i = 1, \ldots, q\right] \geq 1 - \alpha$$

1.5.1 Details

The confidence ellipsoids are of course multiple comparisons in the sense that they provide information about the entire vector of estimable functions under consideration. However it is usually of greater interest to draw conclusions on the individual estimable functions, but the inference should be simultaneous. To this end, the confidence ellipsoids are used as a basis and the intervals are simply deduced from the ellipsoids as follows.
Theorem:

$$P\left[\hat{\psi}_i - \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}_i} < \psi_i < \hat{\psi}_i + \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}_i} \quad i = 1, \ldots, q\right] \geq 1 - \alpha$$

Corollary: Let $L := \{\psi = \sum_1^q h_i \psi_i : h_1, \ldots, h_q \in \mathbb{R}\}$. Then

$$P\left[\hat{\psi} - \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}} < \psi < \hat{\psi} + \sqrt{qF_{q,n-r,1-\alpha}}\hat{\sigma}_{\hat{\psi}} \quad \forall \psi \in L\right] = 1 - \alpha$$

1.5.2 Handout

Several interesting, useful and important methods can be derived from these confidence sets. These sets are attributed to Scheffe and are called the S-sets or S-methods of obtaining simultaneous confidence statements.
The proof of the theorem is not trivial and the reader is referred to Scheffe's book.

1.6 Data-snooping

When $q = 1$ the S-method is the same as a t-test. When $q > 1$, conducting multiple t-tests will ruin the error rate. The S-method permit multiple test:
Can use the S-method for data-snooping
May want to use a large α
Better than LSD: Know explicitly the error rate

1.6.1 Details

Suppose we are interested in **searching for significance** or **data-snooping**. Normally this is not permitted since usually the hypotheses to be tested need to be specified in advance. However, the confidence sets discussed in this tutorial are all simultaneous and can therefore be searched in arbitrary detail.

Suppose Ψ is a set of estimable functions, e.g. a set spanned by q estimable functions: $\Psi = \left\{\psi = k_1\psi_1 + \ldots + k_q\psi_q\right\}$ where $\psi_i = \mathbf{c}_i'\beta$ and $\mathbf{c}_1, \ldots, \mathbf{c}_n$ are linearly independent. Then from the earlier results we can assert

$$P\left[\hat{\psi} - \sqrt{qF^*}\hat{\sigma}_{\hat{\psi}} \le \psi \le \hat{\psi} + \sqrt{qF^*}\hat{\sigma}_{\hat{\psi}} \quad \forall \psi \in \Psi\right] = 1 - \alpha$$

and we are therefore allowed to **search** among all estimable functions within the set to find significant effects.

The "trick" here lies in the cutoff-point, $qF^* = qF_{q,n-r,1-\alpha}$, which takes into account the dimension of the space.

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 Special cases of Scheffes confidence sets: Applications to simple linear regression

2.1 The setup

$$\boxed{y_i \sim n(\alpha + \beta x_i, \sigma^2),\ i = 1, \ldots, n}$$

2.1.1 Handout

We will assume the model to be $y_i \sim n(\alpha + \beta x_i, \sigma^2)$, independent and the x_i are not all the same, so the $\mathbf{X}$-matrix is of full rank.
In this case the OLS estimators, $\hat{\alpha}$, $\hat{\beta}$ are well known linear combinations of the y-values. They can be written as

$$\hat{\beta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1}'\mathbf{y} \\ \mathbf{a_2}'\mathbf{y} \end{pmatrix} = \mathbf{Ay}$$

for an appropriate choice of $\mathbf{a_1}$,$\mathbf{a_2}$ and $\mathbf{A}$.
The variance-covariance matrix has been derived elsewhere as $\sigma^2\mathbf{AA}' = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$.
Since σ^2 can be estimated with $s^2 = MSE$, the variances and covariances of $\hat{\alpha}$ and $\hat{\beta}$ can easily be estimated.

