Testing commutative quadratic structure of covariance matrix under multivariate t distribution

Katarzyna Filipiak[1], Daniel Klein[2], Stepan Mazur[3], Monika Mokrzycka[4]

[1]Institute of Mathematics, Poznań University of Technology, Poland
[2]Institute of Mathematics, P. J. Šafarik University in Košice, Slovakia
[3]School of Business, Örebro University, Sweden
[4]Institute of Plant Genetics, PAS, Poznań, Poland

5[th] Congress of Polish Statistics, Warsaw, 2025

Model

n – number of independent objects

p – number of features

Model

n – number of independent objects

p – number of features

Multivariate t model:

$$\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n), \qquad \mathbf{t}_i \sim t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Model

n – number of independent objects
p – number of features

Multivariate t model:

$$\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n), \qquad \mathbf{t}_i \sim t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

ν – degrees of freedom of multivariate t distribution

$$\mathbb{E}(\mathbf{t}_i) = \boldsymbol{\mu}, \qquad \mathbb{D}(\mathbf{t}_i) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}$$

Model

n – number of independent objects
p – number of features

Multivariate t model:

$$\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n), \qquad \mathbf{t}_i \sim t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

ν – degrees of freedom of multivariate t distribution

$$\mathbb{E}(\mathbf{t}_i) = \boldsymbol{\mu}, \qquad \mathbb{D}(\mathbf{t}_i) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}$$

Kotz, Nadarajah (2004):

$$f(\mathbf{T}) = \left(\frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{p/2}}\right)^n |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \left(\prod_{i=1}^{n}\left[1 + \frac{1}{\nu}(\mathbf{t}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{t}_i - \boldsymbol{\mu})\right]\right)^{-\frac{\nu+p}{2}}$$

Maximum likelihood estimators (MLEs)

EM algorithm (Liu and Rubin, 1995)

$$\delta_i^{(k)} = \left(\mathbf{t}_i - \boldsymbol{\mu}^{(k)}\right)' \boldsymbol{\Sigma}^{(k)^{-1}} \left(\mathbf{t}_i - \boldsymbol{\mu}^{(k)}\right)$$

$$\boldsymbol{\mu}^{(k+1)} = \sum_{i=1}^{n} \frac{\mathbf{t}_i}{\nu+\delta_i^{(k)}} \bigg/ \sum_{i=1}^{n} \frac{1}{\nu+\delta_i^{(k)}}$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{\nu+p}{n} \sum_{i=1}^{n} \frac{\left(\mathbf{t}_i-\boldsymbol{\mu}^{(k+1)}\right)\left(\mathbf{t}_i-\boldsymbol{\mu}^{(k+1)}\right)'}{\nu+\delta_i^{(k)}}$$

Maximum likelihood estimators (MLEs)

EM algorithm (Liu and Rubin, 1995)

$$\delta_i^{(k)} = \left(\mathbf{t}_i - \boldsymbol{\mu}^{(k)}\right)' \boldsymbol{\Sigma}^{(k)^{-1}} \left(\mathbf{t}_i - \boldsymbol{\mu}^{(k)}\right)$$

$$\boldsymbol{\mu}^{(k+1)} = \sum_{i=1}^{n} \frac{\mathbf{t}_i}{\nu+\delta_i^{(k)}} \Bigg/ \sum_{i=1}^{n} \frac{1}{\nu+\delta_i^{(k)}}$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{\nu+p}{n} \sum_{i=1}^{n} \frac{\left(\mathbf{t}_i-\boldsymbol{\mu}^{(k+1)}\right)\left(\mathbf{t}_i-\boldsymbol{\mu}^{(k+1)}\right)'}{\nu+\delta_i^{(k)}}$$

MLE of ν (McLachlan and Krishnan, 1997)

$$\Psi\left(\frac{\nu^{(k)}}{2}\right) - \ln\frac{\nu^{(k)}}{2} = 1 + \frac{1}{n}\sum_{i=1}^{n}\left[\ln\frac{\nu^{(k-1)}+p}{\nu^{(k-1)}+\delta_i^{(k)}} - \frac{\nu^{(k-1)}+p}{\nu^{(k-1)}+\delta_i^{(k)}}\right] + \Psi\left(\frac{\nu^{(k-1)}+p}{2}\right) - \ln\frac{\nu^{(k-1)}+p}{2}$$

$\Psi(\cdot)$ – digamma function

Commutative quadratic subspace (Seely, 1971)

The linear space $\mathcal{V}$ is a quadratic subspace if $\mathbf{V} \in \mathcal{V}$ implies that $\mathbf{V}^2 \in \mathcal{V}$. Moreover, $\mathcal{V}$ is a commutative quadratic subspace if the product of two matrices belonging to $\mathcal{V}$ is commutative.

Commutative quadratic subspace (Seely, 1971)

The linear space $\mathcal{V}$ is a quadratic subspace if $\mathbf{V} \in \mathcal{V}$ implies that $\mathbf{V}^2 \in \mathcal{V}$. Moreover, $\mathcal{V}$ is a commutative quadratic subspace if the product of two matrices belonging to $\mathcal{V}$ is commutative.

If $\boldsymbol{\Sigma}$ belongs to quadratic subspace, then

$$\boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j$$

$\lambda_j > 0$

$\mathbf{V}_j$, $j = 1, \ldots, \ell$ – known, idempotent and strongly orthogonal matrices ($\mathbf{V}_j^2 = \mathbf{V}_j$ and $\mathbf{V}_j \mathbf{V}_{j'} = \mathbf{0}$ for $j \neq j'$).

Commutative quadratic subspace - examples

- sphericity: $\lambda_1 \mathbf{I}_p$
- diagonality: $\mathrm{Diag}(\lambda_1, \ldots, \lambda_p)$
- compound symmetry (CS): $\lambda_1 \mathbf{P}_p + \lambda_2(\mathbf{I}_p - \mathbf{P}_p)$, $\mathbf{P}_p = \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'$
- circular Toeplitz (CT): $\sum_{j=1}^{k} \lambda_j \mathbf{V}_j$, $k = \lfloor \frac{p}{2} \rfloor + 1$
- block-diagonal with CS blocks
- block-CS with CT blocks

Commutative quadratic subspace - examples

- sphericity: $\lambda_1 \mathbf{I}_p$
- diagonality: $\mathrm{Diag}(\lambda_1, \ldots, \lambda_p)$
- compound symmetry (CS): $\lambda_1 \mathbf{P}_p + \lambda_2(\mathbf{I}_p - \mathbf{P}_p)$, $\mathbf{P}_p = \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'$
- circular Toeplitz (CT): $\sum_{j=1}^{k} \lambda_j \mathbf{V}_j$, $k = \lfloor \frac{p}{2} \rfloor + 1$
- block-diagonal with CS blocks
- block-CS with CT blocks

Matrix from commutative quadratic subspace – matrix isomorphic with

$$\mathrm{Bdiag}(\lambda_1 \mathbf{I}_{p_1}, \ldots, \lambda_\ell \mathbf{I}_{p_\ell}), \qquad \sum_{j=1}^{\ell} p_j = p$$

(Filipiak et al., 2025)

Hypothesis and tests

$$H_0: \ \boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j \qquad H_1: \ \boldsymbol{\Sigma} \ \text{ is unstructured}$$

Hypothesis and tests

$$H_0 : \ \boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j \qquad H_1 : \ \boldsymbol{\Sigma} \ \text{ is unstructured}$$

Tests:

- likelihood ratio test (LRT)
- Rao score test (RST)
- Wald test (WT)

Maximum likelihood estimators (MLEs)

ν_0, $\boldsymbol{\mu}_0$ – degrees of freedom and mean under $H_0 : \boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j$

$$
\begin{aligned}
\delta_{i,0}^{(k)} &= \left(\mathbf{t}_i - \boldsymbol{\mu}_0^{(k)}\right)' \left(\sum_{j=1}^{\ell} \frac{1}{\lambda_j^{(k)}} \mathbf{V}_j\right) \left(\mathbf{t}_i - \boldsymbol{\mu}_0^{(k)}\right) \\
\boldsymbol{\mu}_0^{(k+1)} &= \sum_{i=1}^{n} \frac{\mathbf{t}_i}{\nu_0 + \delta_{i,0}^{(k)}} \Big/ \sum_{i=1}^{n} \frac{1}{\nu_0 + \delta_{i,0}^{(k)}} \\
\lambda_j^{(k+1)} &= \frac{\nu_0 + p}{n \operatorname{Tr} \mathbf{V}_j} \sum_{i=1}^{n} \frac{\left(\mathbf{t}_i - \boldsymbol{\mu}_0^{(k+1)}\right)' \mathbf{V}_j \left(\mathbf{t}_i - \boldsymbol{\mu}_0^{(k+1)}\right)}{\nu_0 + \delta_{i,0}^{(k)}}
\end{aligned}
$$

Likelihood ratio test

Theorem

The LRT statistic for testing H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, when no constraints are imposed on $\boldsymbol{\mu}$ and ν, is given by

$$\begin{aligned}
\mathrm{LRT}(\mathbf{T}) \quad = \quad & n\left[\widehat{\nu}\ln\widehat{\nu} - \widehat{\nu}_0\ln\widehat{\nu}_0 + 2\ln\frac{\Gamma(\frac{\widehat{\nu}+p}{2})\Gamma(\frac{\widehat{\nu}_0}{2})}{\Gamma(\frac{\widehat{\nu}_0+p}{2})\Gamma(\frac{\widehat{\nu}}{2})} + \ln\frac{|\widehat{\boldsymbol{\Sigma}}_0|}{|\widehat{\boldsymbol{\Sigma}}|}\right] \\
& +(\widehat{\nu}_0 + p)\sum_{i=1}^{n}\ln\left[\widehat{\nu}_0 + (\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)'\widehat{\boldsymbol{\Sigma}}_0^{-1}(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)\right] \\
& -(\widehat{\nu} + p)\sum_{i=1}^{n}\ln\left[\widehat{\nu} + (\mathbf{t}_i - \widehat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{t}_i - \widehat{\boldsymbol{\mu}})\right],
\end{aligned}$$

where $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n)$ is a random sample of size n from $t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\nu > 2$ is unknown, $\widehat{\nu}$, $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$ are the MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under H_1, and $\widehat{\nu}_0$, $\widehat{\boldsymbol{\mu}}_0$, $\widehat{\lambda}_j$, $j = 1, \ldots, \ell$, (coefficients in $\widehat{\boldsymbol{\Sigma}}_0$) are the MLEs of ν_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$ under H_0.
When $n \to \infty$ and H_0 holds, the distribution of LRT tends to the chi-square distribution with $p(p+1)/2 - \ell$ degrees of freedom.

Rao score test

Theorem

The RST statistics for testing H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, when no constraints are imposed on $\boldsymbol{\mu}$ and ν, is given by

$$\mathrm{RST}(\mathbf{T}) \quad = \quad \tfrac{n(\widehat{\nu}_0+p+2)}{2\widehat{\nu}_0(\widehat{\nu}_0+p)} \left\{ \widehat{\nu}_0 \cdot \mathrm{Tr}[(\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}_0^{-1})^2] + \mathrm{Tr}^2(\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}_0^{-1}) \right\},$$

with

$$\widehat{\mathbf{U}} = \widehat{\mathbf{U}}(\mathbf{T}) = \widehat{\boldsymbol{\Sigma}}_0 - \tfrac{\widehat{\nu}_0+p}{n} \sum_{i=1}^{n} \frac{(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)'}{\widehat{\nu}_0 + (\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)'\widehat{\boldsymbol{\Sigma}}_0^{-1}(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)},$$

where $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n)$ is a random sample of size n from $t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\nu > 2$ is unknown, and $\widehat{\nu}_0$, $\widehat{\boldsymbol{\mu}}_0$, $\widehat{\lambda}_j$, $j = 1, \ldots, \ell$, (coefficients in $\widehat{\boldsymbol{\Sigma}}_0$) are the MLEs of ν_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$ under H_0.
When $n \to \infty$ and H_0 holds, the distribution of RST tends to the chi-square distribution with $p(p+1)/2 - \ell$ degrees of freedom.

Rao score test

Theorem

The RST statistics for testing H_0: $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, *when no constraints are imposed on* $\boldsymbol{\mu}$ *and* ν, *is given by*

$$\mathrm{RST}(\mathbf{T}) \;=\; \tfrac{n(\widehat{\nu}_0+p+2)}{2\widehat{\nu}_0(\widehat{\nu}_0+p)} \left\{ \widehat{\nu}_0 \cdot \mathrm{Tr}[(\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}_0^{-1})^2] + \mathrm{Tr}^2(\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}_0^{-1}) \right\},$$

with

$$\widehat{\mathbf{U}} = \widehat{\mathbf{U}}(\mathbf{T}) = \widehat{\mathbf{\Sigma}}_0 - \tfrac{\widehat{\nu}_0+p}{n} \sum_{i=1}^{n} \frac{(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)'}{\widehat{\nu}_0 + (\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)'\widehat{\mathbf{\Sigma}}_0^{-1}(\mathbf{t}_i - \widehat{\boldsymbol{\mu}}_0)},$$

where $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n)$ *is a random sample of size* n *from* $t_p(\nu, \boldsymbol{\mu}, \mathbf{\Sigma})$, $\nu > 2$ *is unknown, and* $\widehat{\nu}_0$, $\widehat{\boldsymbol{\mu}}_0$, $\widehat{\lambda}_j$, $j = 1, \ldots, \ell$, *(coefficients in* $\widehat{\mathbf{\Sigma}}_0$*) are the MLEs of* ν_0, $\boldsymbol{\mu}_0$, $\mathbf{\Sigma}_0$ *under* H_0.
When $n \to \infty$ *and* H_0 *holds, the distribution of RST tends to the chi-square distribution with* $p(p+1)/2 - \ell$ *degrees of freedom.*

!!! RST can be used also for $n < p$, which is not the case for LRT !!!

