Date:
Page: 29/10

Wk1 - Quiz 1.

1. $f(x_1, x_2) = \frac{1}{2x_2} e^{-\left(\frac{x_2}{4} + \frac{2x_1}{x_2}\right)}, \quad x_1, x_2 > 0$

(a) $f_{x_2}(x_2) = \int_0^\infty \frac{1}{2x_2} e^{-\left(\frac{x_2}{4} + \frac{2x_1}{x_2}\right)} dx_1$

$= \frac{1}{2x_2} \int_0^\infty e^{-\frac{x_2}{4}} \cdot e^{-\frac{2x_1}{x_2}} \, dx_1$

$= \frac{1}{2x_2} \cdot e^{-\frac{x_2}{4}} \int_0^\infty e^{-\frac{2x_1}{x_2}} dx_1$

$= \frac{1}{2x_2} \cdot e^{-\frac{x_2}{4}} \left[-\frac{x_2}{2} e^{-\frac{2x_1}{x_2}} \right]_0^\infty$

$= \frac{1}{2x_2} \cdot e^{-\frac{x_2}{4}} \left[-\frac{x_2}{2}(0) + \frac{x_2}{2}(1) \right]$

$= \frac{1}{2\cancel{x_2}} e^{-\frac{x_2}{4}} \left[\frac{x_2}{2} \right]$

$f_{x_2}(x_2) = \underline{\underline{\frac{1}{4} \cdot e^{-x_2/4}}}$

(b) $f_{x_1|x_2}(x_1|x_2) = \frac{f_{x_1,x_2}(x_1, x_2)}{f_{x_2}(x_2)}$

$= \frac{\frac{1}{2x_2} \cdot e^{-\left(\frac{x_2}{4} + 2x_1/x_2\right)}}{\frac{1}{4} \cdot e^{-x_2/4}}$

$= \frac{\overset{2}{\cancel{4}} \cancel{e^{-x_2/4}} \cdot e^{-2x_1/x_2}}{\cancel{2} x_2 \cdot \cancel{e^{-x_2/4}}}$

$= \frac{2 \cdot e^{-2x_1/x_2}}{x_2}$

Quiz 1....

Date:
Page:

1. (c) $g^*(X_2) = E(X_1 | X_2 = x_2)$

$$= \int_0^\infty x_1 \cdot f_{X_1|X_2}(x_1|x_2)\, dx_1$$

$$= \int_0^\infty x_1 \cdot \frac{2}{x_2} \cdot e^{-2x_1/x_2}\, dx_1$$

$$= \frac{2}{x_2}\int_0^\infty x_1 . e^{-2x_1/x_2}\, dx_1$$

Integration by parts:

$u = x_1$ | $dv = e^{-2x_1/x_2}\, dx_1$

$du = dx_1$ | $v = \int e^{-2x_1/x_2}\, dx_1 = -\frac{x_2}{2} . e^{-2x_1/x_2}$

(Neglect constants of integration here.)

$$\therefore \int u\,dv = uv - \int v\,du$$

$$= \frac{-x_1x_2}{2} . e^{-2x_1/x_2} - \int \frac{-x_2}{2} e^{-2x_1/x_2}\, dx_1$$

$$= \frac{-x_1x_2}{2} e^{-2x_1/x_2} + \frac{x_2}{2}\int e^{-2x_1/x_2}\, dx_1$$

$$= \frac{-x_1x_2}{2} e^{-2x_1/x_2} - \frac{x_2^2}{4} . e^{-2x_1/x_2}$$

$$= -e^{-2x_1/x_2}\left(\frac{x_1x_2}{2} + \frac{x_2^2}{4}\right)$$

$$\therefore g^*(X_2) = \frac{2}{x_2}\left[-e^{-2x_1/x_2}\left(\frac{x_1x_2}{2} + \frac{x_2^2}{4}\right)\right]_0^\infty$$

$$= \left[-e^{-2x_1/x_2}\left(x_1 + \frac{x_2}{2}\right)\right]_0^\infty$$

$$= 0 + 1 . \frac{x_2}{2}$$

$$= \frac{x_2}{2} \,/\!/$$

Date:
Page:

1.(d). $E[(X_1 - g^*(X_2))^2 \mid X_2 = x_2]$

$= E[(X_1 - \frac{x_2}{2})^2 \mid X_2 = x_2]$

$= E[X_1^2 - X_1 x_2 + \frac{x_2^2}{4} \mid X_2 = x_2]$

$= E(X_1^2 \mid X_2 = x_2) - E(x_2 X_1 \mid X_2 = x_2) + E(\frac{x_2^2}{4} \mid X_2 = x_2)$

$= \frac{x_2^2}{2} - \frac{x_2 \cdot x_2}{2} + \frac{x_2^2}{4}$

$= \frac{x_2^2}{2} - \frac{x_2^2}{2} + \frac{x_2^2}{4}$

Used CAS for $E(X_1^2 \mid X_2 = x_2)$

$= \frac{x_2^2}{4}.$ //

Quiz 1...

