Topic 3: Multivariate ANOVA: testing canonical correlations and multivariate linear models

Computations used in the MANOVA tests

In standard (univariate) Analysis of Variance, with usual normality assumptions on the errors, testing about effects of the factors involved in the model description is based on the F test. The F tests are derived from the ANOVA decomposition $\mathrm{SST} = \mathrm{SSA} + \mathrm{SSE}$. The argument goes as follows:

1. SSE and SSA are independent, (up to constant factors involving the variance σ^2 of the errors) χ^2 distributed;
2. By proper norming to account for degrees of freedom, from SSE and SSA one gets statistics that have the following behaviour: the normed SSE always delivers an unbiased estimator of σ^2 no matter if the null hypothesis or alternative is true; the normed SSA delivers an unbiased estimator of σ^2 under the null hypothesis but delivers an unbiased estimator of a "larger" quantity under the alternative.

The above observation is crucial and motivates the F-testing: F statistics are (**suitably normed to account for degrees of freedom**) ratios of SSA/SSE. When taking the ratio, the factors involving σ^2 **cancel out** and σ^2 does not play any role in the distribution of the ratio. Under H_0 their distribution is F. When the null hypothesis is violated, then the same statistics will tend to have "larger" values as compared to the case when H_0 is true. Hence significant (w.r.t. the corresponding F-distribution) values of the statistic lead to rejection of H_0.

Aiming at generalising these ideas to the Multivariate ANOVA (MANOVA) case, we should note that instead of χ^2 distributions we now have to deal with **Wishart** distributions and we need to properly define (a proper functional of) the SSA/SSE ratio which would be a "ratio" of matrices now. Obviously, there are more ways to define suitable statistics in this context! It turns out that such functionals are related to the eigenvalues of the (properly normed) Wishart-distributed matrices that enter the decomposition SST = SSA + SSE in the multivariate case.

Optional viewing: MANOVA

Matthew E. Clapham. (2016). MANOVA. Retrieved from: https://youtu.be/Oec5hwTx-SE

Roots distributions

Let $\boldsymbol{Y}_i, i=1,2,\ldots,n \overset{\text{ind.}}{\sim} N_p(\boldsymbol{\mu}_i, \Sigma)$. Then the following data matrix:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1^\top \\ \boldsymbol{Y}_2^\top \\ \vdots \\ \boldsymbol{Y}_n^\top \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{pmatrix} \in \mathcal{M}_{n,p}$$

is a $n \times p$ matrix containing n p-dimensional (transposed) vectors. Denote: $\mathrm{E}(\boldsymbol{Y}) = M$, $\mathrm{Var}(\overrightarrow{\boldsymbol{Y}}) = \Sigma \otimes I_n$. Let A and B be projectors such that $\boldsymbol{Q}_1 = \boldsymbol{Y}^\top A \boldsymbol{Y}$ and $\boldsymbol{Q}_2 = \boldsymbol{Y}^\top B \boldsymbol{Y}$ are two **independent** $W_p(\Sigma, v)$ and $W_p(\Sigma, q)$ matrices, respectively. Although the theory is general, to keep you on track, you could always think about a multivariate linear model example:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{E}, \hat{\boldsymbol{Y}} = X\hat{\boldsymbol{\beta}}$$
$$A = I_n - X\left(X^\top X\right)^- X^\top, B = X\left(X^\top X\right)^- X^\top - \mathbf{1}_n\left(\mathbf{1}_n^\top \mathbf{1}_n\right)^{-1}\mathbf{1}_n^\top$$

and the corresponding decomposition

$$\boldsymbol{Y}\left[I_n - \mathbf{1}_n\left(\mathbf{1}_n^\top \mathbf{1}_n\right)^{-1}\mathbf{1}_n^\top\right]\boldsymbol{Y} = \boldsymbol{Y}^\top B\boldsymbol{Y} + \boldsymbol{Y}^\top A\boldsymbol{Y} = \boldsymbol{Q}_2 + \boldsymbol{Q}_1$$

of $\mathrm{SST} = \mathrm{SSA} + \mathrm{SSE} = \boldsymbol{Q}_2 + \boldsymbol{Q}_1$ where $\boldsymbol{Q}_2$ is the "hypothesis matrix" and $\boldsymbol{Q}_1$ is the "error matrix".

Now, for hypothesis testing, we need to somehow summarise these two matrices as a single number that we could compare to a critical value or use to compute a p-value. There are various ways to choose such functional transformations (statistics). Some commonly used ones depend on the eigenvalues of $\boldsymbol{Q}_1^{-1}\boldsymbol{Q}_2$, which we will denote as $\lambda_1, \lambda_2, \ldots$, in decreasing order.

- $\Lambda = \left|\boldsymbol{Q}_1\left(\boldsymbol{Q}_1 + \boldsymbol{Q}_2\right)^{-1}\right| = \prod_{i=1}^p \left(1+\lambda_i\right)^{-1}$ (Wilks's Lambda)
- $\operatorname{tr}\left(\boldsymbol{Q}_2\boldsymbol{Q}_1^{-1}\right) = \operatorname{tr}\left(\boldsymbol{Q}_1^{-1}\boldsymbol{Q}_2\right) = \sum_{i=1}^p \lambda_i$ (Lawley-Hotelling trace)
- $\max_i \lambda_i$ (Roy's criterion)
- $V = \operatorname{tr}\left[\boldsymbol{Q}_2\left(\boldsymbol{Q}_1 + \boldsymbol{Q}_2\right)^{-1}\right] = \sum_{i=1}^p \frac{\lambda_i}{1+\lambda_i}$ (Pillai statistic / Pillai's trace)

P-values for these statistics are readily calculated in statistical packages. In these applications, the meaning of $\boldsymbol{Q}_1$ is of the "error matrix" (also denoted by $\boldsymbol{E}$ sometimes) and the meaning of $\boldsymbol{Q}_2$ is that of a "hypothesis matrix" (also denoted by $\boldsymbol{H}$ sometimes).