LRT and RST: known ν

Remark 1

When the degrees of freedom ν of multivariate t distribution is the same under the null and alternative hypothesis (which happens for example when ν is known) the LRT and RST become, respectively,

$$\mathrm{LRT}\,(\mathbf{T}) \;=\; n(\ln|\widehat{\boldsymbol{\Sigma}}_0| - \ln|\widehat{\boldsymbol{\Sigma}}|) + (\nu+p)\sum_{i=1}^{n} \ln \frac{\nu+(\boldsymbol{t}_i-\widehat{\boldsymbol{\mu}}_0)'\widehat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{t}_i-\widehat{\boldsymbol{\mu}}_0)}{\nu+(\boldsymbol{t}_i-\widehat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{t}_i-\widehat{\boldsymbol{\mu}})}$$

and

$$\mathrm{RST}(\mathbf{T}) \;=\; \tfrac{n(\nu+p+2)}{2\nu(\nu+p)}\left\{\nu\cdot\mathrm{Tr}[(\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}_0^{-1})^2] + \mathrm{Tr}^2(\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}_0^{-1})\right\}$$

with

$$\widehat{\mathbf{U}} = \widehat{\mathbf{U}}(\mathbf{T}) = \widehat{\boldsymbol{\Sigma}}_0 - \tfrac{\nu+p}{n}\sum_{i=1}^{n}\frac{(\mathbf{t}_i-\widehat{\boldsymbol{\mu}}_0)(\mathbf{t}_i-\widehat{\boldsymbol{\mu}}_0)'}{\nu+(\mathbf{t}_i-\widehat{\boldsymbol{\mu}}_0)'\widehat{\boldsymbol{\Sigma}}_0^{-1}(\mathbf{t}_i-\widehat{\boldsymbol{\mu}}_0)}.$$

LRT and RST: $\nu \to \infty$

Remark 2

When the degrees of freedom ν of multivariate t distribution tend to infinity (i.e., multivariate t distribution tends to the normal distribution), respective MLEs change into $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}_0 = \frac{1}{n}\mathbf{T}\mathbf{1}_n = \overline{\mathbf{T}}$, $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{T}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{T}' = \mathbf{S}$, $\widehat{\lambda}_j = \frac{1}{\operatorname{Tr}\boldsymbol{V}_j}\operatorname{Tr}(\mathbf{V}_j\mathbf{S})$, and the test statistics become, respectively,

$$\mathrm{LRT}\,(\mathbf{T}) = n\ln\frac{|\widehat{\boldsymbol{\Sigma}}_0|}{|\mathbf{S}|} \qquad \text{and} \qquad \mathrm{RST}(\mathbf{T}) = \frac{n}{2}\operatorname{Tr}[(\mathbf{I}_p - \mathbf{S}\widehat{\boldsymbol{\Sigma}}_0^{-1})^2].$$

LRT and RST: $\nu \to \infty$

Remark 2

When the degrees of freedom ν of multivariate t distribution tend to infinity (i.e., multivariate t distribution tends to the normal distribution), respective MLEs change into $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}_0 = \frac{1}{n}\mathbf{T}\mathbf{1}_n = \overline{\mathbf{T}}$, $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{T}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{T}' = \mathbf{S}$, $\widehat{\lambda}_j = \frac{1}{\operatorname{Tr}\boldsymbol{V}_j}\operatorname{Tr}(\mathbf{V}_j\mathbf{S})$, and the test statistics become, respectively,

$$\mathrm{LRT}\,(\mathbf{T}) = n\ln\frac{|\widehat{\boldsymbol{\Sigma}}_0|}{|\mathbf{S}|} \quad \text{and} \quad \mathrm{RST}(\mathbf{T}) = \frac{n}{2}\operatorname{Tr}[(\mathbf{I}_p - \mathbf{S}\widehat{\boldsymbol{\Sigma}}_0^{-1})^2].$$

Above results are in line with the forms of test statistics determined under the multivariate normal distribution for testing compound symmetry structure (Filipiak et al., 2017, Roy et al. 2018)

Wald test for simple and composite hypothesis

Simple hypothesis: (Rao, 2005)

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \mathrm{WT} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{F}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Wald test for simple and composite hypothesis

Simple hypothesis: (Rao, 2005)

