MAST30025: Linear Statistical Models

Week 3 Lab

1. Let X be a 10×5 matrix of full rank and let $H = X(X^TX)^{-1}X^T$.
   Find $tr(H)$ and $r(H)$.
   Solution: $tr(H) = tr(X(X^TX)^{-1}X^T) = tr(X^TX(X^TX)^{-1}) = tr(I_5) = 5$. Since H is symmetric and idempotent, $r(H) = tr(H) = 5$.

2. Let
$$A = \begin{bmatrix} 3 & 1 & 8 \\ 1 & 0 & 1 \\ 2 & 1 & -4 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$
   Let $z = \mathbf{y}^TA\mathbf{y}$. Write out z in full, then find $\frac{\partial z}{\partial \mathbf{y}}$ directly *and* using the matrix formula.
   Solution: $z = 3y_1^2 - 4y_3^2 + 2y_1y_2 + 10y_1y_3 + 2y_2y_3$.
$$\frac{\partial z}{\partial \mathbf{y}} = A\mathbf{y} + A^T\mathbf{y} = \begin{bmatrix} 6y_1 + 2y_2 + 10y_3 \\ 2y_1 + 2y_3 \\ 10y_1 + 2y_2 - 8y_3 \end{bmatrix}.$$

3. Let $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T$ be a random vector with mean $\boldsymbol{\mu} = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}^T$, and assume that $\mathrm{Var}\, y_i = 4$ and $\mathrm{Cov}\,(y_i, y_j) = 0$.

   (a) Write down $\mathrm{Var}\,\mathbf{y}$.
   Solution:
$$\mathrm{Var}\,\mathbf{y} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

   (b) Let
$$A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix}$$
   and find $\mathrm{Var}\, A\mathbf{y}$ and $\mathbb{E}[\mathbf{y}^TA\mathbf{y}]$.
   Solution:
$$\begin{aligned} \mathrm{Var}\, A\mathbf{y} &= AVA^T = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ -3 & 2 & 6 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 8 & 4 & -4 \\ -12 & 8 & 24 \\ 4 & 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 56 & -16 & -76 \\ -16 & 20 & 44 \\ -76 & 44 & 152 \end{bmatrix}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\mathbf{y}^TA\mathbf{y}] &= tr(AV) + \boldsymbol{\mu}^TA\boldsymbol{\mu} \\ &= tr\left(\begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ 1 & 2 & 0 \\ -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \\ &= tr\left(\begin{bmatrix} 8 & -12 & 4 \\ 4 & 8 & 0 \\ -4 & 24 & 4 \end{bmatrix} \right) + \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} -5 \\ 7 \\ 19 \end{bmatrix} \\ &= 20 + 54 = 74. \end{aligned}$$

4. Prove corollaries 3.6 and 3.7 from the lectures.

   Solution:

   (a) Just observe that if $\mu = \mathbf{0}$ then $\lambda = \frac{1}{2}\mu^T A \mu = 0$.

   (b) From the distribution of $\mathbf{y}$, we know that $\frac{1}{\sigma}\mathbf{y} \sim MVN(\frac{1}{\sigma}\boldsymbol{\mu}, I)$. Therefore $(\frac{1}{\sigma}\mathbf{y})^T A(\frac{1}{\sigma}\mathbf{y}) = \frac{1}{\sigma^2}\mathbf{y}^T A\mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter

   $$\lambda = \frac{1}{2}(\frac{1}{\sigma}\boldsymbol{\mu})^T A(\frac{1}{\sigma}\boldsymbol{\mu}) = \frac{1}{2\sigma^2}\boldsymbol{\mu}^T A\boldsymbol{\mu}$$

   if and only if A is idempotent and has rank k.

5. Let $\mathbf{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^T$ be a normal random vector with mean and variance

   $$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

   Let

   $$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

   (a) Find the distributions of $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$.

   Solution: A and B are both idempotent and have rank 1, so $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ have noncentral χ^2 distributions with 1 degree of freedom each and noncentrality parameters

   $$\frac{1}{2}\begin{bmatrix} 2 & 4 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = 9$$

   and

   $$\frac{1}{2}\begin{bmatrix} 2 & 4 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = 1$$

   respectively.

   (b) Are $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ independent?

   Solution: $AB = 0$, so they are independent.

   (c) What is the distribution of $\mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y}$?

   Solution: $\mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{y}$ has a noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter 10.

6. Let $y_1, \ldots, y_n$ be an i.i.d. normal sample. Show that

   $$\bar{y} = \tfrac{1}{n}\sum_{i=1}^{n} y_i \text{ and } s^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

   are independent. (*Hint*: Express them as a random "vector" and quadratic form respectively.)

   Solution: We can do this directly or via our general theory for linear models. To do it directly, let $\mathbf{1}$ be the vector made up entirely of 1's, then

   $$\begin{aligned} \bar{y} &= \tfrac{1}{n}\mathbf{1}^T\mathbf{y} \\ \mathbf{y} - \bar{y}\mathbf{1} &= (I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y} \\ s^2 &= \tfrac{1}{n-1}[(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y}]^T(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y} \\ &= \tfrac{1}{n-1}\mathbf{y}^T(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)^T(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y} \\ &= \tfrac{1}{n-1}\mathbf{y}^T(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y} \end{aligned}$$

   noting that $I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is symmetric and idempotent. It is now easy to check that $B = \frac{1}{n}\mathbf{1}^T$ and $A = \frac{1}{n-1}(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$ satisfy $BVA = 0$, where $V = \sigma^2 I = \operatorname{Var}\mathbf{y}$, whence $\bar{y} = B\mathbf{y}$ and $s^2 = \mathbf{y}^T A\mathbf{y}$ are independent.

