Problem B1.5 Let $\boldsymbol{X}_{p\times n} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \sim \mathcal{N}_{p\times n}(\boldsymbol{\mu}\mathbf{1}_n^T, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$. Since $\bar{\boldsymbol{X}}$ and $\boldsymbol{S} = \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T$ are complete, sufficient statistics for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we simply need to find an unbiased estimator (if it exists) of the parameter $t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

(a) Note that:

- $\mathbb{E}[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T(\bar{\boldsymbol{X}} - \boldsymbol{\mu})] = \frac{\sum \sigma_{ii}^2}{n} = \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{n} = \mathbb{E}\left(\frac{\mathrm{tr}(\boldsymbol{S})}{n(n-1)}\right)$
- $\mathbb{E}[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T(\bar{\boldsymbol{X}} - \boldsymbol{\mu})] = \mathbb{E}[\bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}} - 2\boldsymbol{\mu}^T\bar{\boldsymbol{X}} + \boldsymbol{\mu}^T\boldsymbol{\mu}] = \mathbb{E}(\bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}}) - \boldsymbol{\mu}^T\boldsymbol{\mu}$,

and so the estimator $\delta_1 = \bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}} - \frac{\mathrm{tr}(\boldsymbol{S})}{n(n-1)}$ is unbiased, and so therefore, UMVU for $\boldsymbol{\mu}^T\boldsymbol{\mu}$.

(b) Similarly, notice:

- $\mathbb{E}[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T] = \frac{\boldsymbol{\Sigma}}{n} = \mathbb{E}\left(\frac{\boldsymbol{S}}{n(n-1)}\right)$
- $\mathbb{E}[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^T] = \mathbb{E}[\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^T - 2\boldsymbol{\mu}\bar{\boldsymbol{X}}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T]^T = \mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$,

and so the estimator $\delta_2 = \bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^T - \frac{\boldsymbol{S}}{n(n-1)}$ is UMVU for $\boldsymbol{\mu}^T\boldsymbol{\mu}$.

(c) Let $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then we can write $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I}_p)$. Thus

- $\bar{\boldsymbol{X}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\bar{\boldsymbol{Z}}$
- $\mathbb{E}\left\{[(\boldsymbol{Z} - \bar{\boldsymbol{Z}})(\boldsymbol{Z} - \bar{\boldsymbol{Z}})^T]^{-1}\right\} = c\boldsymbol{I}_p$, and so:

$$\begin{aligned}
\mathbb{E}(\boldsymbol{S}^{-1}) &= \mathbb{E}\left\{[(\boldsymbol{X} - \bar{\boldsymbol{X}})(\boldsymbol{X} - \bar{\boldsymbol{X}})^T]^{-1}\right\} \\
&= \mathbb{E}\left\{\boldsymbol{\Sigma}^{-1/2}[(\boldsymbol{Z} - \bar{\boldsymbol{Z}})(\boldsymbol{Z} - \bar{\boldsymbol{Z}})^T]^{-1}\boldsymbol{\Sigma}^{-1/2}\right\} \\
&= \boldsymbol{\Sigma}^{-1/2}\mathbb{E}\left\{[(\boldsymbol{Z} - \bar{\boldsymbol{Z}})(\boldsymbol{Z} - \bar{\boldsymbol{Z}})^T]^{-1}\right\}\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{\Sigma}^{-1/2}(c\boldsymbol{I}_p)\boldsymbol{\Sigma}^{-1/2} \\
&= c\boldsymbol{\Sigma}^{-1},
\end{aligned}$$

which proves unbiasedness. This estimator is of course UMVU because it is a function of the sufficient statistic $\boldsymbol{S}$. ♦

Problem B1.7 Let $\boldsymbol{X}_{p\times n} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \sim \mathcal{N}_{p\times n}(\boldsymbol{\mu}\mathbf{1}_n^T, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$. Then $\boldsymbol{Z}_{p\times 1} = \boldsymbol{X}\boldsymbol{B}_{n\times 1}$, where $\boldsymbol{B} = \mathbf{1}_n/\sqrt{n}$. Applying Theorem 1 on the bottom of page B1.13 gives:

$$\boldsymbol{Z} \sim \mathcal{N}_p\left((\boldsymbol{\mu}\mathbf{1}_n^T)\mathbf{1}_n/\sqrt{n}, \boldsymbol{\Sigma} \otimes \frac{1}{n}\mathbf{1}_n^T\mathbf{1}_n\right) \;=\; \mathcal{N}_p(\sqrt{n}\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

indicating that the covariance of $\boldsymbol{Z}$ is $\boldsymbol{\Sigma}$. But since we assumed $\mathrm{diag}(\boldsymbol{\Sigma}) = \mathbf{1}_p$ to begin with, then $\boldsymbol{\Sigma}$ is also the correlation matrix. ♦

Problem B1.9 Let $\overset{\perp}{\boldsymbol{P}} = \boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^T/n$ and $\overset{\perp}{\boldsymbol{Q}} = \boldsymbol{I}_p - \mathbf{1}_p\mathbf{1}_p^T/p$.

(a) Note that $\overset{\perp}{\boldsymbol{P}}\mathbf{1}_n = (\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^T/n)\mathbf{1}_n = \mathbf{0}$. So:

$$\begin{pmatrix} \sum_{j=1}^n \boldsymbol{Y}_{1j} \\ \vdots \\ \sum_{j=1}^n \boldsymbol{Y}_{pj} \end{pmatrix} = \boldsymbol{Y}_{p\times n}\mathbf{1}_n = \overset{\perp}{\boldsymbol{Q}}\boldsymbol{X}\overset{\perp}{\boldsymbol{P}}\mathbf{1}_n = \mathbf{0}_p.$$

Moreover,

$$\left(\sum_{i=1}^p \boldsymbol{Y}_{i1}, \ldots, \sum_{i=1}^p \boldsymbol{Y}_{in}\right) = \mathbf{1}_p^T\boldsymbol{Y} = \mathbf{1}_p^T\overset{\perp}{\boldsymbol{Q}}\boldsymbol{X}\overset{\perp}{\boldsymbol{P}} = \mathbf{0}_n.$$

so $\sum_{j=1}^n Y_{ij} = \sum_{i=1}^p Y_{ij} = 0 \;\; \forall i, j$.

(b) From Theorem 1 on the bottom of page B1.13, we have $\boldsymbol{Y} \sim \mathcal{N}_{p\times n}(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}^{\boldsymbol{Y}} \otimes \Delta^{\boldsymbol{Y}})$, where

$$\boldsymbol{\Sigma}^{\boldsymbol{Y}} = \overset{\perp}{\boldsymbol{Q}}\boldsymbol{\Sigma}\overset{\perp}{\boldsymbol{Q}}^T = \overset{\perp}{\boldsymbol{Q}}\boldsymbol{\Sigma}\overset{\perp}{\boldsymbol{Q}}; \quad \Delta^{\boldsymbol{Y}} = \overset{\perp}{\boldsymbol{P}}^T\Delta\overset{\perp}{\boldsymbol{P}} = \overset{\perp}{\boldsymbol{P}}\Delta\overset{\perp}{\boldsymbol{P}}$$

Thus:

$$\begin{aligned} \boldsymbol{\Sigma}^{\boldsymbol{Y}} &= (\boldsymbol{I}_p - \mathbf{1}_p\mathbf{1}_p^T/p)\boldsymbol{\Sigma}(\boldsymbol{I}_p - \mathbf{1}_p\mathbf{1}_p^T/p) \\ &= \boldsymbol{\Sigma} - \frac{\mathbf{1}_p\mathbf{1}_p^T}{p}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\frac{\mathbf{1}\mathbf{1}_p^T}{p} + \frac{\mathbf{1}_p\mathbf{1}_p^T}{p}\boldsymbol{\Sigma}\frac{\mathbf{1}_p\mathbf{1}_p^T}{p} \\ &= \boldsymbol{\Sigma} - \frac{\mathbf{1}_p\mathbf{1}_p^T}{p}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\frac{\mathbf{1}\mathbf{1}_p^T}{p} + \sigma_{..}\mathbf{1}_p\mathbf{1}_p^T \end{aligned}$$