Date:

Page:

2. $\Sigma = \begin{bmatrix} 8 & 3 \\ 3 & 9 \end{bmatrix}$

(a) Characteristic polynomial

$= |\Sigma - \lambda I|$

$= \left| \begin{pmatrix} 8 & 3 \\ 3 & 9 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right|$

$= \left| \begin{pmatrix} 8 & 3 \\ 3 & 9 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right|$

$= \begin{vmatrix} 8-\lambda & 3-0 \\ 3-0 & 9-\lambda \end{vmatrix}$

$= (8-\lambda)(9-\lambda) - (3)^2$

$= 72 - 8\lambda - 9\lambda + \lambda^2 - 9$

$= \lambda^2 - 17\lambda + 63$ ✓

(b) $\lambda = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ $a = 1$, $b = -17$, $c = 63$

$= \dfrac{17 \pm \sqrt{17^2 - (4 \times 1 \times 63)}}{2}$

$= \dfrac{17 \pm \sqrt{37}}{2} = \underbrace{\dfrac{17 + \sqrt{37}}{2}}_{\lambda_1}, \underbrace{\dfrac{17 - \sqrt{37}}{2}}_{\lambda_2}$

Now for the eigenvectors: $(A - \lambda I)\vec{x} = 0$

For λ_1: $(\Sigma - \lambda_1 I)\vec{x} = 0$

$\begin{pmatrix} 8 - \frac{17+\sqrt{37}}{2} & 3 \\ 3 & 9 - \frac{17+\sqrt{37}}{2} \end{pmatrix} \vec{x} = 0$

$\begin{pmatrix} \frac{-1-\sqrt{37}}{2} & 3 \\ 3 & \frac{1-\sqrt{37}}{2} \end{pmatrix} \vec{x} = 0$

Date:
Page:

$$\frac{-1-\sqrt{37}}{2}x_1 + 3x_2 = 0$$

$$3x_1 + \frac{1-\sqrt{37}}{2}x_2 = 0$$

$$x_1 + \frac{1-\sqrt{37}}{6}x_2 = 0$$

$$x_1 = -\left(\frac{1-\sqrt{37}}{6}\right)x_2 = \frac{-1+\sqrt{37}}{6}x_2$$

$$\therefore \vec{x} = \begin{pmatrix} \frac{-1+\sqrt{37}}{6} \\ 1 \end{pmatrix}$$

$$\therefore \vec{x}_{norm} \approx \begin{pmatrix} 0.6464 \\ 0.763 \end{pmatrix}$$

Used Sympy.

For λ_2: $(\Sigma - \lambda_2 I)\vec{x} = 0$

$$\begin{pmatrix} 8-\lambda_2 & 3 \\ 3 & 9-\lambda_2 \end{pmatrix}\vec{x} = 0$$

$$\begin{pmatrix} 8-\frac{17-\sqrt{37}}{2} & 3 \\ 3 & 9-\frac{17-\sqrt{37}}{2} \end{pmatrix}\vec{x} = 0$$

$$\begin{pmatrix} \frac{-1+\sqrt{37}}{2} & 3 \\ 3 & \frac{1+\sqrt{37}}{2} \end{pmatrix}\vec{x} = 0$$

$$\frac{-1+\sqrt{37}}{2}x_1 + 3x_2 = 0$$

$$3x_1 + \frac{1+\sqrt{37}}{2}x_2 = 0$$

$$x_1 + \frac{1+\sqrt{37}}{6}x_2 = 0$$

$$x_1 = -\left(\frac{1+\sqrt{37}}{6}\right)x_2 = \frac{-1-\sqrt{37}}{6}x_2$$

$$\therefore \vec{x} = \begin{pmatrix} \frac{-1-\sqrt{37}}{6} \\ 1 \end{pmatrix}$$

$$\therefore \vec{x}_{norm} \approx \begin{pmatrix} -0.763 \\ 0.6464 \end{pmatrix}$$

Used Sympy.

Quiz 1...

2.(c). $X^{\frac{1}{2}} = \sum_{i=1}^{p} \sqrt{\lambda_i}\, \vec{e}_i \cdot \vec{e}_i^{\,T}$

$= \sqrt{\lambda_1}\, \vec{e}_1 \cdot \vec{e}_1^{\,T} + \sqrt{\lambda_2}\, \vec{e}_2 \vec{e}_2^{\,T}$

$= \sqrt{\frac{17+\sqrt{37}}{2}} \cdot \vec{e}_1 \cdot \vec{e}_1^{\,T} + \sqrt{\frac{17-\sqrt{37}}{2}} \cdot \vec{e}_2 \cdot \vec{e}_2^{\,T}$

from (b)

$\approx \begin{bmatrix} 2.7796 & 0.5232 \\ 0.5232 & 2.9540 \end{bmatrix}$

(d). $X^{-1} = \sum_{i=1}^{p} \lambda_i^{-\frac{1}{2}} \cdot \vec{e}_i \vec{e}_i^{\,T}$

$= \lambda_1^{-\frac{1}{2}} \vec{e}_1 \vec{e}_1^{\,T} + \lambda_2^{-\frac{1}{2}} \cdot \vec{e}_2 \vec{e}_2^{\,T}$

$\approx \begin{bmatrix} 0.3722 & -0.0659 \\ -0.0659 & 0.3502 \end{bmatrix}$

Date:
Page:

3. $S = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}$ $n = 12$ $\vec{x} = \begin{bmatrix} -1.2 \\ -3.9 \end{bmatrix}$ $\alpha = 0.1$

For eigenvalues:

$|S - \lambda I| = 0$

$\left| \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = 0$

$\begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = 0$

$(6-\lambda)^2 - 1 = 0$

$(6-\lambda-1)(6-\lambda+1) = 0$

$-\lambda+5 = 0, \ -\lambda+7 = 0$

$\underline{\underline{\lambda = 5, 7}}$ ✓

For eigenvectors:

$\lambda = 5$

$(S - \lambda I)\vec{x} = 0$

$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$

$x_1 + x_2 = 0$

$x_1 + x_2 = 0$

$x_1 = -x_2$

$\therefore \vec{x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \div \sqrt{1^2+1^2}$

$= \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ normalised ✓

$\lambda = 7$

$(S - \lambda I)\vec{x} = 0$

$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$

$-x_1 + x_2 = 0$

$x_1 - x_2 = 0$

$x_1 = x_2$

$\therefore \vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \div \sqrt{1^2+1^2}$

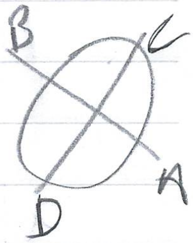
$= \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ normalised ✓

Quiz 1....

3....

3(a) To locate the ellipse:

Centre is at $\vec{\bar{x}}$ i.e. determined by sample mean vector.



then endpoints stretch out from centre, using a scale.

Major axis determined by $\lambda = 7$, and $\vec{e} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$:

endpoint $D = \vec{\bar{x}} - \vec{e}\sqrt{7 \times \text{scale}}$

$C = \vec{\bar{x}} + \vec{e}\sqrt{7 \times \text{scale}}$

Minor axis determined by $\lambda = 5$, and $\vec{e} = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$:

endpoint $B = \vec{\bar{x}} - \vec{e}\sqrt{5 \times \text{scale}}$

$A = \vec{\bar{x}} + \vec{e}\sqrt{5 \times \text{scale}}$

$$\text{Scale} = \sqrt{\lambda \cdot \frac{p(n-1) \cdot F_{1-\alpha, p, n-p}}{n(n-p)}}$$

(b) Same as before, but scale becomes $\chi^2_{1-\alpha, p}$ as this is the distribution when Σ is known.

(c) For simultaneous CIs, for each variable: $\bar{x} \pm \sqrt{\sigma^2 \times \text{scale}}$ from (a)

(d) Bonferroni method:

Use "naive" t-statistic to get CI for each variable but use $\alpha_{new} = \alpha/m$ $\quad m = \#$ linear combinations $= 4$.

$\therefore$ For each variable, $\bar{x} \pm t_{1-\alpha, p} \sqrt{\frac{\sigma^2}{n}}$.

Date:
Page:

4. Pairwise plots & MVN tests.

SW's year built : YES (normal)
SW's length : NO (not normal — looks skewed)

mardia's skewness: NO
mardia's kurtosis: NO

5. $\alpha = 0.05$

2 independent samples:
- bridges in Monogahela (M)
- bridges in Allegheny. (A).

Inference on joint means of year built & length.

$$\vec{\mu}_1 = \begin{pmatrix} \mu_{1,M} \\ \mu_{2,M} \end{pmatrix} \qquad \vec{\mu}_2 = \begin{pmatrix} \mu_{1,A} \\ \mu_{2,A} \end{pmatrix}$$

Means of year built & length for M. — means of year built & length for A.

$H_0: \vec{\mu}_1 = \vec{\mu}_2$

$H_1: \vec{\mu}_1 \neq \vec{\mu}_2$

Unknown Σ. Assume $\Sigma_M = \Sigma_A$ i.e. pooled.

First find: $S_p = \dfrac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$

$$T^2 = (\bar{X}_1 - \bar{X}_2 - \vec{\delta}_0)^T \cdot \left(S_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)^{-1} \cdot (\bar{X}_1 - \bar{X}_2 - \delta_0)$$

$$= (\bar{X}_1 - \bar{X}_2)^T \cdot \left(S_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)^{-1} \cdot (\bar{X}_1 - \bar{X}_2)$$ because $\vec{\delta}_0 = \vec{\mu}_1 - \vec{\mu}_2 = 0$ (under H_0)

$= 1.196022$

$$T^2_{critical} = F_{p,\, n_1+n_2-p-1,\, 1-\alpha} \cdot \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} = 6.366864$$

Quiz1...

5.... $T^2 < T_{critical}$

$\therefore$ Do not reject H_0.

Alternatively, compare F & $F_{critical}$

$$F = T^2 \times \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} \approx 0.5891$$

$$F_{critical} = F_{p,\,n_1+n_2-p-1,\,1-\alpha} \approx 3.1359.$$

$F < F_{critical}$. $\therefore$ Do not reject H_0.

Alternatively, compare p-value of F with level 0.05:

$$\text{p-value} = P(F > 0.5891) \approx 0.56$$

p-value > 0.05 $\therefore$ Do not reject H_0.

6. $\boxed{\text{False}}$ (as seen in Week 1 R demo.)

7. $\boxed{\text{Always}}$ - Only the error has to be normal.

8. $\boxed{\text{Normal}}$

In general, $\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$

↳ Unbiased but not necessarily MLE.

For normal, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$.

↳ this is MLE.

So ρ_{ij} can only be MLE if the $\hat{\Sigma}$ is MLE,

which won't happen unless the dist is normal.

9. $\boxed{\text{Always}}$. $Var(A\vec{X}) = A \cdot Var(\vec{X}) \cdot A^T$ for $A \in M_{q,p}$

10. $\boxed{\text{Always}}$. General property of covariances.

Quiz 1 - Questions 1 & 2

Amanjit Gill

In [44]:
```
from sympy import *
from sympy.integrals import *
```

Question 1

(a)

In [45]:
```
# define symbols
x1 = Symbol('x_1')
x2 = Symbol('x_2')
```

In [46]:
```
# define joint density
f = (1/(2*x2)) * exp(-(x2/4 + 2*x1/x2))
f
```

Out[46]:

$$\frac{e^{-\frac{2x_1}{x_2}-\frac{x_2}{4}}}{2x_2}$$

In [47]:
```
# integrate to get marginal density of x2

wrt = x1
lower = 0
upper = oo

f_x2 = integrate(f, (wrt, lower, upper), conds='none')
f_x2
```

Out[47]:

$$\frac{e^{-\frac{x_2}{4}}}{4}$$

(b)

In [48]:
```
# divide f by f_x2 to get conditional probability of x1|x2

f_x1_cond = simplify(f/f_x2)
f_x1_cond
```

Out[48]:

$$\frac{2e^{-\frac{2x_1}{x_2}}}{x_2}$$

(c)

In [49]:
```
# obtain g*(X1) by E(X1 | X2 = x2)

wrt = x1
lower = 0
upper = oo
```

```
g_star = integrate(x1*f_x1_cond, (wrt, lower, upper), conds='none')
g_star
```

Out[49]: $\frac{x_2}{2}$

(d)

In [76]:
```
# MSE

wrt = x1
lower = 0
upper = oo

integrate((x1 - x2/2)**2 * f_x1_cond, (wrt, lower, upper), conds='none')
```

Out[76]: $\frac{x_2^2}{4}$

Question 2

(a)

In [83]:
```
# characteristic polynomial

lam = Symbol('lambda')

Sigma = Matrix([[8, 3],[3, 9]])

char_poly = det(Sigma - lam*eye(2))
char_poly
```

Out[83]: $\lambda^2 - 17\lambda + 63$

(b)

In [117…
```
# eigenvalues

eigs = solve(char_poly, lam)
eigs
```

Out[117]:
```
[17/2 - sqrt(37)/2, sqrt(37)/2 + 17/2]
```

In [163…
```
# eigenvectors

Sigma.eigenvects()
```

Out[163]:
```
[(17/2 - sqrt(37)/2,
  1,
  [Matrix([
   [-sqrt(37)/6 - 1/6],
   [                1]])]),
 (sqrt(37)/2 + 17/2,
  1,
  [Matrix([
   [-1/6 + sqrt(37)/6],
   [                1]])])]
```

In [219…
```
# normalise eigenvector (17 - sqrt(37))/2
```

```
vect1 = Sigma.eigenvects()[0][2][0]
length = sqrt(vect1[0]**2 + vect1[1]**2)
vect1_norm = cancel(vect1/length)
vect1_norm
```

Out[219]:

$$\begin{bmatrix} \frac{-\sqrt{37}-1}{\sqrt{2\sqrt{37}+74}} \\ \frac{6}{\sqrt{2\sqrt{37}+74}} \end{bmatrix}$$

In [227…

```
vect1_norm.applyfunc(lambda x : round(x, 4))
```

Out[227]:

$$\begin{bmatrix} -0.763 \\ 0.6464 \end{bmatrix}$$

In [221…

```
# normalise eigenvector (17 + sqrt(37))/2

vect2 = Sigma.eigenvects()[1][2][0]
length = sqrt(vect2[0]**2 + vect2[1]**2)
vect2_norm = cancel(vect2/length)
vect2_norm
```

Out[221]:

$$\begin{bmatrix} \frac{-1+\sqrt{37}}{\sqrt{74-2\sqrt{37}}} \\ \frac{6}{\sqrt{74-2\sqrt{37}}} \end{bmatrix}$$

In [228…

```
vect2_norm.applyfunc(lambda x : round(x, 4))
```

Out[228]:

$$\begin{bmatrix} 0.6464 \\ 0.763 \end{bmatrix}$$

(c)

In [251…

```
# use eigenvalues and eigenvectors to compute X^(1/2)
eigval1 = Sigma.eigenvects()[0][0]
eigval2 = Sigma.eigenvects()[1][0]

(sqrt(eigval1)*vect1_norm*vect1_norm.T + sqrt(eigval2)*vect2_norm*vect2_norm.T).eva
```

Out[251]:

$$\begin{bmatrix} 2.77961 & 0.523229 \\ 0.523229 & 2.95402 \end{bmatrix}$$

(d)

In [254…

```
# use eigenvalues and eigenvectors to compute X^(1/2)

(eigval1**-0.5*vect1_norm*vect1_norm.T + eigval2**-0.5*vect2_norm*vect2_norm.T).eva
```

Out[254]:

$$\begin{bmatrix} 0.372171 & -0.0659206 \\ -0.0659206 & 0.350198 \end{bmatrix}$$

QUIZ 1 - QUESTION 3

Amanjit Gill

October 31, 2022

```
library("ellipse")
```

```
##
## Attaching package: 'ellipse'
## The following object is masked from 'package:graphics':
##
##     pairs
```

```
par(mar = c(4, 4, 0.5, 0.5))
dev.new(width = 2.5, height = 2.5, unit = "in")

n <- 12
p = 2
xbar <- cbind(c(-1.2, -3.9))
S <- rbind(c(6,1), c(1,6))
Sinv <- solve(S)

# PART A - confidence ellipsoid at level 90%

alpha <- 0.1
S.eig <- eigen(S)
scale <- p*(n-1)*qf(1-alpha, p, n-p)/(n*(n-p))

# major axis endpoints
(major_pt1 <- xbar - S.eig$vectors[,1] * sqrt(S.eig$values[1]*scale))
```

```
##           [,1]
## [1,] -2.569866
## [2,] -5.269866
```

```
(major_pt2 <- xbar + S.eig$vectors[,1] * sqrt(S.eig$values[1]*scale))
```

```
##            [,1]
## [1,]  0.1698658
## [2,] -2.5301342
```

```
# minor axis endpoints
(minor_pt1 <- xbar - S.eig$vectors[,2] * sqrt(S.eig$values[2]*scale))
```

```
##             [,1]
## [1,] -0.04225209
## [2,] -5.05774791
```

```
(minor_pt2 <- xbar + S.eig$vectors[,2] * sqrt(S.eig$values[2]*scale))
```

```
##           [,1]
## [1,] -2.357748
## [2,] -2.742252
```

```
# vector of x-coordinates for major axis
major_xvals <- c(major_pt1[1], major_pt2[1])

# vector of y-coordinates for major axis
major_yvals <- c(major_pt1[2], major_pt2[2])

# vector of x-coordinates for minor axis
minor_xvals <- c(minor_pt1[1], minor_pt2[1])

# vector of y-coordinates for minor axis
minor_yvals <- c(minor_pt1[2], minor_pt2[2])

# plot ellipse and major and minor axes
plot(ellipse(cov2cor(S), centre=c(-1.2, -3.9), t=qt(1-alpha/2, n-1), npoints=1000),
    pch=".", cex.lab=.8, cex.axis=.7, xlim=c(-8,6), ylim=c(-10,2),
    xlab=expression(mu*"1"), ylab=expression(mu*"2"))
points(c(major_xvals,minor_xvals), c(major_yvals,minor_yvals),
     pch=4, col="blue", lwd=2)
lines(major_xvals, major_yvals)
lines(minor_xvals, minor_yvals)
points(xbar[1], xbar[2], pch=16, col="blue")

# PART B - confidence ellipsoid for known variance

# everything stays the same except for the scale
scale <- qchisq(1-alpha, p)

# major axis endpoints
(major_pt1 <- xbar - S.eig$vectors[,1] * sqrt(S.eig$values[1]*scale))
```

```
##           [,1]
## [1,] -5.214735
## [2,] -7.914735
```

```
(major_pt2 <- xbar + S.eig$vectors[,1] * sqrt(S.eig$values[1]*scale))
```

```
##           [,1]
## [1,] 2.8147348
## [2,] 0.1147348
```

```
# minor axis endpoints
(minor_pt1 <- xbar - S.eig$vectors[,2] * sqrt(S.eig$values[2]*scale))
```

```
##          [,1]
## [1,]  2.19307
## [2,] -7.29307
```

```
(minor_pt2 <- xbar + S.eig$vectors[,2] * sqrt(S.eig$values[2]*scale))
```

```
##            [,1]
## [1,] -4.5930702
## [2,] -0.5069298
```

```r
# vector of x-coordinates for major axis
major_xvals <- c(major_pt1[1], major_pt2[1])

# vector of y-coordinates for major axis
major_yvals <- c(major_pt1[2], major_pt2[2])

# vector of x-coordinates for minor axis
minor_xvals <- c(minor_pt1[1], minor_pt2[1])

# vector of y-coordinates for minor axis
minor_yvals <- c(minor_pt1[2], minor_pt2[2])

# plot major and minor axes
lines(ellipse(cov2cor(S), centre=c(-1.2, -3.9),
    scale=c(sqrt(qchisq(1-alpha, p)),sqrt(qchisq(1-alpha, p))), npoints=1000),
    pch=".", cex=1, lty="dashed")
points(c(major_xvals,minor_xvals), c(major_yvals,minor_yvals), pch=4, lwd=1.5)
lines(major_xvals, major_yvals, lty="dashed")
lines(minor_xvals, minor_yvals, lty="dashed")
points(xbar[1], xbar[2], pch=16, col="blue")

# PART C - projection to get confidence interval

# change back to Part A scale
scale <- p*(n-1)*qf(1-alpha, p, n-p)/(n*(n-p))

# for variable x1
l_x1 <- c(1,0)
(x1_ci_low <- c(crossprod(l_x1, xbar)) - sqrt(scale * t(l_x1) %*% S %*% l_x1))
```

```
##           [,1]
## [1,] -2.993575
```

```r
(x1_ci_high <- c(crossprod(l_x1, xbar)) + sqrt(scale * t(l_x1) %*% S %*% l_x1))
```

```
##           [,1]
## [1,] 0.5935754
```

```r
# for variable x2
l_x2 <- c(0,1)
(x2_ci_low <- c(crossprod(l_x2, xbar)) - sqrt(scale * t(l_x2) %*% S %*% l_x2))
```

```
##           [,1]
## [1,] -5.693575
```

```r
(x2_ci_high <- c(crossprod(l_x2, xbar)) + sqrt(scale * t(l_x2) %*% S %*% l_x2))
```

```
##           [,1]
## [1,] -2.106425
```

```r
# plot x1 CI as vertical lines
# plot x2 CI as horizontal lines
abline(v=c(x1_ci_low, x1_ci_high), col="blue", lty="dashed")
abline(h=c(x2_ci_low, x2_ci_high), col="blue", lty="dashed")

# PART D - Bonferroni method to get confidence intervals
```

```r
# adapt Part C to do this efficiently

m <- 2*p

scale <- qt(1-alpha/m, n-1)

# for variable x1
l_x1 <- c(1,0)
(x1_ci_low <- c(crossprod(l_x1, xbar)) - scale*sqrt(t(l_x1) %*% S %*% l_x1 / n))
```

```
##           [,1]
## [1,] -2.756332
```

```r
(x1_ci_high <- c(crossprod(l_x1, xbar)) + scale*sqrt(t(l_x1) %*% S %*% l_x1 / n))
```

```
##           [,1]
## [1,] 0.3563315
```

```r
# for variable x2
l_x2 <- c(0,1)
(x2_ci_low <- c(crossprod(l_x2, xbar)) - scale*sqrt(t(l_x2) %*% S %*% l_x2 / n))
```

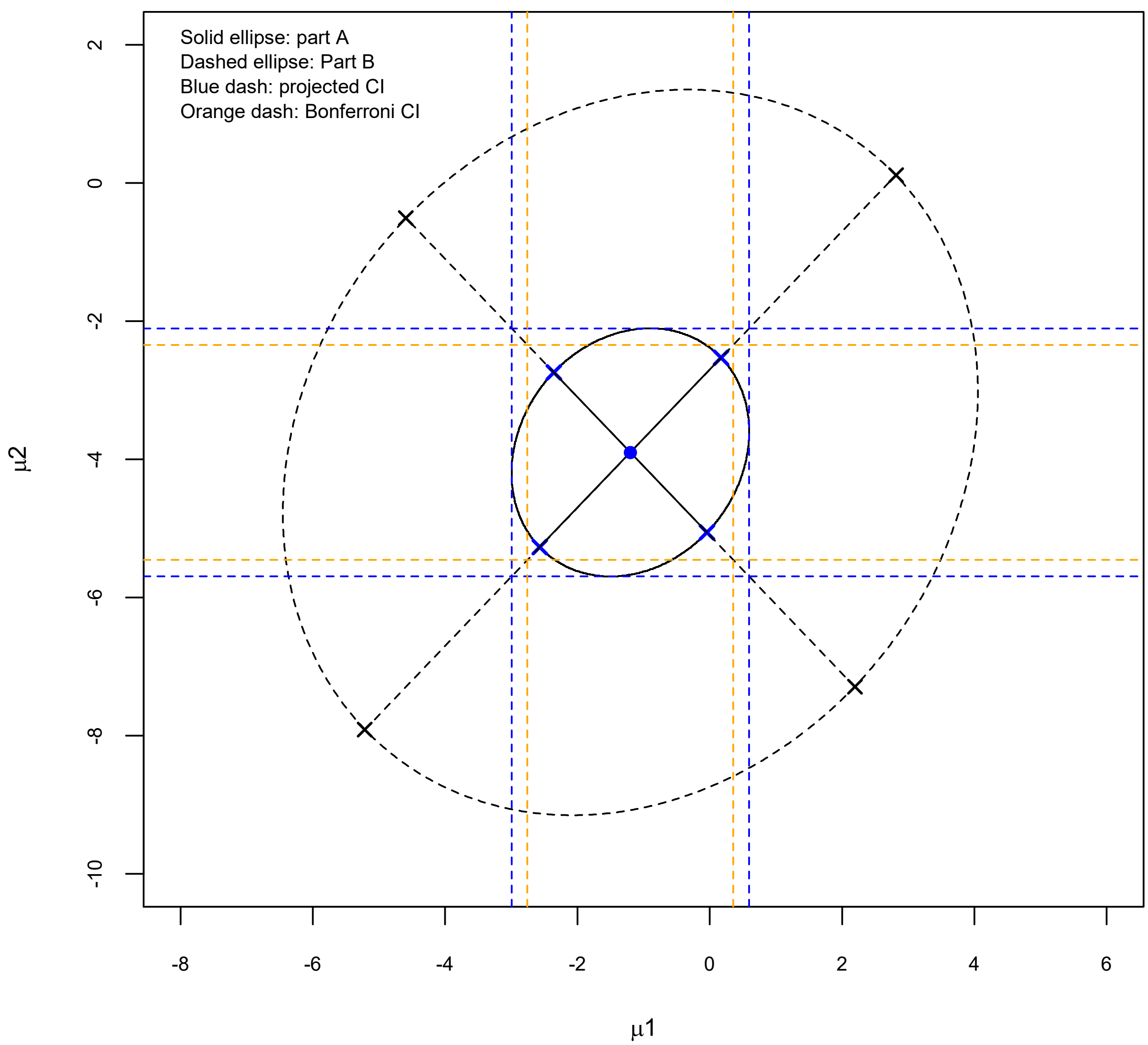
```
##           [,1]
## [1,] -5.456332
```

```r
(x2_ci_high <- c(crossprod(l_x2, xbar)) + scale*sqrt(t(l_x2) %*% S %*% l_x2 / n))
```

```
##           [,1]
## [1,] -2.343668
```

```r
# plot x1 CI as vertical lines
# plot x2 CI as horizontal lines
abline(v=c(x1_ci_low, x1_ci_high), col="orange", lty="dashed")
abline(h=c(x2_ci_low, x2_ci_high), col="orange", lty="dashed")

# add legend for all 4 parts
legend(
    x="topleft",
    legend=c("Solid ellipse: part A", "Dashed ellipse: Part B",
             "Blue dash: projected CI", "Orange dash: Bonferroni CI"),
    bty='n',
    cex=0.55
)
```


Solid ellipse: part A
Dashed ellipse: Part B
Blue dash: projected CI
Orange dash: Bonferroni CI
μ1
μ2

QUIZ 1 - QUESTION 4

Amanjit Gill

October 31, 2022

```r
library(GGally)
```

```
## Loading required package: ggplot2
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(MVN)

dev.new(width = 2.5, height = 2.5, unit = "in")

b <- as_tibble(read.csv("bridges.csv"))
head(b)
```

```
## # A tibble: 6 x 5
##   river erected purpose  length material
##   <chr>   <int> <chr>     <int> <chr>
## 1 A        1819 HIGHWAY    1037 WOOD
## 2 A        1837 HIGHWAY    1000 WOOD
## 3 A        1840 HIGHWAY     990 WOOD
## 4 A        1844 AQUEDUCT   1000 IRON
## 5 M        1846 HIGHWAY    1500 IRON
## 6 A        1851 HIGHWAY    1000 WOOD
```

```r
# PART A - pairwise plots and normality test

b.subset <- b %>% select(erected, length)
b.subset
```

```
## # A tibble: 81 x 2
##    erected length
##      <int>  <int>
##  1    1819   1037
```

```
##  2    1837   1000
##  3    1840    990
##  4    1844   1000
##  5    1846   1500
##  6    1851   1000
##  7    1856   1200
##  8    1859   1030
##  9    1863   1000
## 10    1864   1200
## # ... with 71 more rows
```

```r
ggpairs(b.subset)

mvn(b.subset, mvnTest = "mardia", univariateTest = "SW")
```

```
## $multivariateNormality
##              Test        Statistic              p value Result
## 1 Mardia Skewness 44.4848070981332 5.08777517344877e-09     NO
## 2 Mardia Kurtosis 3.78231712161238 0.000155375239301536     NO
## 3             MVN             <NA>                 <NA>     NO
##
## $univariateNormality
##           Test  Variable Statistic   p value Normality
## 1 Shapiro-Wilk  erected     0.9832  0.3674      YES
## 2 Shapiro-Wilk  length      0.8188  <0.001      NO
##
## $Descriptives
##          n     Mean   Std.Dev Median  Min  Max 25th 75th       Skew   Kurtosis
## erected 81 1910.222  34.76852   1910 1819 1978 1890 1931 -0.3205326 -0.3713876
## length  81 1567.469 747.49152   1300  804 4558 1000 2000  1.6723769  3.1325645
```

```r
# PART B - Hypothesis Test

alpha <- 0.05

# separate into separate sets by river (M or A)
# river 1 = M, river 2 = A
x1 <- b[b$river == 'M', ] %>% select(erected, length)
x2 <- b[b$river == 'A', ] %>% select(erected, length)

p <- ncol(x1)

(n1 <- nrow(x1))
```

```
## [1] 32
```

```r
(n2 <- nrow(x2))
```

```
## [1] 37
```

```r
xbar1 <- colMeans(x1)
S1 <- cov(x1)

xbar2 <- colMeans(x2)
S2 <- cov(x2)

(Sp <- ((n1 - 1)*S1 + (n2 - 1)*S2)/(n1+n2-2))
```

```
##           erected     length
## erected 1240.460   4091.154
## length  4091.154 411748.065
```

```r
# approach 1: compare T^2 with critical T^2

(Tsq <- t(xbar1-xbar2) %*% solve(Sp*(1/n1 + 1/n2)) %*% (xbar1-xbar2))
```

```
##          [,1]
## [1,] 1.196022
```

```r
(Tsq_crit <- qf(1-alpha, p, n1+n2-p-1) * p*(n1+n2-2)/(n1+n2-p-1))
```

```
## [1] 6.366864
```

```r
# approach 2: compare F with critical F

(F <- Tsq * (n1+n2-p-1)/(p*(n1+n2-2)))
```

```
##           [,1]
## [1,] 0.5890852
```

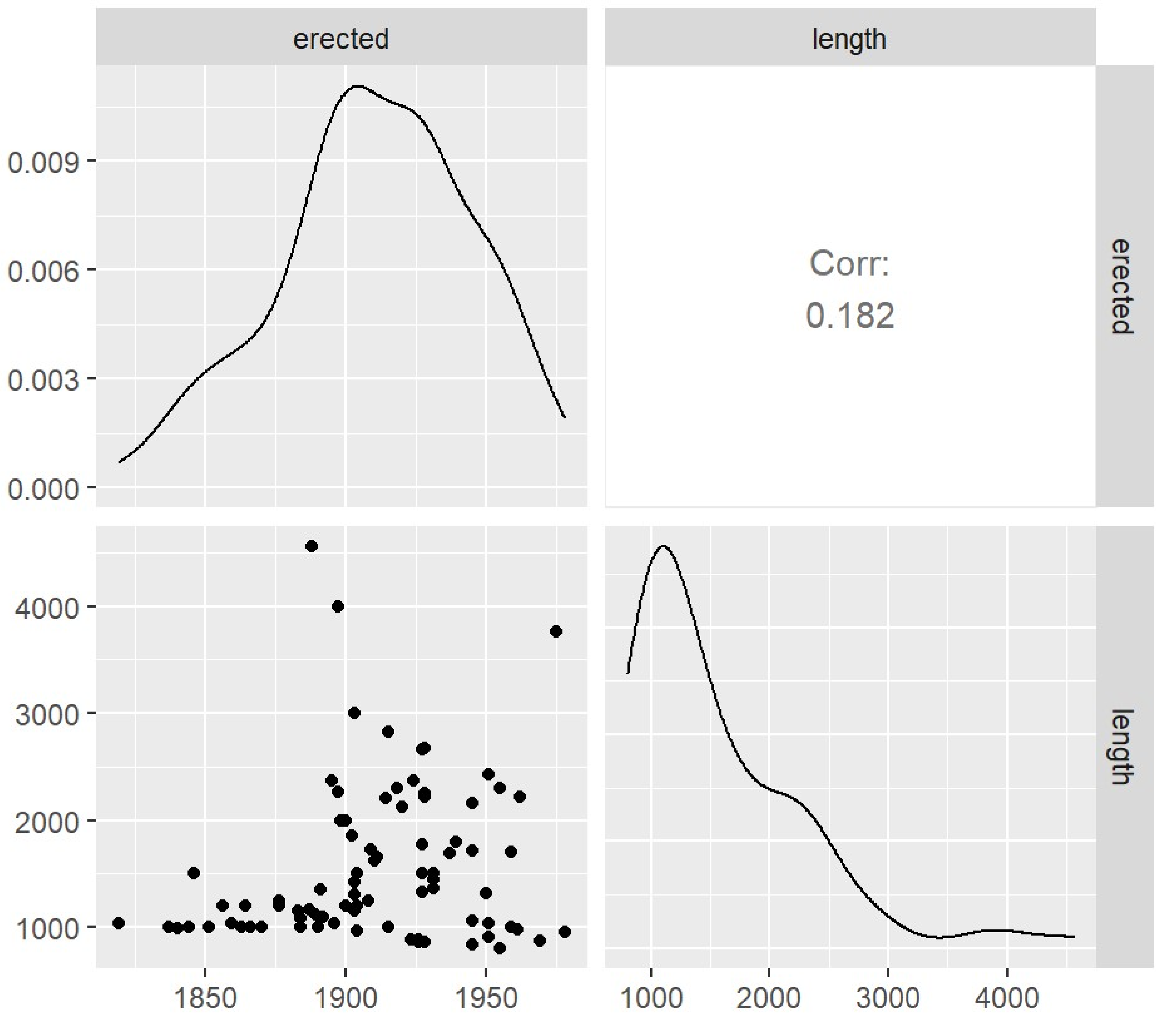
```r
(F_crit <- qf(1-alpha, p, n1+n2-p-1))
```

```
## [1] 3.135918
```

```r
# approach 3: compare p-value with critical p (0.05)

(prob_of_F <- pf(F, p, n1+n2-p-1, lower.tail = FALSE))
```

```
##           [,1]
## [1,] 0.5577251
```



erected
length
Corr:
0.182
erected
length
0.009
0.006
0.003
0.000
4000
3000
2000
1000
1850
1900
1950
1000
2000
3000
4000