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \mathrm{WT} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{F}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Composite hypothesis:

$$H_0 : \ \mathbf{h}(\boldsymbol{\theta}) = \mathbf{c} \qquad \mathrm{WT} = (\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c})'\mathbf{A}^{-1}(\widehat{\boldsymbol{\theta}})(\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c})$$

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{H}'(\boldsymbol{\theta}), \quad \mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Wald test for simple and composite hypothesis

Simple hypothesis: (Rao, 2005)

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \mathrm{WT} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{F}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Composite hypothesis:

$$H_0 : \ \mathbf{h}(\boldsymbol{\theta}) = \mathbf{c} \qquad \mathrm{WT} = (\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c})'\mathbf{A}^{-1}(\widehat{\boldsymbol{\theta}})(\mathbf{h}(\widehat{\boldsymbol{\theta}}) - \mathbf{c})$$

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{H}'(\boldsymbol{\theta}), \quad \mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$$H_0 : \boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j - \text{ composite hypothesis}$$

Alternative representation of H_0

$$\boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j \quad \text{isomorphic with} \quad \boldsymbol{\Sigma}_* = \mathrm{Diag}(\lambda_1 \mathbf{1}'_{v_1}, \ldots, \lambda_\ell \mathbf{1}'_{v_\ell})$$

Alternative representation of H_0

$$\boldsymbol{\Sigma} = \sum_{j=1}^{\ell} \lambda_j \mathbf{V}_j \quad \text{isomorphic with} \quad \boldsymbol{\Sigma}_* = \mathrm{Diag}(\lambda_1 \mathbf{1}'_{v_1}, \ldots, \lambda_\ell \mathbf{1}'_{v_\ell})$$

$$\mathbf{CPD}_p^+(\mathbf{V}' \otimes \mathbf{V}')\mathrm{vec}\boldsymbol{\Sigma} = \mathbf{0}$$

$$\mathbf{C} = \left(\mathbf{0}_{q^* \times q} : \mathrm{Diag}(\mathbf{C}_{q+1}, \ldots, \mathbf{C}_\ell, \mathbf{I}_{p(p-1)/2})\right), \qquad q^* = p(p-1)/2 + \sum_{j=q+1}^{\ell} (\mathrm{Tr}\,\mathbf{V}_j - 1)$$

$$\mathbf{C}_j = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix} : \qquad (\mathrm{Tr}\,\mathbf{V}_j - 1) \times \mathrm{Tr}\,\mathbf{V}_j$$

$\mathbf{P}$ – particular permutation matrix of order $\frac{p(p+1)}{2}$, $\quad \mathbf{D}_p$ – $p \times \frac{p(p+1)}{2}$ duplication matrix

$$\mathbf{V} = (\mathbf{V}_1^*, \ldots, \mathbf{V}_\ell^*), \qquad \mathbf{V}_j^* \mathbf{V}_j^{*'} = \mathbf{V}_j$$

Wald test (WT)

$$H_0: \ \mathbf{CPD}_p^+(\mathbf{V}' \otimes \mathbf{V}')\text{vec}\boldsymbol{\Sigma} = \mathbf{0}$$

Wald test (WT)

$$H_0 : \ \mathbf{CPD}_p^+(\mathbf{V}' \otimes \mathbf{V}')\mathrm{vec}\boldsymbol{\Sigma} = \mathbf{0}$$

Theorem

The WT statistic for testing H_0, when no constraints are imposed on $\boldsymbol{\mu}$ and ν, is given by

$$\mathrm{WT}(\mathbf{T}) = \mathrm{vech}'\widehat{\boldsymbol{\Sigma}}_*\mathbf{P}'\mathbf{C}'\left(\mathbf{CP}\widehat{\mathbf{F}}_*^{-1}\mathbf{P}'\mathbf{C}'\right)^{-1}\mathbf{CP}\mathrm{vech}\widehat{\boldsymbol{\Sigma}}_*$$

with

$$\widehat{\mathbf{F}}_* = \tfrac{n}{2(\widehat{\nu}+p+2)}\mathbf{D}_p'\left((\widehat{\nu}+p)(\widehat{\boldsymbol{\Sigma}}_*^{-1} \otimes \widehat{\boldsymbol{\Sigma}}_*^{-1}) - \mathrm{vec}\widehat{\boldsymbol{\Sigma}}_*^{-1}\mathrm{vec}'\widehat{\boldsymbol{\Sigma}}_*^{-1}\right)\mathbf{D}_p,$$

where $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_n)$ is a random sample of size n from $t_p(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\nu > 2$ is unknown, $\widehat{\nu}$ is the MLE of ν, $\widehat{\boldsymbol{\Sigma}}_* = \mathbf{V}'\widehat{\boldsymbol{\Sigma}}\mathbf{V}$, and $\widehat{\boldsymbol{\Sigma}}$ is the MLE of $\boldsymbol{\Sigma}$ under H_1. When $n \to \infty$ and H_0 holds, the distribution of WT statistic tend to the chi-square distribution with $p(p+1)/2 - \ell$ degrees of freedom.