Now for $1 \le i, i' \le p$, we have for $\sigma_{ii'}^{\boldsymbol{Y}} \overset{\text{def}}{=} (\boldsymbol{\Sigma}^{\boldsymbol{Y}})_{ii'}$:

$$\begin{aligned} \sigma_{ii'}^{\boldsymbol{Y}} &= \boldsymbol{\Sigma}_{ii'} - \sum_j \left[\left(\frac{\mathbf{1}_p\mathbf{1}_p^T}{p}\right)_{ij} \boldsymbol{\Sigma}_{ji'}\right] - \sum_j \left[\boldsymbol{\Sigma}_{ij}\left(\frac{\mathbf{1}_p\mathbf{1}_p^T}{p}\right)_{ji'}\right] + \sigma_{..}(\mathbf{1}_p\mathbf{1}_p^T)_{ii'} \\ &= \sigma_{ii'} - \frac{1}{p}\sum_j \sigma_{ji'} - \frac{1}{p}\sum_j \sigma_{ij} + \sigma_{..} \\ &= \sigma_{ii'} - \sigma_{.i'} - \sigma_{i.} + \sigma_{..}, \end{aligned}$$

as desired. The algebra for $\Delta_{jj'}^{\boldsymbol{Y}}$ is the essentially the same.

(c) For $1 \le i \le p$:

$$\begin{aligned} \sum_i \sigma_{ii'}^{\boldsymbol{Y}} &= \sum_i \sigma_{ii'} - \sum_i \sigma_{.i'} - \sum_i \sigma_{i.} + \sum_i \sigma_{..} \\ &= p\sigma_{.i'} - p\sigma_{.i'} - p\sigma_{..} + p\sigma_{..} \\ &= 0. \end{aligned}$$

Similar algebra shows $\sum_j \Delta_{jj'}^{\boldsymbol{Y}} = 0$. ♦

Problem B1.11 By definition, we have:

$$t = \text{sign}(\bar{x})\sqrt{n-1}\frac{||\hat{\boldsymbol{x}}||}{||\overset{\perp}{\boldsymbol{x}}||} = \sqrt{n-1}\cot A$$

The one-sided t-test at level $\alpha = .05$ rejects for $t \geq t_{n-1}^{(.95)}$, where $t_{n-1}^{(.95)}$ is the .95 quantile of the t-distribution on $n-1$ degrees of freedom. So

$$t \geq t_{n-1}^{(.95)} \implies \sqrt{n-1}\cot A \geq t_{n-1}^{(.95)} \implies \boxed{A \leq \arctan\left(\frac{\sqrt{n-1}}{t_{n-1}^{(.95)}}\right) \triangleq A_{n,05}}.$$

```
> n <- c(10, 20, 40, 80)
> t.95 <- qt(.95, n-1)
> print(t.95)
[1] 1.8331 1.7291 1.6849 1.6644
> A05.rad <- atan(sqrt(n-1)/t.95)
> print(A05.rad)
[1] 1.0223 1.1931 1.3073 1.3857
> A05.deg <- A05.rad*(180/pi)
> print(A05.deg)
[1] 58.573 68.362 74.901 79.394
```

These values are summarized in the following table:

n	10	20	40	80
$A_{n,05}$	$58.6°$	$68.4°$	$74.9°$	$79.4°$

Problem B2.3 Let $\boldsymbol{T}$ be a $p \times p$ upper triangular matrix, and define $\boldsymbol{S} = \boldsymbol{T}^T\boldsymbol{T}$. As suggested by the hint, we prove the lemma by induction. The case $j = 1$ is trivial, so we also consider the case $j = 2$ to be thorough.

- Case $j = 1$ (immediately follows): Then $\boldsymbol{T} = t_{11}$ is a scalar, with $\boldsymbol{S} = t_{11}^2$. Then $\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{T}} = 2t_{11}$, implying

  $$J(\boldsymbol{T} \to \boldsymbol{S}) = [J(\boldsymbol{S} \to \boldsymbol{T})]^{-1} = [\det(\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{T}})]^{-1} = (2t_{11})^{-1},$$

  and so the identity holds for the case $j = 1$.

- Case $j = 2$: Let $\boldsymbol{T} = \begin{pmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{pmatrix}$. Then:

  $$\boldsymbol{S} = \boldsymbol{T}^T\boldsymbol{T} = \begin{pmatrix} t_{11} & 0 \\ t_{12} & t_{22} \end{pmatrix}\begin{pmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{pmatrix} = \begin{pmatrix} t_{11}^2 & t_{11}t_{12} \\ t_{11}t_{12} & t_{12}^2 + t_{22}^2 \end{pmatrix}$$

  Because $\boldsymbol{S} = \boldsymbol{T}^T\boldsymbol{T}$ is symmetric, it has $\frac{p(p+1)}{2} = \frac{2(2+1)}{2} = 3$ distinct parameters that can be indexed by the upper triangular part of the matrix. So let $\boldsymbol{S}$ have coordinates

$$\boldsymbol{S} = \begin{pmatrix} s_{11} & s_{12} \\ \cdot & s_{22} \end{pmatrix}$$

Now consider the derivative matrix:

$$\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{T}} = (\nabla s_{11}, \nabla s_{12}, \nabla s_{22}) = \begin{pmatrix} \partial s_{11}/\partial t_{11} & \partial s_{12}/\partial t_{11} & \partial s_{22}/\partial t_{11} \\ \partial s_{11}/\partial t_{12} & \partial s_{12}/\partial t_{12} & \partial s_{22}/\partial t_{12} \\ \partial s_{11}/\partial t_{22} & \partial s_{12}/\partial t_{22} & \partial s_{22}/\partial t_{22} \end{pmatrix}$$

Note the order of the coordinates in which the derivatives are taken. (Start with element s_{1j} and proceed down the column until the main diagonal is reached.)

Therefore:

$$\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{T}} = \begin{pmatrix} 2t_{11} & t_{12} & 0 \\ 0 & t_{11} & 2t_{12} \\ 0 & 0 & 2t_{22} \end{pmatrix}$$

Then $J(\boldsymbol{S} \to \boldsymbol{T}) = \det(\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{T}}) = 4t_{11}^2 t_{22}$, implying $J(\boldsymbol{T} \to \boldsymbol{S}) = J(\boldsymbol{S} \to \boldsymbol{T})^{-1} = (4t_{11}^2 t_{22})^{-1}$, which verifies the lemma for the case $j = 2$.

- Case $j = (p+1)$: Now suppose the lemma holds for $j = 1, \cdots, p$. Consider the $(p+1) \times (p+1)$ upper triangular matrix $\boldsymbol{T}_{p+1}$:

$$\boldsymbol{T}_{p+1} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1p} & t_{1,p+1} \\ 0 & t_{22} & \cdots & t_{2p} & t_{2,p+1} \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & t_{pp} & t_{p,p+1} \\ 0 & 0 & 0 & 0 & t_{p+1,p+1} \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{T}_p & \boldsymbol{t}_{(1:p),p+1} \\ \boldsymbol{0}^T & t_{p+1,p+1} \end{pmatrix},$$

where $\boldsymbol{t}_{(1:p),p+1} = (t_{1,p+1}, t_{2,p+1}, \cdots, t_{p,p+1})$ is a $p \times 1$ column vector. Now:

$$\begin{aligned} \boldsymbol{S}_{p+1} &= \boldsymbol{T}_{p+1}^T \boldsymbol{T}_{p+1} \\ &= \begin{pmatrix} \boldsymbol{T}_p^T & \boldsymbol{0} \\ \boldsymbol{t}_{(1:p),p+1}^T & t_{p+1,p+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{T}_p & \boldsymbol{t}_{(1:p),p+1} \\ \boldsymbol{0}^T & t_{p+1,p+1} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{T}_p^T \boldsymbol{T} & \boldsymbol{T}_p^T \boldsymbol{t}_{(1:p),p+1} \\ \boldsymbol{t}_{(1:p),p+1}^T \boldsymbol{T}_p & \boldsymbol{t}_{(1:p),p+1}^T \boldsymbol{t}_{(1:p),p+1} + t_{p+1,p+1}^2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{T}_p^T \boldsymbol{T} & \boldsymbol{T}_p^T \boldsymbol{t}_{(1:p),p+1} \\ \boldsymbol{t}_{(1:p),p+1}^T \boldsymbol{T}_p & \sum_{j=1}^{p+1} t_{j,p+1}^2 \end{pmatrix} \\ &\triangleq \begin{pmatrix} \boldsymbol{S}_p & \boldsymbol{s}_{(1:p),p+1} \\ \cdot & s_{p+1,p+1} \end{pmatrix}, \end{aligned}$$

where $\boldsymbol{s}_{(1:p),p+1} = (s_{1,p+1}, s_{2,p+1}, \ldots, s_{p,p+1})^T$.

Now consider the derivative matrix, where the order of differentiation is described above in induction step $j = 2$:

$$\frac{\partial \boldsymbol{S}_{p+1}}{\partial \boldsymbol{T}_{p+1}} = (\nabla s_{11}, \nabla s_{12}, \nabla s_{22}, \nabla s_{13}, \cdots, \nabla s_{1p}, \nabla s_{2p}, \cdots, \nabla s_{p,p}, \nabla s_{1,p+1}, \nabla s_{2,p+1}, \cdots, \nabla s_{p+1,p+1})$$

Note that the submatrix $\boldsymbol{S}_p$ consists of terms s_{ij}, for $1 \leq i \leq j \leq p$, which are **not** functions of $t_{k,p+1}$. So the first $\frac{p(p+1)}{2}$ columns of $\frac{\partial \boldsymbol{S}_{p+1}}{\partial \boldsymbol{T}_{p+1}}$ is of the form:

$$(\nabla s_{11}, \nabla s_{12}, \nabla s_{22}, \nabla s_{13}, \cdots, \nabla s_{1p}, \nabla s_{2p}, \cdots, \nabla s_{p,p}) = \begin{pmatrix} \frac{\partial \boldsymbol{S}_p}{\partial \boldsymbol{T}_p} \\ \boldsymbol{O}, \end{pmatrix}$$

where $\frac{\partial \boldsymbol{S}_p}{\partial \boldsymbol{T}_p}$ is of dimension $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ and $\boldsymbol{O}$ is matrix of zeros having dimension $(p+1) \times \frac{p(p+1)}{2}$. Therefore, our overall derivative matrix has the form:

$$\frac{\partial \boldsymbol{S}_{p+1}}{\partial \boldsymbol{T}_{p+1}} = \begin{pmatrix} \frac{\partial \boldsymbol{S}_p}{\partial \boldsymbol{T}_p} & \boldsymbol{B} \\ \boldsymbol{O} & \boldsymbol{C}, \end{pmatrix}$$

where

$$\boldsymbol{C} = \begin{pmatrix} \partial s_{1,p+1}/\partial t_{1,p+1} & \partial s_{2,p+1}/\partial t_{1,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{1,p+1} & \partial s_{p+1,p+1}/\partial t_{1,p+1} \\ \partial s_{1,p+1}/\partial t_{2,p+1} & \partial s_{2,p+1}/\partial t_{2,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{2,p+1} & \partial s_{p+1,p+1}/\partial t_{2,p+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \partial s_{1,p+1}/\partial t_{p,p+1} & \partial s_{p,p+1}/\partial t_{p,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{p,p+1} & \partial s_{p+1,p+1}/\partial t_{p,p+1} \\ \partial s_{1,p+1}/\partial t_{p+1,p+1} & \partial s_{2,p+1}/\partial t_{p+1,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{p+1,p+1} & \partial s_{p+1,p+1}/\partial t_{p+1,p+1} \end{pmatrix}$$

consists of the derivatives of the last column of $\boldsymbol{S}_{p+1}$ with respect to the last column of $\boldsymbol{T}_{p+1}$. Now note the following:

(a) The bottom row (except for the last term of $\partial s_{p+1,p+1}/\partial t_{p+1,p+1}$) is 0, since the terms of $\boldsymbol{s}_{(1:p),p+1}$ from the last column of $\boldsymbol{S}_{p+1}$ are **not** functions of $t_{p+1,p+1}$.

(b) The upperleft $p \times p$ corner of $\boldsymbol{C}$ (that is, eliminating the last row and column of $\boldsymbol{C}$) is exactly:

$$\begin{pmatrix} \partial s_{1,p+1}/\partial t_{1,p+1} & \partial s_{2,p+1}/\partial t_{1,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{1,p+1} \\ \partial s_{1,p+1}/\partial t_{2,p+1} & \partial s_{2,p+1}/\partial t_{2,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{2,p+1} \\ \vdots & \vdots & \vdots & \vdots \\ \partial s_{1,p+1}/\partial t_{p,p+1} & \partial s_{p,p+1}/\partial t_{p,p+1} & \cdots & \partial s_{p,p+1}/\partial t_{p,p+1} \end{pmatrix} = \boldsymbol{T}_p,$$

since this portion of the matrix is simply $\frac{\partial \boldsymbol{T}_p^T \boldsymbol{t}_{(1:p),p+1}}{\partial \boldsymbol{t}_{(1:p),p+1}} = \boldsymbol{T}_p$.

(c) We have $s_{p+1,p+1} = \sum_{j=1}^{p+1} t_{j,p+1}^2$. Thus, the last column of $\boldsymbol{C}$ is $\frac{\partial s_{p+1,p+1}}{\partial \boldsymbol{t}_{(1:p+1),p+1}} = 2\boldsymbol{t}_{(1:p+1),p+1} = 2(t_{1,p+1}, t_{2,p+1}, \ldots, t_{p,p+1}, t_{p+1,p+1})^T$.
Hence,

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{T}_p & 2\boldsymbol{t}_{(1:p),p+1} \\ \boldsymbol{0}^T & 2t_{p+1,p+1} \end{pmatrix}.$$

Then since $\frac{\partial \boldsymbol{S}_{p+1}}{\partial \boldsymbol{T}_{p+1}}$ and $\boldsymbol{C}$ are both upper triangular, we have:

$$\begin{aligned}
J(\boldsymbol{S} \to \boldsymbol{T}) &= \det\left(\frac{\partial \boldsymbol{S}_{p+1}}{\partial \boldsymbol{T}_{p+1}}\right) \\
&= \det\begin{pmatrix} \frac{\partial \boldsymbol{S}_p}{\partial \boldsymbol{T}_p} & \boldsymbol{B} \\ \boldsymbol{O} & \boldsymbol{C}, \end{pmatrix} \\
&= \det\left(\frac{\partial \boldsymbol{S}_p}{\partial \boldsymbol{T}_p}\right)\det(\boldsymbol{C}) \\
&\overset{\text{(induction hypothesis)}}{=} \left[2^p \prod_{j=1}^{p} t_{jj}^{p-j+1}\right] \det(\boldsymbol{T}_p)(2t_{p+1,p+1}) \\
&\overset{(\boldsymbol{T}_p \text{ upper triangular})}{=} \left[2^{p+1} \prod_{j=1}^{p} t_{jj}^{p-j+1}\right] \left[\prod_{j=1}^{p} t_{jj}\right] t_{p+1,p+1} \\
&= 2^{p+1} \prod_{j=1}^{p+1} t_{jj}^{p+1-j+1},
\end{aligned}$$

and taking inverses gives $J(\boldsymbol{T} \to \boldsymbol{S})$, as desired. This completes the proof. ♦

Problem B3.8: Suppose $\boldsymbol{S} \sim \mathcal{W}(\boldsymbol{I}_p; n, p)$. From the change of variables formula, we then have:

$$\begin{aligned}
f(\boldsymbol{D}, \boldsymbol{R}) &= f^{\boldsymbol{S}}(\boldsymbol{S}) J(\boldsymbol{S} \to \boldsymbol{D}, \boldsymbol{R}) \\
&\overset{(1)}{=} f^{\boldsymbol{S}}(\boldsymbol{S}) 2^P |\boldsymbol{D}|^p \\
&\overset{(2)}{=} \left(c_3 |\boldsymbol{S}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\operatorname{tr} S}\right) 2^p |\boldsymbol{D}|^p
\end{aligned}$$

where equality (1) follows from the lemma on the top of page B3.14 and equality (2) uses the Wishart density (Theorem 2, page B3.1)

$$f^{\boldsymbol{S}}(\boldsymbol{s}) = c_3 |\boldsymbol{s}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})} / |\boldsymbol{\Sigma}|^{n/2}$$

with $\boldsymbol{\Sigma} \equiv \boldsymbol{I}_p$. Moreover, since $\boldsymbol{S} = \boldsymbol{DRD}$:

- $|\boldsymbol{S}|^{\frac{n-1-p}{2}} = |\boldsymbol{D}|^{n-p-1} |\boldsymbol{R}|^{\frac{n-p-1}{2}}$
- $\operatorname{tr}(\boldsymbol{S}) = \operatorname{tr}(\boldsymbol{DRD}) = \operatorname{tr}(\boldsymbol{D}^2\boldsymbol{R}) = \sum_i \sum_j (\boldsymbol{D}^2)_{ij} R_{ij} = \sum_i \boldsymbol{D}_{ii}^2$, since $\boldsymbol{D}$ is diagonal and the correlation matrix has diagonal elements $\boldsymbol{R}_{ii} \equiv 1$. Continuing equality (2), we have:
$$f(\boldsymbol{D}, \boldsymbol{R}) = c_3 2^p |\boldsymbol{R}|^{\frac{n-p-1}{2}} \prod_{i=1}^{n} \left[\boldsymbol{D}_{ii}^{n-1} e^{-\frac{1}{2}\boldsymbol{D}_{ii}^2}\right]$$

which factors as separable functions of $\boldsymbol{D}_{11}, \ldots, \boldsymbol{D}_{nn}$ and $\boldsymbol{R}$, and so these terms are independent. If $\boldsymbol{\Sigma} \neq \boldsymbol{I}_p$, then we would have to worry about the exponential term $e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})}$ from the Wishart distribution. In this case, we cannot reduce $\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})$ to separable functions of $\boldsymbol{D}$ and $\boldsymbol{R}$. ♦

Problem B4.6: The R code for computing Hotelling's T^2 test is found below (thanks to Bokyung Choi for his code).

(a) Hotelling's two-sample T^2 statistic for the first 10 genes is 25.8753, with a p-value of .016.

(b) After removing the second gene, we have $T^2 = 11.5817$, with a p-value of .3147.

(c) The p-values from the individual t-tests are found in the R code below. The list indicates that Gene 2 is significant, which explains why Hotelling's test gives markedly different results between parts (a) and (b).

```
source("prostdata.r")
TSQtest = function(data)
{
   X = as.matrix(data[, colnames(prostdata) == "0"])
   Y = as.matrix(data[, colnames(prostdata) == "1"])
   Xmean = as.matrix(rowMeans(X))
   Ymean = as.matrix(rowMeans(Y))

   n1 = ncol(X)
   n2 = ncol(Y)
   n = n1 + n2
   p = nrow(X)
   S1 = matrix(0, p, p)
   S2 = matrix(0, p, p)
   for(i in 1:n1)
   S1 = S1 + (X[, i, drop = F] - Xmean) %*% t(X[, i, drop = F] - Xmean)
   for(i in 1:n2)
   S2 = S2 + (Y[, i, drop = F] - Ymean) %*% t(Y[, i, drop = F] - Ymean)
   Tsq = t(Ymean - Xmean)%*%solve((n/n1/n2 * (S1+S2)/(n-2)), diag(p))%*%(Ymean - Xmean)
   pvalue = 1 - pf(Tsq * (n-p-1)/p/(n-2), p, n-p-1)
   return(pvalue)
}
(a)
> TSQtest(prostdata[1:10, ])
[,1]
[1,] 0.01597193
(b)
> TSQtest(prostdata[c(1, 3:10), ])
[,1]
[1,] 0.3146653
(c)
> for (i in 1:10) print(t.test(prostdata[i, 1:50], prostdata[i,
+ 51:102])[[3]])
[1] 0.1409327
[1] 0.0003652258
[1] 0.9778594
[1] 0.2579435
[1] 0.8884974
[1] 0.3377298
[1] 0.2854839
[1] 0.1968688
[1] 0.2198858
```

```
[1] 0.2396724
```

Problem B4.8: Here, we assume $n = n_1 + n_2$. Hotelling's two-sample T^2-statistic is given by:

$$T^2 \stackrel{\text{def}}{=} (\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})^T \left(\frac{n}{n_1 n_2} \frac{\boldsymbol{S}_1 + \boldsymbol{S}_2}{n-2} \right)^{-1} (\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})$$