2.2 The intercept

$$\boxed{\text{C.I. for } \alpha \text{ alone is the same as before.}}$$

2.2.1 Handout

The intercept alone is a simple linear function of the full parameter vector, i.e.

$$\psi = \alpha = (1,0)\begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

the corresponding estimate is

$$\hat{\psi} = \hat{\alpha} = (1,0)\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \mathbf{a_1}'\mathbf{y}.$$

The variance, $\sigma^2_{\hat{\psi}}$, corresponding estimator, $\hat{\sigma}^2_{\hat{\psi}}$ and estimate $s^2_{\hat{\psi}}$ of this particular estimable function is well known.
Since this is a single estimable function we have $q = 1$. If the x-values are not all the same then $\mathbf{X}$ has full rank so $r = p = 2$ and we obtain the same CI as before.
To actually derive the quantities we can either use the matrix version, define $\mathbf{X}$ as usual and see that the estimate becomes

$$s^2_{\hat{\psi}} = s^2\mathbf{a_1}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a_1}$$

where $s^2 = MSE$.
Alternatively we can go from the usual elementary formulae

$$\hat{\beta} = \frac{(x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

and

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

and rewrite first the top and then the bottom one as linear combinations of y_i to find the actual a_{1i}-components of $\mathbf{a_1}$. From that we obtain $\sigma^2_{\hat{\psi}} = \sigma^2 \sum a^2_{1i}$ as the true variance, the estimator follows by substituting $\hat{\sigma}^2$ for σ^2 and the estimate is obtained using

$$s^2 = MSE = ||\mathbf{y} - \mathbf{X}\beta||^2 = \sum \left(y_i - \hat{\alpha} - \hat{\beta}x_i\right).$$

Details are left to the reader.

2.3 The slope

C.I. for slope is the same as before

2.3.1 Handout

The slope alone is a simple linear function of the full parameter vector, i.e.

$$\psi = \beta = (0,1)\begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

the corresponding estimate is

$$\hat{\psi} = \hat{\beta} = (0,1)\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \mathbf{a_2}'\mathbf{y}.$$

The variance, $\hat{\sigma}^2_{\hat{\psi}}$, of this particular estimable function is well known.
Since this is a single estimable function we have $q = 1$. If the x-values are not all the same then $\mathbf{X}$ has full rank so $r = p = 2$ and we obtain the same confidence intervals as before.
Details of the derivations are left to the reader.
the same CI as before.

2.4 A simultaneous confidence set for the slope and intercept

Confidence ellipse in the α-β plane

$$\left\{\psi : \frac{(\hat{\psi} - \psi)'\mathbf{B}^{-1}(\hat{\psi} - \psi)/q}{||\mathbf{y} - \mathbf{X}\hat{\beta}||^2/(n-r)} \le F_{q,n-r,1-\alpha}\right\}$$

where

$$\psi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

2.4.1 Handout

Recall that the vector $\psi = (\hat{\alpha}, \hat{\beta})'$ is estimable in simple linear regression if the x-values are not all the same. A simultaneous confidence set for ψ is based on the point estimate $\psi = (\hat{\alpha}, \hat{\beta})'$ and the corresponding covariance matrix and the earlier result

$$\frac{(\hat{\psi} - \psi)'\mathbf{B}^{-1}(\hat{\psi} - \psi)/q}{||\mathbf{y} - \mathbf{X}\hat{\beta}||^2/(n-r)} \sim F_{q,n-r} \tag{2}$$

where B is defined by

$$V[\hat{\psi}] = \sigma^2 B,$$

and we also have

$$V[\hat{\psi}] = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

It follows that in this case $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}$ so in particular $\mathbf{B}^{-1} = \mathbf{X}'\mathbf{X}$
Equation 2 provides a confidence set,

$$\left\{\psi : \frac{(\hat{\psi}-\psi)'\mathbf{B}^{-1}(\hat{\psi}-\psi)/q}{||\mathbf{y}-\mathbf{X}\hat{\beta}||^2/(n-r)} \le F_{q,n-r,1-\alpha}\right\}, \tag{3}$$

which describes an ellipse in the (α,β)-plane.
In terms of the original SLR parameters, the $100(1-\alpha')\%$ confidence set becomes

$$\left\{\begin{pmatrix}\alpha\\ \beta\end{pmatrix} : \left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)' X'X \left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right) \le qs^2 F_{q,n-r,1-\alpha'}\right\}.$$