The distribution of the statistics defined above depends on the following three parameters:

- $p =$ the number of responses
- $q = \nu_h =$ degrees of freedom for the hypothesis
- $v = \nu_e =$ degrees of freedom for the error

Based on these, the following quantities are calculated: $s = \min(p, q)$, $m = 0.5(|p - q| - 1)$, $n = 0.5(v - p - 1)$, $r = v - 0.5(p - q + 1)$, $u = 0.25(pq - 2)$. Moreover, we define: $t = \sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}$ if $p^2 + q^2 - 5 > 0$ and $t = 1$ otherwise. Let us order the eigenvalues of $\boldsymbol{E}^{-1}\boldsymbol{H} = \boldsymbol{Q}_1^{-1}\boldsymbol{Q}_2$ according to: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

Then the following distribution results are true: (Note: the F -statistic quoted below is **exact** if $s = 1$ or 2, otherwise the F-distribution is an **approximation**):

- Wilks's test. The test statistics, Wilks's lambda, is $\Lambda = \frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|} = \prod_{i=1}^{p} \frac{1}{1+\lambda_i}$. Then it holds: $F = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} . \frac{rt-2u}{pq} \sim F_{pq, rt-2u}$ df (Rao's F).
- Lawley–Hotelling trace Test. The Lawley–Hotelling statistic is $U = \operatorname{tr}(\boldsymbol{E}^{-1}\boldsymbol{H}) = \lambda_1 + \cdots + \lambda_p$, and $F = 2(sn+1)\frac{U}{s^2(2m+s+1)} \sim F_{s(2m+s+1), 2(sn+1)}$ df.
- Pillai's test. The test-statistics, Pillai trace, is $V = \operatorname{tr}(\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{E})^{-1}) = \frac{\lambda_1}{1+\lambda_1} + \cdots + \frac{\lambda_p}{1+\lambda_p}$ and $F = \frac{2n+s+1}{2m+s+1} \times \frac{V}{s-V} \sim F_{s(2m+s+1), s(2n+s+1)}$ df.
- Roy's maximum root criterion. The test-statistic is just the largest eigenvalue λ_1.

Finally, we shall mention one historically older and very universal approximation to the distribution of the Λ statistic due to Bartlett (1927):

It holds: level of $-\left[\nu_e - \frac{p-\nu_h+1}{2}\right] \log \Lambda = c\left(p, \nu_h, M\right) \times$ level of $\chi^2_{p\nu_h}$, where the constant $c\left(p, \nu_h, M = \nu_e - p + 1\right)$ is given in accompanying tables. Such tables are prepared for levels $\alpha = 0.10, 0.05, 0.025$ etc..

In the context of testing the hypothesis about the significance of the first canonical correlation, we have:

$$\boldsymbol{E} = \boldsymbol{S}_{22} - \boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}, \boldsymbol{H} = \boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}$$

The Wilks's statistic becomes $\frac{|\boldsymbol{S}|}{|\boldsymbol{S}_{11}||\boldsymbol{S}_{22}|}$.

We also see that in this case, if μ_i^2 were the squared canonical correlations then μ_1^2 was defined as the maximal eigenvalue to $\boldsymbol{S}_{22}^{-1}\boldsymbol{H}$, that is, it is a solution to $|(\boldsymbol{E}+\boldsymbol{H})^{-1}\boldsymbol{H} - \mu_1^2 I| = 0$ However, setting $\lambda_1 = \frac{\mu_1^2}{1-\mu_1^2}$ we see that:

$$
\begin{aligned}
&\left|(\boldsymbol{E}+\boldsymbol{H})^{-1}\boldsymbol{H}-\mu_1^2 I\right|=0\\
\Longrightarrow &\left|\boldsymbol{H}-\mu_1^2(\boldsymbol{E}+\boldsymbol{H})\right|=0\\
\Longrightarrow &\left|\boldsymbol{H}-\frac{\mu_1^2}{1-\mu_1^2}\boldsymbol{E}\right|=0\\
\Longrightarrow &\left|\boldsymbol{E}^{-1}\boldsymbol{H}-\lambda_1 I\right|=0
\end{aligned}
$$

holds and λ_1 is an eigenvalue of $\boldsymbol{E}^{-1}\boldsymbol{H}$. Similarly you can argue for the remaining $\lambda_i=\frac{\mu_i^2}{1-\mu_i^2}$ values.

Comparisons

From all statistics discussed, Wilks's lambda has been most widely applied. One important reason for this is that this statistic has the virtue of being convenient to use and, more importantly, being related to the Likelihood Ratio Test! Despite the above, the fact that so many different statistics exist for the same hypothesis testing problem, indicates that there is no universally best test. Power comparisons of the above tests are almost lacking since the distribution of the statistic under alternatives is hardly known.

Demonstration: Multivariate LM and MANOVA

Start by watching the following demonstration by Dr Krivitsky.

Transcript

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete this task select the 'MLM_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

MLM_Examples.demo.Rmd

The output of the RMD file is also displayed below:

Challenge: Multivariate LM and MANOVA

If you choose to complete this task in your own RStudio, upload the following file:

MLM_Examples.challenge.Rmd

Select the 'MLM_Examples.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The solution will also be available here on Friday of Week 3 by clicking on the 'Solution' tab.

The output of the RMD file is also displayed below: