P-set 3(s23) (For your reference for this course only. Do not circulate or post.)

1. (a) i. $\bar{\boldsymbol{x}} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ ii. $S = \begin{bmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{bmatrix}$ iii. $S^{-1} = \frac{1}{|S|}\begin{bmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{bmatrix} = \frac{9}{44}\begin{bmatrix} 2 & \frac{10}{3} \\ \frac{10}{3} & 8 \end{bmatrix} = \frac{3}{44}\begin{bmatrix} 6 & 10 \\ 10 & 24 \end{bmatrix}$

   iv. $T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'S^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) = 4[-1 \ \ -1]'\frac{3}{44}\begin{bmatrix} 6 & 10 \\ 10 & 24 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \frac{150}{11}$

   v. Under $H_0 : \mu = [7, 11]'$, T^2 is distributed as $\frac{(n-1)p}{(n-p)}F_{p,n-p} = 3F_{2,2}$.
   Note: the p-value is 0.18 `(1-pf(13.64/3,2,2) = 0.1802885`

   (b) i. $\bar{\boldsymbol{y}} = C\bar{\boldsymbol{x}} = \begin{bmatrix} 16 \\ 4 \end{bmatrix}$ ii. $S_y = C\boldsymbol{S}C' = \frac{1}{3}\begin{bmatrix} 10 & -18 \\ -18 & 50 \end{bmatrix}$

   iii. Under $H_o : \boldsymbol{\mu}_y = \boldsymbol{\mu}_o^*$, Hotelling's T^2 for $\{\boldsymbol{y}_j\} = \frac{150}{11}$, the same as in (a)-iv.

   (c) The T^2 for $\{\boldsymbol{y}_j\}$ is

$$T^2(\boldsymbol{y}) = n(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_y)\boldsymbol{S}_y^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_y) = n(C\bar{\boldsymbol{x}} - C\boldsymbol{\mu})'(C\boldsymbol{S}C')^{-1}(C\bar{\boldsymbol{x}} - C\boldsymbol{\mu}) = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'C'\left(C'^{-1}\boldsymbol{S}^{-1}C^{-1}\right)C(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \qquad (since\ (AB)' = B'A')$$
$$= n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) = T^2(\boldsymbol{x})$$

   which illustrates that T^2 is invariant under non-singular linear transformation.

2. (a) ρ_1^{*2} is the common eigenvalue of symmetric matrices $A, B, C'C$ and their transposes $A', B', C'C$, where

$$A = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \quad B = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \quad CC' = \Sigma_{11}^{1/2}A\Sigma_{11}^{-1/2} = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$$

   (b) ρ_i^{*2} are the common eigenvalues of matrices: A, B, CC' and $C'C$. Of which

$$A = \frac{1}{36}\begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix}\frac{1}{38}\begin{bmatrix} 7 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix} = \frac{1}{1368}\begin{bmatrix} 379 & -33 \\ -58 & 366 \end{bmatrix} \approx \begin{bmatrix} .277 & -.024 \\ -.042 & .268 \end{bmatrix}$$

   and

$$B = \frac{1}{38}\begin{bmatrix} 7 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}\frac{1}{36}\begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix} = \frac{1}{1368}\begin{bmatrix} 405 & -45 \\ -20 & 340 \end{bmatrix} \approx \begin{bmatrix} .296 & -.032 \\ -.015 & .249 \end{bmatrix}$$

   From the eigenvalues of A we obtain $\rho_1^* = \sqrt{\rho_1^{*2}} = \sqrt{0.3046} = 0.55, \quad \rho_2^* = \sqrt{\rho_2^{*2}} = \sqrt{0.2400} = 0.49$

   Remarks For computational purpose, it is often easier to find the eigenvalues of A or B instead of that of CC' or $C'C$. Here we will also need to find the eigenvectors of A and B in Part (c) below.

   (c) $U_i = \boldsymbol{a}_i'\boldsymbol{X}_1$, $\boldsymbol{a}_i$ is an eigenvector of A with eigenvalue ρ_i^{*2}, subject to $Var(U_i) = \boldsymbol{a}_i'\Sigma_{11}\boldsymbol{a}_i = 1$.
   Obtain (e.g. in `R`) and normalize the eigenvectors of A, we have

$$\boldsymbol{a}_1 = (0.658, -0.753)'/\sqrt{4.318} = (0.32, -0.36)', \quad \boldsymbol{a}_2 = (0.26, 0.40)'$$

   $V_i = \boldsymbol{b}_i'\boldsymbol{X}_2$, $\boldsymbol{b}_i$ is an eigenvector of B with eigenvalue ρ_i^{*2}, subject to $Var(V_i) = \boldsymbol{b}_i'\Sigma_{22}\boldsymbol{b}_i = 1$.
   Normalize the eigenvectors of B, we obtain $\boldsymbol{b}_1 = (0.36 - 0.10)'$, $\boldsymbol{b}_2 = (0.23, 0.39)'$.
   Therefore the first and second pairs of canonical variables are

$$(U_1, V_1) = (\boldsymbol{a}_1'\boldsymbol{X}_1, \boldsymbol{b}_1'\boldsymbol{X}_2) = \left(0.32X_{11} - 0.36X_{12},\ 0.36X_{21} - 0.10X_{22}\right),$$
$$(U_2, V_2) = (\boldsymbol{a}_2'\boldsymbol{X}_1, \boldsymbol{b}_2'\boldsymbol{X}_2) = \left(0.20X_{11} + 0.30X_{12},\ 0.23X_{21} + 0.39X_{22}\right).$$

   (d) $E(U_1) = \boldsymbol{a}_1'\boldsymbol{\mu}_1 = 0.32 \times (-3) - 0.36 \times 2 = -1.68$. Similar calculations result in

$$E\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} = E\begin{bmatrix} U_1 \\ U_2 \\ V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} -1.675 \\ 0.015 \\ -0.095 \\ 0.386 \end{bmatrix} \quad Cov\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0.55 & 0 \\ 0 & 1 & 0 & 0.49 \\ 0.55 & 0 & 1 & 0 \\ 0 & 0.49 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \rho_1^* & 0 \\ 0 & 1 & 0 & \rho_2^* \\ \rho_1^* & 0 & 1 & 0 \\ 0 & \rho_2^* & 0 & 1 \end{bmatrix}$$

   (e) The covariance matrix in (d) is based on the properties of the canonical variables:
   Normalization $V(U_i) = V(V_i) = 1$;
   Between and within uncorrelatedness: $Cov(U_1, U_2) = 0$, $Cov(V_1, V_2) = 0$, $Cov(U_i, V_j) = 0$ if $i \neq j$.
   Therefore $Cov(U) = I_2, Cov(V) = I_2$, $Cov(U, V)$ is diagonal.
   Correlation between the $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ variables, which is not very strong:
   $Conv(U_1, V_1) = \rho_1^* = 0.55$, $Cov(U_2, V_2) = \rho_2^* = 0.49$.

```
# Example R commands for Q2
S11=matrix(c(8,2,2,5),2,2); S22=matrix(c(6,-2,-2,7),2,2); S21=matrix(c(3,1,-1,3),2,2); S12=t(S21)
#  Matrix A, B
A=solve(S11)%*%S12%*%solve(S22)%*%S21
B=solve(S22)%*%S21%*%solve(S11)%*%S12
# get rho_i^*
sqrt(eigen(A)$value)      #0.5519301 0.4898610
sqrt(eigen(B)$value)      #same
# get a_i, b_i
av=eigen(A)$vector
```



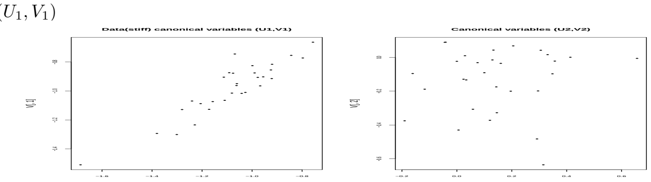
```
a1=av[,1]/sqrt(t(av[,1])%*%S11%*%av[,1])      # 0.3168206 -0.3622269
a2=av[,2]/sqrt(t(av[,2])%*%S11%*%av[,2])      # 0.1962489 0.3016851
bv=eigen(B)$vector
b1=bv[,1]/sqrt(t(bv[,1])%*%S22%*%bv[,1])      # 0.36470579 -0.09506271
b2=bv[,2]/sqrt(t(bv[,2])%*%S22%*%bv[,2])      # 0.2262746 0.3858210
#  E(U_i]
c(-3,2)%*%cbind(a1,a2)
            a1         a2
[1,] -1.674915 0.01462369
# E[V_i]
c(0,1)%*%cbind(b1,b2)
              b1       b2
[1,] -0.09506271 0.385821
```

3. (a) Canonical correlation analysis from R output: $\rho_1^* = 0.913$, $\rho_2^* = 0.068$.

```
stiff = read.table("stiffness.DAT").
X1 = stiff[,1:2]; X2 = stiff[,3:4]
cancor(X1,X2)
...
$xcoef
            [,1]         [,2]
X1 -0.0006687933 -0.001237328
X2  0.0001106253  0.001430402
$ycoef
            [,1]         [,2]
X3 -0.0002497238  0.001573032
X4 -0.0003515941 -0.001453802
$xcenter
      V1       V2
1906.100 1749.533
$ycenter
      V3       V4
1509.133 1724.967
```

   (b) $U_1 = 0.0001(-6.7X_1^{(1)} + 1.1X_2^{(1)})$, $V_1 = 0.0001(-2.5X_1^{(2)} - 3.5X_2^{(2)})$

   (c) Plot (U_1, V_1)




```
U=as.matrix(X1)%*%cancor(X1,X2)$xcoef; V=as.matrix(X2)%*%cancor(X1,X2)$ycoef
par(mfrow=c(1,2))
plot(U[,1],V[,1],cex.lab=.8,cex.axis=.5,cex=.6,pch=16)
title(main="Data(stiff) canonical variables (U1,V1)",cex.main=.8)
plot(U[,2],V[,2],cex.lab=.8,cex.axis=.5,cex=.6,pch=16)
title(main="Canonical variables (U2,V2)",cex.main=.8)
```

   (d) Comments: the first pair of canonical variates capture most of the correlation between the two sets of variables. The correlation of the static measures $\boldsymbol{X}^{(1)}$ and dynamic measures $\boldsymbol{X}^{(2)}$ is well explained by the first canonical variate pair (U_1, V_1), as indicated in the first scatterplot by the obvious linear form of the data with correlation $\rho_1^* = 0.913$. There is not much variation left for the second canonical variate pair (U_2, V_2) with correlation $\rho_2^* = 0.068$, as shown in the feature-less random pattern in the second scatterplot.

4. (a) i. Evaluated from the data, the Hotelling's T^2 is

$$T^2 = (\bar{Y}_{Af} - \bar{Y}_{Apf})'\left[(\frac{1}{n_1} + \frac{1}{n_2})S_{pool}\right]^{-1}(\bar{Y}_{Af} - \bar{Y}_{Apf}) = 55.88$$

   where $\bar{Y}_{Af} = [Y_{1,Af}\ Y_{2,Af}]'$ is the measurement vector $[Y_1\ Y_2]$ for Species Af, etc.

$$Under \quad H_0 : \mu_{Af} = \mu_{Apf}, \qquad T^2 \sim \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)}F_{p,n_1+n_2-p-1} = \frac{26}{12}F_{2,12}$$

   with $n_1 = 9, n_2 = 6, p = 2$. Thus $55.88(12/26) = 25.8 \sim F_{2,12} > F_{2,12,\alpha=0.01} = 6.9$.
   The null hypothesis of equality of the means is rejected at test level much less than $\alpha = 0.01$.

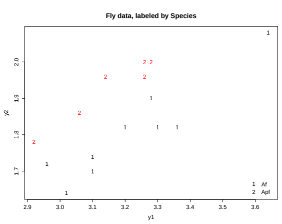
   ii. The results would be the same using the original values (X_1, X_2), since $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, and the Hotelling T^2 statistic preserves under non-singular linear transformations (as we have shown in the previous exercise Q1(c)).

```
# R commands for (a) Q4
fly=read.table("fly.dat")
y1=fly[,2]+fly[,3]; y2=fly[,3]; y=cbind(y1,y2)
## By hand Hotelling's T2
yAf=y[1:9,]; barAf=colMeans(yAf)
yApf=y[10:15,]; barApf=colMeans(yApf)
diffmean = barAf - barApf
n1=9; n2=6
Spool=(n1-1)/(n1+n2-2)*cov(yAf)+(n2-1)/(n1+n2-2)*cov(yApf)
T2=t(diff)%*%solve((1/n1+1/n2)*Spool)%*%diff  # = 55.8807
## By manova Hotelling's T2
summary(manova(y~Species),test="Hotelling")
            Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
  Species    1           4.2985   25.791      2     12 4.519e-05 ***
  Residuals 13
qf(.99,2,12)  # = 6.9
```

(b) Marginal t-test, unequal variance: For Y_1, the univariate 2-sample t-test on the two species does not reject the equal mean null hypothesis with a p-value $= 0.49$. Y_2 has p-value $= 0.049$ from the t-test, the null would be rejected at $\alpha = 0.05$ but accepted at $\alpha = 0.01$ or any $\alpha < 0.488$.
Marginal t-test, equal variance : The equal mean null hypothesis is not rejected by the equal variance t-test for both Y_1 (p-value 0.52) and Y_2 (p-value -.066).

```
t.test(y1[1:9], y1[10:15])   # t = 0.7132, df = 12.943, p-value = 0.4884 (unequal variance)
t.test(y1~Species)   # same as above
t.test(y1~Species, var.equal=TRUE) # t = 0.661, df = 13, p-value = 0.5202
t.test(y2~Species)  # t = -2.1697, df = 12.967, p-value = 0.0492 (unequal variance)
t.test(y2~Species, var.equal=TRUE) # t = -2.0047, df = 13, p-value = 0.06628
```

(c) We can see from the scatterplot of Y_1 vs. Y_2 that the two species are very similar with respect to their Y_1

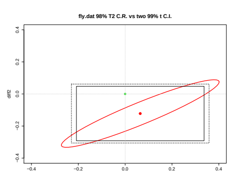


values, and are slightly apart by species with respect to Y_2. Therefore, the individual t-test can barely detect the difference in the means in Y_2, but cannot reject the equality of means for Y_1 at all.

```
# Scatter plots Part(c)
plot(y1,y2,type="n")
text(y1,y2,label=c(rep(1,9),rep(2,6)),col=c(rep(1,9),rep(2,6)))
title(main ="Fly data, labeled by Species")
legend('bottomright',c('Af','Apf'),pch=c("1","2"),bty='n')
```

(d)

```
# Plot t-rectangle and T2 ellipse
library(car)
p=2
# Radius of 98% conf. level by F(p, n1+n2-p-1)
Rad= sqrt((n1+n2-2)/(n1+n2-p-1)*p*qf(0.98,p,n1+n2-p-1))
plot(c(-0.4,0.4), c(-0.4,0.4), xlab="diff1", ylab="diff2", type="n")
# t rectangle, unequal variance
xpt1 = t.test(y1~Species,conf.level=.99)$conf.int[1]; xpt2 = t.test(y1~Species,conf.level=.99)$conf.int[2]
ypt1 = t.test(y2~Species,conf.level=.99)$conf.int[1]; ypt2 = t.test(y2~Species,conf.level=.99)$conf.int[2]
ellipse(diffmean,shape=(1/n1+1/n2)*Spool,radius=Rad,lwd=3)
rect(xpt1,ypt2,xpt2,ypt1,lwd=2)
abline(v=0,lty=3,col="grey"); abline(h=0,lty=3,col="grey"); points(0,0,col=3)
title(main="fly.dat 98% T2 C.R. vs two 99% t C.I.")
# Add t rectangle, equal variance
xpt1e = t.test(y1~Species,conf.level=.99, var.equal=T)$conf.int[1]
xpt2e = t.test(y1~Species,conf.level=.99, var.equal=T)$conf.int[2]
ypt1e = t.test(y2~Species,conf.level=.99, var.equal=T)$conf.int[1]
ypt2e = t.test(y2~Species,conf.level=.99, var.equal=T)$conf.int[2]
rect(xpt1e,ypt2e,xpt2e,ypt1e, lty=2)
```



i. T^2 produces an ellipse confidence region.
ii. Univariate method produces a rectangle. The rectangle is slightly larger under equal variance assumption in producing t statistic induced marginal confidence interval.
iii. The T^2 98% confidence region has $\alpha = 0.02$. For dimension $p = 2$, Bonferroni method uses $\alpha/p = \alpha/2 = 0.01$ to yield $(1-\alpha)100\% = 99\%$ simultaneous C.I.
iv. The origin $\mathbf{0} = (0,0)$ is not in the ellipse but in the rectangle.
The ellipse is more reasonable than the rectangle, since points in the ellipse are close to the mean, according to the more reasonable (Mahalabonis) statistical distance. The rectangles are easier to calculate, however they cover some points quite far from the mean, such as the regions near the upper left and lower right corners (from data one can see that those points are not feasible means).
From the figure, we can see that higher dimension multivariate confidence regions for correlated variables are not well approximated by the intersection of lower dimensional univariate confidence intervals.

5. $\boldsymbol{X}$ and $\boldsymbol{Y}$ have covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ with } \Sigma_{11} = \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}, \ \Sigma_{22} = \begin{bmatrix} 1 & \rho_y \\ \rho_y & 1 \end{bmatrix}, \ \Sigma_{21} = \Sigma_{12} = \begin{bmatrix} r & r \\ r & r \end{bmatrix}.$$

(a) Derive ρ_1^*, the largest canonical correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

$\Sigma_{11}^{-1} = \frac{1}{1-\rho_x^2}\begin{bmatrix} 1 & -\rho_x \\ -\rho_x & 1 \end{bmatrix}$. ρ_1^{*2} is the largest eigenvalue of $A = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} c & c \\ c & c \end{bmatrix}$, with $c = \frac{2r^2}{(1+\rho_x)(1+\rho_y)}$.
From $|A - \lambda I| = 0$ we get A's eigenvalues $2c$ and 0, with e-vectors $\left[\begin{smallmatrix}\alpha \\ \alpha\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}\alpha \\ -\alpha\end{smallmatrix}\right]$. Thus $\rho_1 * = \frac{2r}{\sqrt{1+\rho_x)(1+\rho_y)}}$.

(b) Derive the canonical variate pairs $(U_1, V_1) = (\boldsymbol{a}_1'\boldsymbol{X}, \boldsymbol{b}_1'\boldsymbol{Y})$ corresponding to ρ_1^*, with $\boldsymbol{a}_1'\Sigma_{11}\boldsymbol{a}_1 = 1, \boldsymbol{b}_1'\Sigma_{22}\boldsymbol{b}_1 = 1$.
$U_1 = \left[\begin{smallmatrix}\alpha \\ \alpha\end{smallmatrix}\right]'\boldsymbol{X} = \alpha(X_1 + X_2)$. $\boldsymbol{a}_1'\Sigma_{11}\boldsymbol{a}_1 = 1$ yields $\alpha = \frac{1}{\sqrt{2(1+\rho_x)}}$. Analogously, $V_1 = \beta(Y_1 + Y_2)$, $\beta = \frac{1}{\sqrt{2(1+\rho_y)}}$.

6. (a) Write

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})' = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{x}_j\boldsymbol{x}_j' - \boldsymbol{x}_j\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}\boldsymbol{x}_j' + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}') = \frac{1}{n-1}\left(\sum_{j=1}^{n}\boldsymbol{x}_j\boldsymbol{x}_j' - n\,\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'\right)$$

Note that $\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{x}_j' = \boldsymbol{X}'\boldsymbol{X}$, and $\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{1}_n$, so $\bar{\boldsymbol{x}}' = \frac{1}{n}\boldsymbol{1}_n'\boldsymbol{X}$. Bring these expressions into the above,

$$\boldsymbol{S} = \frac{1}{n-1}\left(\boldsymbol{X}'\boldsymbol{X} - n(\frac{1}{n}\boldsymbol{X}'\boldsymbol{1}_n)(\frac{1}{n}\boldsymbol{1}_n'\boldsymbol{X})\right) = \frac{1}{n-1}\boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n'\right)\boldsymbol{X} = \frac{1}{n-1}\boldsymbol{X}'\boldsymbol{H}\boldsymbol{X}$$

(b) Since $\mathbb{E}(\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c}) = \boldsymbol{A}\,\mathbb{E}(\boldsymbol{Y}) + \boldsymbol{c}$,

$$\begin{aligned}
Cov(\boldsymbol{W}) &= \mathbb{E}\Big([\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c} - \mathbb{E}(\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c})]\,[\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c} - \mathbb{E}(\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c})]'\Big) \\
&= \mathbb{E}\Big([\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c} - \boldsymbol{A}\,\mathbb{E}(\boldsymbol{Y}) - \boldsymbol{c}]\,[\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{c} - \boldsymbol{A}\,\mathbb{E}(\boldsymbol{Y}) - \boldsymbol{c}]'\Big) \\
&= \mathbb{E}\Big([\boldsymbol{A}\{\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})\}]\,[\boldsymbol{A}\{\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})\}]'\Big) = \mathbb{E}\Big(\boldsymbol{A}[\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]\,[\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]'\boldsymbol{A}'\Big) \\
&= \boldsymbol{A}\,\mathbb{E}\Big([\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]\,[\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]'\Big)\boldsymbol{A}' = \boldsymbol{A}\,Cov(\boldsymbol{Y})\boldsymbol{A}'
\end{aligned}$$

where the 4th equality is from $(BC)' = C'B'$.

(c) $\boldsymbol{a}'\boldsymbol{Y}, \boldsymbol{b}'\boldsymbol{W}$ are univariate random variables. Use $\mathbb{E}(\boldsymbol{a}'\boldsymbol{Y}) = \boldsymbol{a}'\mathbb{E}(\boldsymbol{Y})$ and $\boldsymbol{b}'\{\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})\} = \{\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})\}'\boldsymbol{b}$, we have

$$\begin{aligned}
Cov(\boldsymbol{a}'\boldsymbol{Y}, \boldsymbol{b}'\boldsymbol{W}) &= \mathbb{E}\Big([\boldsymbol{a}'\boldsymbol{Y} - \mathbb{E}(\boldsymbol{a}'\boldsymbol{Y})]\,[\boldsymbol{b}'\boldsymbol{W} - \mathbb{E}(\boldsymbol{b}'\boldsymbol{W})]\Big) \\
&= \mathbb{E}\Big([\boldsymbol{a}'\{\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})\}]\,[\boldsymbol{b}'\{\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})\}]\Big) \\
&= \mathbb{E}\Big(\boldsymbol{a}'[\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]\,[\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})]'\boldsymbol{b}\Big) \\
&= \boldsymbol{a}'\mathbb{E}\Big([\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})]\,[\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})']\Big)\boldsymbol{b} = \boldsymbol{a}'Cov(\boldsymbol{Y}, \boldsymbol{W})\boldsymbol{b}
\end{aligned}$$

Since $\boldsymbol{a}'Cov(\boldsymbol{Y}, \boldsymbol{W})\boldsymbol{b}$ is a number which equals to its transpose, and that $Cov(\boldsymbol{Y}, \boldsymbol{W})' = Cov(\boldsymbol{W}, \boldsymbol{Y})$, we have

$$Cov(\boldsymbol{a}'\boldsymbol{Y}, \boldsymbol{b}'\boldsymbol{W}) = \boldsymbol{a}'Cov(\boldsymbol{Y}, \boldsymbol{W})\boldsymbol{b} = (\boldsymbol{a}'Cov(\boldsymbol{Y}, \boldsymbol{W})\boldsymbol{b})' == \boldsymbol{b}'Cov(\boldsymbol{Y}, \boldsymbol{W})'\boldsymbol{a} = \boldsymbol{b}'Cov(\boldsymbol{W}, \boldsymbol{Y})\boldsymbol{a}$$