This confidence set can be used to obtain simultaneous bound on the two parameters:
To find simultaneous confidence intervals for α and β, consider the following optimisation problem:

$$\min_{\alpha,\beta} \beta$$

w.r.t.

$$\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)' X'X \left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right) \le qs^2 F_{q,n-r,1-\alpha'}$$

This will find the smallest possible β within the simultaneous confidence set for α and β.
Repeating this for the four cases of minimising and maximising the values of α and β gives confidence intervals for each of α and β, which hold **simultaneously**.
These simultaneous intervals are of course wider than the usual t-based intervals.
Note that here, $qs^2F_{q,n-r,1-\alpha'}$ is $2s^2F_{2,n-2,1-\alpha'}$ and this does **not** correspond to a t-interval.
There is more than one way to solve this. One is to use the Lagrange function

$$\beta + \lambda\left\{\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)' X'X \left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right) - qs^2 F_{q,n-r,1-\alpha'}\right\},$$

differentiate w.r.t. α, β, λ and set all derivatives to zero.
Another method is to look at the ellipse itself, write $\hat{\alpha}-\alpha$ as a function of $\hat{\beta}-\beta$ and see for which values of β there is a solution for α. The above ellipse needs to first be written out, aka

$$n(\hat{\alpha}-\alpha)^2 + 2\sum_i x_i(\hat{\alpha}-\alpha)\left(\hat{\beta}-\beta\right) + \sum_i x_i^2\left(\hat{\beta}-\beta\right)^2 = qs^2F_{q,n-r,1-\alpha'}.$$

Whichever method is chosen, the deduced confidence intervals should be of the form

$$\hat{\alpha} \pm \hat{\sigma}_{\hat{\alpha}}\sqrt{qF_{q,n-r,1-\alpha'}}$$

and

$$\hat{\beta} \pm \hat{\sigma}_{\hat{\beta}}\sqrt{qF_{q,n-r,1-\alpha'}}$$

and it is worth repeating that these are **simultaneous** confidence intervals which will have a joint confidence greater that 100(1-α')%. It is also worth nothing, that although these intervals are written here in terms of estimators, a strictly correct notation would be to use the estimates instead, as in

$$a \pm s_a \sqrt{qF_{q,n-r,1-\alpha'}}$$

and

$$b \pm s_b \sqrt{qF_{q,n-r,1-\alpha'}}.$$

Writing out this ellipse is also a useful start to compare the univariate (t-based) intervals, the joint ellipse and the simultaneous (F-based) intervals. Writing an R script to draw this is a good exercise for the reader, as is a simulation exercise to compare the actual coverage probability of these various confidence sets.

2.5 Confidence band for the regression line

Simultaneous band for the entire regression line:

$$\left\{ a + bx \pm s\sqrt{2F_{2,n-2,1-\alpha}\left\{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}\right\}} : x \in \mathbb{R} \right\}$$

2.5.1 Handout

The simultaneous confidence set for the two parameters in SLR can be used to obtain a confidence band for the regression line.
The confidence band for the regression line is a simultaneous statement on all points in the set

$$\mathcal{C} = \{\alpha + \beta x : x \in \mathbb{R}\}$$

Now, the variance of the estimates $\hat{\alpha} + \hat{\beta}x$ is well known and it is also clear that the above confidence set is a subset of

$$\mathcal{L} = \{\psi = c_1\alpha + c_2\beta : c_1, c_2 \in \mathbb{R}\}.$$

This is the set of all linear combinations of the two-dimensional parameter vector $(\alpha,\beta)'$, which is an estimable function,

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

The above set, $\mathcal{L}$, consists of all linear combinations of ψ_1 and ψ_2 and can be written as

$$\mathcal{L} = \left\{ \psi = \sum_{h=1}^{2} h_i \psi_i : h_1, h_2 \in \mathbb{R} \right\}.$$

This demonstrates that $\mathcal{L}$ is spanned by two estimable functions, $\psi_1 = \alpha$ and $\psi_2 = \beta$ and $\mathcal{L}$, as in the corollary earlier. It therefore has dimension $q = 2$ and one can use a corresponding F-cutoff to obtain simultaneous confidence bounds for the entire regression line.
To derive the actual formulae, note that a generic point on the regression line, $\psi = \psi_x = \alpha + \beta x$ (an element of $\mathcal{L}$) is predicted with $\hat{\psi} = \hat{\alpha} + \hat{\beta}x$, which has variance

$$\sigma^2_{\hat{\psi}} = V\left[(1,x)\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}\right] = \sigma^2(1,x)\left(X'X\right)^{-1}\begin{pmatrix} 1 \\ x \end{pmatrix}$$

and as usual, this variance is estimated using

$$\hat{\sigma}^2_{\hat{\psi}} = s^2(1,x)\left(X'X\right)^{-1}\begin{pmatrix} 1 \\ x \end{pmatrix}.$$

The confidence band for the entire regression line thus becomes

$$a+bx \pm s\sqrt{2F_{2,n-2,1-\alpha}(1,x)\,(X'X)^{-1}\begin{pmatrix}1\\x\end{pmatrix}},$$

where a, b and s^2 are the usual numerical estimates of the intercept, slope and residual variance (MSE), respectively.
Note the several "tricks"here, where we know the appropriate variances and can use them directly.
Note also that as in earlier examples, we do not need to use the matrix inverse since it is not at all difficult to derive the variance of $\hat{\psi}$ by rewriting it as a linear combination of the y_i:

$$\hat{\alpha}+\hat{\beta}x = \left(\bar{y}-\hat{\beta}\bar{x}\right)+\hat{\beta}x = \bar{y}+(x-\bar{x})\,\hat{\beta}$$

and now insert the two equations $\bar{y}=\frac{1}{n}\sum_i y_i$,

$$\hat{\beta}=\sum_i \frac{x_i-\bar{x}}{\sum_j \left(x_j-\bar{x}\right)^2} y_i$$

to obtain

$$\hat{\alpha}+\hat{\beta}x = \frac{1}{n}\sum_i y_i + (x-\bar{x})\sum_i \frac{x_i-\bar{x}}{\sum_j \left(x_j-\bar{x}\right)^2} y_i = \sum_i \left\{\frac{1}{n}+(x-\bar{x})\frac{x_i-\bar{x}}{\sum_j \left(x_j-\bar{x}\right)^2}\right\} y_i.$$

The right hand side is a linear combination of independent y_i, i.e. $\mathbf{a}'\mathbf{y}$ and the variance becomes $\sigma^2 \sum_i a_i^2$.
The next trick is to note that when squaring the term in the curly brackets, the cross-product involves $\sum_i (x_i-\bar{x}) = 0$ so it goes away and the result becomes:

$$V\left[\hat{\alpha}+\hat{\beta}x\right] = \sigma^2 \sum_i \left\{\left(\frac{1}{n}\right)^2 + (x-\bar{x})^2\left(\frac{x_i-\bar{x}}{\sum_j \left(x_j-\bar{x}\right)^2}\right)^2\right\}$$

which simplifies easily to

$$V\left[\hat{\alpha}+\hat{\beta}x\right] = \left\{\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}\right\}\sigma^2$$

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 The Bonferroni approach to multipe comparisons

3.1 The multiplicity issue

Consider testing k independent hypotheses, each at level α.
Then, since

$$P[\text{conclusion i is incorrect}] = \alpha$$

we obtain

$$P[\text{conclusion i is correct}] = 1 - \alpha,$$

and therefore

$$P[\text{an error occurs}] = 1 - P[\text{all correct}] = 1 - \Pi_{i=1}^{k} P[\text{conclusion i is correct}] = 1 - (1-\alpha)^k$$

3.2 LSD

Consider just doing a whole bunch of t-tests
This amounts to saying "There is something significant going on if there is anything significant seen"
This is the method of **Least Significant Difference** and has a very large potential error rate