We know that $\bar{\boldsymbol{x}}$, $\bar{\boldsymbol{y}}$, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are all statistically independent of each other. Therefore:

(a) Since $\bar{\boldsymbol{x}} \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}/n_1)$ and $\bar{\boldsymbol{y}} \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}/n_2)$, we have:

$$\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}} \;\sim\; \mathcal{N}_p(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1; \boldsymbol{\Sigma}/n_2 + \boldsymbol{\Sigma}/n_1) = \mathcal{N}_p(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1, \frac{n}{n_1 n_2}\boldsymbol{\Sigma}),$$

(b) Furthermore, $\boldsymbol{S}_1 \sim \mathcal{W}(\boldsymbol{\Sigma}; n_1 - 1, p)$ and $\boldsymbol{S}_2 \sim \mathcal{W}(\boldsymbol{\Sigma}; n_2 - 1, p)$, and so:

$$\boldsymbol{S}_1 + \boldsymbol{S}_2 \;\sim\; \mathcal{W}(\boldsymbol{\Sigma}; n_1 + n_2 - 2, p) = \mathcal{W}(\boldsymbol{\Sigma}; n-2, p)$$

Applying the theorem on the top of page B4.18 to $(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})$, $(\boldsymbol{S}_1 + \boldsymbol{S}_2)$, $c = \frac{n}{n_1 n_2}$, and $m = n - 2$, we then have:

$$T^2 \;=\; (\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})^T \left(\frac{n}{n_1 n_2} \frac{\boldsymbol{S}_1 + \boldsymbol{S}_2}{n-2} \right)^{-1} (\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}}) \;\sim\; \frac{(n-2)p}{n-p-1} F_{p,\, n-p-1}(\delta^2),$$

where

$$\delta^2 = \frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) n_1 n_2}{n}. \;\blacklozenge$$

Problem B4.9: Let $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. We can write the Kullback-Leibler (KL) divergence as:

$$KL(\boldsymbol{X}||\boldsymbol{Y}) \;=\; \int_{\mathbb{R}^n} f^{\boldsymbol{X}}(\boldsymbol{z}) \log \frac{f^{\boldsymbol{X}}(\boldsymbol{z})}{f^{\boldsymbol{Y}}(\boldsymbol{z})} d\boldsymbol{z} \;=\; \mathbb{E}_{f^{\boldsymbol{X}}} \left\{ \log \frac{f^{\boldsymbol{X}}(\boldsymbol{Z})}{f^{\boldsymbol{Y}}(\boldsymbol{Z})} \right\} \tag{1}$$

The exponential term of $f^{\boldsymbol{X}}(\boldsymbol{z})$ is

$$-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_1) \;=\; -\frac{1}{2}[\boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z} - 2\boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1], \tag{2}$$

and likewise, the exponential of $f^{\boldsymbol{Y}}(\boldsymbol{z})$ is

$$-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_2) \;=\; -\frac{1}{2}[\boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z} - 2\boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 + \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2] \tag{3}$$

Subtracting Eqn. (3) from Eqn. (2) then yields:

$$\begin{aligned}\log \frac{f^{\boldsymbol{X}}(\boldsymbol{z})}{f^{\boldsymbol{Y}}(\boldsymbol{z})} &= \boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \frac{1}{2} \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 \\ &= \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z} - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z} - \frac{1}{2} \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2\end{aligned}$$

Thus,

$$\begin{aligned}KL(\boldsymbol{X}||\boldsymbol{Y}) &= \mathbb{E}_{f^{\boldsymbol{X}}} \left\{ \log \frac{f^{\boldsymbol{X}}(\boldsymbol{Z})}{f^{\boldsymbol{Y}}(\boldsymbol{Z})} \right\} \\ &= \mathbb{E}_{f^{\boldsymbol{X}}} \left\{ \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} - \frac{1}{2} \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 \right\} \\ &= \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \frac{1}{2} \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 \} \\ &= \frac{1}{2} \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 + \frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 \\ &= \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= \frac{1}{2} \Delta^2. \ \blacklozenge\end{aligned}$$