   The alternative approach is to notice that $\bar{y}$ and s^2 are just the usual estimates of $\boldsymbol{\beta}$ and σ^2 for the linear model $\mathbf{y} = \mathbf{1}\boldsymbol{\beta} + \varepsilon$, with $\boldsymbol{\beta} = \mu = \mathbb{E}y_i$ and $\operatorname{Var}\varepsilon = \sigma^2 I$. (This is sometimes called the *null* model.)

R exercises

The following are taken from Chapter 3 of spuRs (Introduction to Scientific Programming and Simulation Using R).

1. Consider the function $y = f(x)$ defined by

x	≤ 0	$\in (0, 1]$	> 1
$f(x)$	$-x^3$	x^2	$\sqrt{x}$

   Supposing that you are given x, write an R expression for y using `if` statements.

   Add your expression for y to the following program, then run it to plot the function f.

```
# input
x.values <- seq(-2, 2, by = 0.1)

# for each x calculate y
n <- length(x.values)
y.values <- rep(0, n)
for (i in 1:n) {
    x <- x.values[i]
    # your expression for y goes here
    y.values[i] <- y
}

# output
plot(x.values, y.values, type = "l")
```

   Do you think f has a derivative at 1? What about at 0?

   Solution: replace the comment `# your expression for y goes here` with the following lines

```
  if (x <= 0) {
    y <- -x^3
  } else if (x <= 1) {
    y <- x^2
  } else {
    y <- sqrt(x)
  }
```

   The function clearly has no derivative at 1, but it does have a derivative (equal to 0) at 0. It does not have a second derivative at 0 however.

2. Let $h(x, n) = 1 + x + x^2 + \cdots + x^n = \sum_{i=0}^{n} x^i$. Write an R program to calculate $h(x, n)$ using a `for` loop.

   Solution:

```
> x <- 0.8
> n <- 10
> # using a for loop
> h <- 0
> for (i in 0:n) h <- h + x^i
> print(h)

[1] 4.570503
```

3. The function $h(x, n)$ from Exercise 2 is the finite sum of a geometric sequence. It has the following explicit formula, for $x \neq 1$,

$$h(x, n) = \frac{1 - x^{n+1}}{1 - x}.$$

Test your program from Exercise 2 against this formula using the following values

x	n	$h(x,n)$
0.3	55	1.428571
6.6	8	4243335.538178

You should use the computer to calculate the formula rather than doing it yourself.

Solution:

```
> # using the formula
> if (x == 1) {
+   h <- n + 1
+ } else {
+   h <- (1 - x^(n+1))/(1 - x)
+ }
> print(h)

[1] 4.570503
```

4. First write a program that achieves the same result as in Exercise 2 but using a `while` loop. Then write a program that does this using vector operations (and no loops).

   If it doesn't already, make sure your program works for the case $x = 1$.

   Solution:

```
> # using a while loop
> h <- 0
> i <- 0
> while (i <= n) {
+   h <- h + x^i
+   i <- i + 1
+ }
> print(h)

[1] 4.570503

> # vectorised
> (h <- sum(x^(0:n)))

[1] 4.570503
```

5. To rotate a vector $(x, y)^T$ anticlockwise by θ radians, you premultiply it by the matrix

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

   Write a program in R that does this for you.

   Solution:

```
> x.old <- matrix(c(1, 0), 2, 1)
> theta <- pi/4
> A <- matrix(c(cos(theta), sin(theta), -sin(theta), cos(theta)), 2, 2)
> (x.new <- A %*% x.old)

          [,1]
[1,] 0.7071068
[2,] 0.7071068
```

6. Given a vector `x`, calculate its geometric mean using both a for loop and vector operations. (The geometric mean of $x_1, \ldots, x_n$ is $\left(\prod_{i=1}^n x_i\right)^{1/n}$.)

   You might also like to have a go at calculating the harmonic mean, $\left(\sum_{i=1}^n 1/x_i\right)^{-1}$, and then check that if the x_i are all positive, the harmonic mean is always less than or equal to the geometric mean, which is always less than or equal to the arithmetic mean.

   Solution:

```
> x <- c(2, 1, 3, 5, 1, 3)
> # geometric mean using a loop
> mg <- 1
> for (xi in x) mg <- mg*xi
> (mg <- mg^(1/length(x)))

[1] 2.116933

> # geometric mean vectorised
> (mg <- prod(x)^(1/length(x)))

[1] 2.116933

> # harmonic mean
> (hm <- 1/sum(1/x))

[1] 0.2970297

> # arithmetic mean
> mean(x)

[1] 2.5
```

7. A room contains 100 toggle switches, originally all turned off. 100 people enter the room in turn. The first toggles every switch, the second toggles every second switch, the third every third switch, and so on, to the last person who toggles the last switch only.

   At the end of this process, which switches are turned on?

   Solution:

```
> x <- rep(FALSE, 100)          # the state of the switches
> for (i in 1:100) {            # we go once round the loop for each person
+   for (j in seq(i, 100, i)) { # the switches toggled by person i are the j's
+     x[j] <- !x[j]             # toggles the switch
+   }
+ }
> which(x)

 [1]   1   4   9  16  25  36  49  64  81 100
```

   You might speculate (correctly) that if you have n switches and n people, then the switches left on would correspond to the square numbers. The proof depends on the number of distinct factors of a number.