WT: known ν, $\nu \to \infty$

Remark 3

When the degrees of freedom ν of multivariate t distribution is known, the WT statistic remains the same, however, in the formula for $\widehat{\mathbf{F}}_*$ the estimator $\widehat{\nu}$ is replaced by ν.

WT: known ν, $\nu \to \infty$

Remark 3

When the degrees of freedom ν of multivariate t distribution is known, the WT statistic remains the same, however, in the formula for $\widehat{\mathbf{F}}_*$ the estimator $\widehat{\nu}$ is replaced by ν.

Remark 4

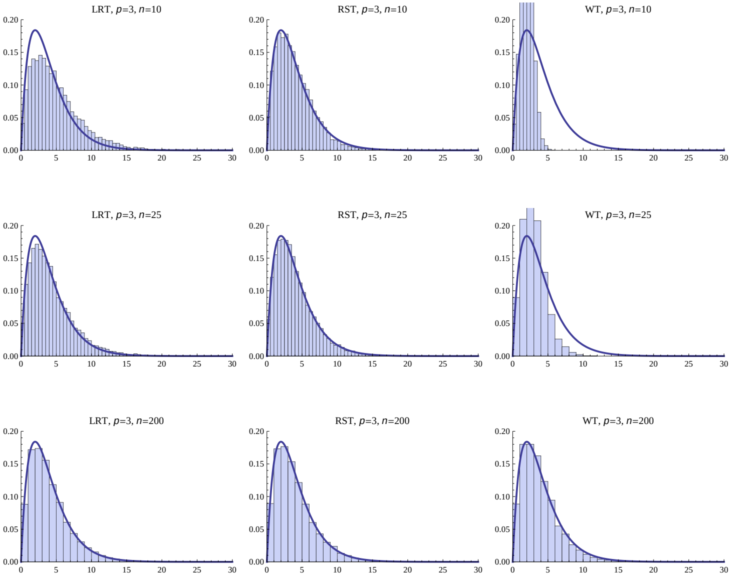
When the degrees of freedom ν of multivariate t distribution tend to infinity in the above remark (i.e., multivariate t distribution tends to the normal distribution), the MLE of $\boldsymbol{\Sigma}$ becomes $\mathbf{S}$ and hence $\widehat{\boldsymbol{\Sigma}}_* = \mathbf{V}'\mathbf{S}\mathbf{V}$. Then, the WT statistic has the form

$$\mathrm{WT}(\mathbf{T}) = \mathrm{vech}'(\mathbf{V}'\mathbf{S}\mathbf{V})\mathbf{P}'\mathbf{C}'\left(\mathbf{C}\mathbf{P}\widehat{\mathbf{F}}_{\#}^{-1}\mathbf{P}'\mathbf{C}'\right)^{-1}\mathbf{C}\mathbf{P}\mathrm{vech}(\mathbf{V}'\mathbf{S}\mathbf{V})$$

with

$$\widehat{\mathbf{F}}_{\#} = \tfrac{n}{2}\mathbf{D}_p'(\mathbf{V}' \otimes \mathbf{V}')(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})(\mathbf{V} \otimes \mathbf{V})\mathbf{D}_p.$$

Empirical null distributions (unknown ν)



Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:

Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:
    - the speed of convergence to the limiting chi-square distribution is the highest for the distribution of RST, and the lowest for WT

Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:
    - the speed of convergence to the limiting chi-square distribution is the highest for the distribution of RST, and the lowest for WT
    - the empirical type I error tends to the nominal significance level ($\alpha = 0.05$) quicker than its counterparts for LRT and WT

Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:
    - the speed of convergence to the limiting chi-square distribution is the highest for the distribution of RST, and the lowest for WT
    - the empirical type I error tends to the nominal significance level ($\alpha = 0.05$) quicker than its counterparts for LRT and WT
    - the LRT is the most liberal test, while the WT - the most conservative

Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:
   - the speed of convergence to the limiting chi-square distribution is the highest for the distribution of RST, and the lowest for WT
   - the empirical type I error tends to the nominal significance level ($\alpha = 0.05$) quicker than its counterparts for LRT and WT
   - the LRT is the most liberal test, while the WT - the most conservative
2. The estimation of ν is challenging; we used the EM algorithm presented by McLachlan and Krishnan (1997, Sect. 5.8). Nevertheless, the EM algorithm does not always converge, especially when the sample size is small.