3.3 Bonferroni confidence intervals

Bonferroni intervals:
Simple
Always work
Conservative

3.3.1 Details

In general, consider two events, A and B having the same probability, $P[A] = P[B] = \alpha'$. In the current situation, A is the event "confidence interval 1 is wrong" and B is the event "confidence interval 2 is wrong".
The probability of both confidence intervals being correct is

$$\begin{aligned} P[A^c \cap B^c] &= P[(A \cup B)^c] \\ &= 1 - P[A \cup B] \\ &= 1 - (P[A] + P[B] - P[A \cap B]) \\ &\geq 1 - P[A] - P[B] \\ &= 1 - 2\alpha' \end{aligned}$$

It follows that if two confidence statements are made, each with error rate $\alpha' = \alpha/2$, or confidence $100(1-\alpha/2)\%$, then the overall confidence is at least $100(1-\alpha)\%$, i.e. the probability of any error is reduced to α.

3.3.2 Example

In the one-way layout one can use the Bonferroni method to compare all the means in a pairwise manner. Since there are $c = I(I-1)/2$ comparisons, the corresponding confidence intervals become:

$$\bar{y}_{i.} - \bar{y}_{k.} \pm t_{1-\alpha/(2c),ab(n-1)} s\sqrt{\frac{2}{n}}$$

where s^2 is the usual estimate of variance, $s^2 = MSE$.
In some cases, for example the two-way layout, there may be a very large number of potential pairwise comparisons and not all may be of interest.
So suppose only a specific collection of c pairwise differences are of interest in the two-way layout. In this case Bonferroni confidence intervals may be preferable to other methods:

$$\bar{y}_{ij.} - \bar{y}_{kl.} \pm t_{1-\alpha/(2c),ab(n-1)} \sqrt{\frac{2MSE}{n}}$$

References ISBN: 0412982811 **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 Tukeys confidence intervals

4.1 Pairwise multiple comparisons

Tukey's method for pairwise comparisons works!

4.1.1 Details

When all pairwise comparisons are of equal importance, the interest is in being able to make statements of the form

$$P[|\bar{X}_i - \bar{X}_j| \leq d_{ij} \text{ for all } i, j] \geq 1 - \alpha$$

Usually, d_{ij} is taken proportional to the common standard deviation, s and written either as $qs/\sqrt{n}$ or $ws/\sqrt{1/n_i + 1/m_j}$ in the case of unequal sample sizes.
The function TukeyHSD in R and the procedure "proc glm" in SAS (with the Tukey option) can be used for general, and valid, pairwise multiple comparisons.

4.2 Tukeys confidence intervals

The Tukey test is used in the one-way layout, when there is an interest in all pairwise comparisons.
For equal sample sizes J in each cell, the simultaneous confidence intervals using Tukey's method in the one-way layout are as follows:

$$\bar{y}_{i.} - \bar{y}_{j.} \pm q_{1-\alpha,g,N-g} \times \sqrt{MSE/J}$$

where $q^* = q_{1-\alpha,g,N-g}$ is the appropriate quantile from Tukey's studentized range distribution.

4.2.1 Details

If all pairwise comparisons are of interest in the two-way layout the following procedure can be followed to calculate simultaneous $1-\alpha$ confidence limit of the difference.

$$\bar{y}_{ij.} - \bar{y}_{kl.} \pm \frac{1}{\sqrt(2)} q_{1-\alpha,ab,ab(n-1)} \sqrt{MSE\frac{2}{n}}$$

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 Simultaneous confidence intervals for all contrasts

5.1 Scheffes method

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.

6 Comparing confidence sets

6.1 Scheffe, Tukey and Bonferroni

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.

7 Applications

7.1 Background

This lecture is a placeholder for a collection of examples and applications of the theory

7.2 One regression line or two?

are two regression lines really the same?
two level factor...
the question generates 4 models
what comparisons can be made?

7.3 the lack of fit test

taken from Neter et al
the approach reverses the usual logic (do I need a line) to "is a line enough"

7.4 Smoothers

consider the cubic spline...
can be a linear model
can be used to test whether a line is appropriate
if the knots are chosen based on data then we have a GAM, not the usual linear model
other splines are normally used, but this is a simple introduction

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.