Simulation studies - main conclusions

1. For each test statistic, the empirical null distributions do not differ significantly when ν is known or unknown:
    - the speed of convergence to the limiting chi-square distribution is the highest for the distribution of RST, and the lowest for WT
    - the empirical type I error tends to the nominal significance level ($\alpha = 0.05$) quicker than its counterparts for LRT and WT
    - the LRT is the most liberal test, while the WT - the most conservative
2. The estimation of ν is challenging; we used the EM algorithm presented by McLachlan and Krishnan (1997, Sect. 5.8). Nevertheless, the EM algorithm does not always converge, especially when the sample size is small. The number of non-converging (nc) cases to obtain 10000 results:

p	3				7			
n	10	25	50	200	10	25	50	200
nc	19524	1895	191	0	379411 (651)	666	2	0

Data analysis (Liang et al., 2015)

11 Kalanchoe plants with 3 flowers per plant, 4 petals per flower

$\mathbf{t}_i$ – vector of the petals length, $i = 1, \ldots, n$

Data analysis (Liang et al., 2015)

11 Kalanchoe plants with 3 flowers per plant, 4 petals per flower

$\mathbf{t}_i$ – vector of the petals length, $i = 1, \ldots, n$

(C1): There is no hierarchical structure in measurements, all 33 flowers are treated equally, despite the fact that they are allocated on the same plant; in this setup the 4×4 compound symmetry (CS) structure is tested, which means that the lengths of all four petals are equally correlated with each other

Data analysis (Liang et al., 2015)

11 Kalanchoe plants with 3 flowers per plant, 4 petals per flower

$\mathbf{t}_i$ – vector of the petals length, $i = 1, \ldots, n$

(C1): There is no hierarchical structure in measurements, all 33 flowers are treated equally, despite the fact that they are allocated on the same plant; in this setup the 4×4 compound symmetry (CS) structure is tested, which means that the lengths of all four petals are equally correlated with each other

(C2): There is no hierarchical structure in measurements, but since 3 flowers are allocated on the same plant, corresponding measurements are not independent; thus, the subset of 11 flowers is randomly chosen (1 flower from each plant) and the 4×4 CS structure is tested

Data analysis (Liang et al., 2015)

11 Kalanchoe plants with 3 flowers per plant, 4 petals per flower

$\mathbf{t}_i$ – vector of the petals length, $i = 1, \ldots, n$

(C1): There is no hierarchical structure in measurements, all 33 flowers are treated equally, despite the fact that they are allocated on the same plant; in this setup the 4×4 compound symmetry (CS) structure is tested, which means that the lengths of all four petals are equally correlated with each other

(C2): There is no hierarchical structure in measurements, but since 3 flowers are allocated on the same plant, corresponding measurements are not independent; thus, the subset of 11 flowers is randomly chosen (1 flower from each plant) and the 4×4 CS structure is tested

(C1) and (C2) structure - results

Table: The values of LRT, RST and WT statistics under multivariate t_3 distribution and under multivariate normality, with respective 95% empirical and limiting quantiles for testing CS covariance structure for Kalanchoe plants data.

	(C1): $n = 33$			(C2): $n = 11$		
	LRT	RST	WT	LRT	RST	WT
under multivariate t_3	4.500	4.739	3.569	13.291	9.103	5.358
empirical quantile	17.078	15.692	9.963	20.560	15.828	5.574
under normality	6.572	6.784	5.406	13.922	7.759	11.321
empirical quantile	17.056	15.709	14.313	21.086	15.519	11.764
limiting quantile	15.507	15.507	15.507	15.507	15.507	15.507

(C1) and (C2) structure - results

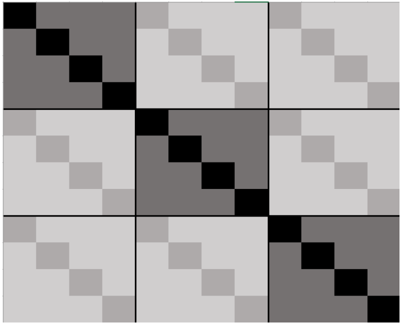
Table: The values of LRT, RST and WT statistics under multivariate t_3 distribution and under multivariate normality, with respective 95% empirical and limiting quantiles for testing CS covariance structure for Kalanchoe plants data.

	(C1): $n = 33$			(C2): $n = 11$		
	LRT	RST	WT	LRT	RST	WT
under multivariate t_3	4.500	4.739	3.569	13.291	9.103	5.358
empirical quantile	17.078	15.692	9.963	20.560	15.828	5.574
under normality	6.572	6.784	5.406	13.922	7.759	11.321
empirical quantile	17.056	15.709	14.313	21.086	15.519	11.764
limiting quantile	15.507	15.507	15.507	15.507	15.507	15.507

CS structure is not rejected

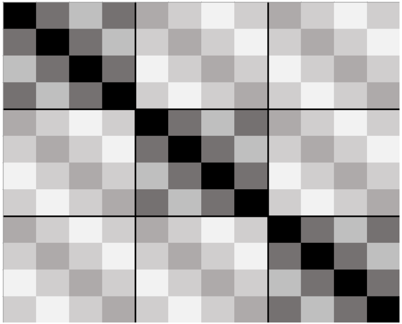
Data analysis

(C3): Hierarchical structure of measurements has an impact on covariance structure; in this setup the sample of size 11 (number of plants) is taken to test the hypothesis about 12×12 block-compound symmetry covariance structure with compound symmetry structure of each 4×4 block (BCS_CS). The block-compound symmetry is motivated by the circular arrangement of three flowers on the plant, while compound symmetry of each block follows from the assumption that the lengths of all four petals are equally correlated with each other



Data analysis

(C4): The same hierarchical structure of measurements as in Case (C3), however, the structure of 4×4 blocks of covariance matrix corresponds to circular Toeplitz matrix structure (BCS_CT). It means, that the neighboring petals are equally correlated, while the petals arranged opposite each other are equally correlated, but not in the same way as neighboring petals



(C3) and (C4) structures - results

Table: The values of RST statistics under multivariate t_3 distribution, under multivariate t_ν distribution with unknown ν and under multivariate normality, with respective 95% empirical and limiting quantiles for testing BCS_CS and BCS_CT covariance structure for Kalanchoe plants data.

	BCS_CS	BCS_CT
under multivariate t_3	99.405	91.654
empirical quantile	100.419	98.045
under multivariate t_ν	108.078	101.671
empirical quantile of t_8 / t_9	101.422	99.595
under normality	131.428	80.602
empirical quantile	103.025	100.399
limiting quantile	95.082	92.808

Summary

1. Multivariate t distribution:

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined
   - it is shown that the RST outperforms remaining tests

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined
   - it is shown that the RST outperforms remaining tests
2. Multivariate normal distribution:

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined
   - it is shown that the RST outperforms remaining tests
2. Multivariate normal distribution:
   - LRT, RST statistics for testing covariance structure from commutative quadratic subspace are determined (so far only the results for testing CS have been presented)

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined
   - it is shown that the RST outperforms remaining tests
2. Multivariate normal distribution:
   - LRT, RST statistics for testing covariance structure from commutative quadratic subspace are determined (so far only the results for testing CS have been presented)
   - WT statistic for testing covariance structure from commutative quadratic subspace is determined

Summary

1. Multivariate t distribution:
   - the algorithm for determination of MLEs of ν, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is presented
   - LRT, RST and WT statistics for testing covariance structure from commutative quadratic subspace are determined
   - it is shown that the RST outperforms remaining tests
2. Multivariate normal distribution:
   - LRT, RST statistics for testing covariance structure from commutative quadratic subspace are determined (so far only the results for testing CS have been presented)
   - WT statistic for testing covariance structure from commutative quadratic subspace is determined
   - it is shown that the RST outperforms remaining tests

Main references

- Filipiak, K., Klein, D., Roy, A. (2017). A comparison of likelihood ratio tests and Rao's score test for three separable covariance matrix structures. *Biom. J.* 59(1), 192–215.
- Filipiak, K., Markiewicz, A., Mrowińska, M. (2025). Characterization of the quadratic space of partitioned matrices. Submitted.
- Kotz, S., Nadarajah, S. (2004). *Multivariate t distributions and their applications.* Cambridge University Press, Cambridge.
- Liang, Y., von Rosen, D., von Rosen, T. (2015). On estimation in hierarchical models with block circular covariance structures. *Ann. Inst. Stat. Math.* 67(4), 773–791.
- Liu, C., Rubin, D. (1995). Ml estimation of the t distribution using EM and its extensions. *Stat. Sin.* 5, 19–39.
- McLachlan, G., Krishnan, T. (1997). *The EM Algorithm and Extensions.* Wiley.
- Rao, C. R. (2005). Score test: historical review and recent developments. In: Balakrishnan, N., Kannan, N., Nagaijuna, H.N. (Eds.), *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, pp. 3–20. Birkhäuser, Boston, MA.
- Roy, A., Filipiak, K., Klein, D. (2018). Testing a block exchangeable covariance matrix. *Statistics* 52(2), 393–408.
- Seely, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Stat.* 42(2), 710–721.