# Analisis de los datos de migraciones de la mosca de las frutas

Anna Sikov

UNI

June 15, 2020

# Datos

El archivo medfly.txt contiene datos coletados para investigar las migraciónes de la mosca mediterránea de la fruta (la descripción completa de los datos está en el archivo Medfly descripcion.pdf), donde la pregunta de investigación es si la mosca mediterránea de la fruta pasa el invierno en regiones más frias de Israel o migran a lugares más calientes y regresan cuando pasa el invienro.

Ajustar un modelo lineal para predecir el número de moscas atrapadas (A), dependiendo de la localización de la trampa.

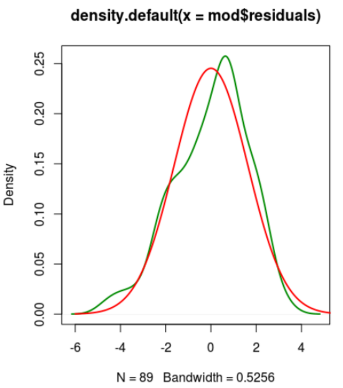
1. Definir el modelo lineal para responder a la pregunta de la investigación.
2. Estimar los coeficientes de la regresión.
3. Calcular intervalos de confianza de 95% para los coeficientes de la regresión.
4. Calcular el $R^2$ del modelo.

Para este ejercício ustedes tienen que hacer el cálculo en el software R, sin utilizar las funcciónes como lm, glm, etc. Para esto ustedes tienen que definir todas las matrices y vectores relevantes y utilizar las fórmulas que ustedes estudiaron en el curso de modelos lineales.

Cómo responderían a la pregunta de la investigación?

# Modelo de predicción del número de moscas atrapadas

```
> summary(lm(log(1+medfly$A)~medfly$south+medfly$alt+medfly$dist+medfly$loc_host+W1+W2+
+            medfly$alt*medfly$south
+ ))

Call:
lm(formula = log(1 + medfly$A) ~ medfly$south + medfly$alt +
    medfly$dist + medfly$loc_host + W1 + W2 + medfly$alt * medfly$south)

Residuals:
   Min     1Q Median     3Q    Max
-3.982 -1.132  0.202  1.332  3.102

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.4170963  0.8414955   4.061 0.000112 ***
medfly$south            -2.3028513  0.8739786  -2.635 0.010081 *
medfly$alt              -0.0007046  0.0012562  -0.561 0.576407
medfly$dist             -0.0604571  0.0205610  -2.940 0.004271 **
medfly$loc_host          0.8506964  0.4651676   1.829 0.071111 .
W1                       2.2879518  0.7160395   3.195 0.001992 **
W2                       2.5481120  0.5555326   4.587 1.62e-05 ***
medfly$south:medfly$alt  0.0043576  0.0016674   2.613 0.010682 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.628 on 81 degrees of freedom
Multiple R-squared:  0.4964,    Adjusted R-squared:  0.4528
F-statistic:  11.4 on 7 and 81 DF,  p-value: 5.559e-10

>
```



# Ejercício

Cuál sería su predicción del número de moscas atrapadas en una trampa que esta en la parte central, en la altura de 200 metros sobre el nivel del mar, 25 km de la región caliente, y donde no se encuentra ningun huesped dentro de 50 metros de la trampa, y la primera captura ocurrió en la semana 17?

# Modelo de Regresión Lineal

$$Y_i = X_i\beta + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma^2),$$

donde $X_i = (1\ x_{i1}\ x_{i2}\ ...\ x_{ip})$, $\beta = (\beta_0, ..., \beta_p)^t$ $i = 1, ..., n$, donde $n$ es el número de observaciones y $p$ es el número de variables explicativas; los residuos $\epsilon_i$ son independientes.
Otra forma para presentar un modelo de regresión lineal es:

$$Y = X\beta + \epsilon,$$

donde $Y = (Y_1, ..., Y_n)^t$ es el vector de las variables dependientes, $\epsilon = (\epsilon_1, ..., \epsilon_n)^t$ es el vector de los residuos,

$$X = \begin{pmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & x_{21} & ... & x_{2p} \\ 1 & ... & ... & ... \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix}$$

# Modelo de Regresión Lineal

$$Y = X\beta + \epsilon$$

Utilizando el método de máxima verosimilitud o el método de mínimos cuadrados obtenemos:

$\hat{\beta} = (X^t X)^{-1} X^t Y$

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{j=1}^{n} (Y_j - X_j \hat{\beta})^2$$

- El estimador $\hat{\beta}$ es insesgado (Ejercício)
- $\hat{\beta} \sim N(\beta, (X^t X)^{-1} \sigma^2)$ (Ejercício)

Además, se puede calcular el $R^2$ del modelo que es el porcentaje de la varianza de $Y$, explicada por el modelo:

$$R^2 = \frac{\sum_{j=1}^{n} (X_j \hat{\beta} - \bar{Y})^2}{\sum_{j=1}^{n} (Y_j - \bar{Y})^2} = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \bar{Y})^2}{\sum_{j=1}^{n} (Y_j - \bar{Y})^2}$$

# Modelo de Regresión Lineal

Console Terminal Jobs

/cloud/project/

```
> X=cbind(1,medfly$south, medfly$alt,medfly$dist,medfly$loc_host,W1,W2,medfly$south*medfly$alt)
> X[1:8,]
                          W1 W2
[1,] 1 1 726.6240 29.5 1  0  1 726.6240
[2,] 1 1 240.7184 30.0 1  0  1 240.7184
[3,] 1 1 240.4684 30.0 1  0  1 240.4684
[4,] 1 1 244.8809 30.0 0  0  1 244.8809
[5,] 1 1 246.4829 30.0 0  1  0 246.4829
[6,] 1 1 232.6738 29.0 0  0  0 232.6738
[7,] 1 1 240.7439 29.0 0  0  0 240.7439
[8,] 1 1 240.6470 29.0 0  0  1 240.6470
> Y=log(1+medfly$A)
> (solve(t(X)%*%X))%*%t(X)%*%Y->est.beta
> est.beta
            [,1]
    3.4170963162
   -2.3028513346
   -0.0007046326
   -0.0604570894
    0.8506963844
W1  2.2879518231
W2  2.5481119898
    0.0043575780
>
```

Los valores de los coeficientes son iguales a aquellos que fueron obtenidos usando la función "lm" de R

# Modelo de Regresión Lineal

```
> Y-X%*%est.beta->residuos
> sqrt(sum(residuos^2)/(89-8))->sigma
> sigma
[1] 1.627858
> (solve(t(X)%*%X))*sigma^2
                                                                                        W1            W2
     0.7081147584 -0.3867858139 -4.248317e-04 -4.666420e-03 -8.788334e-02 -4.386517e-01 -2.277801e-01  5.023519e-04
    -0.3867858139  0.7638386508  4.040097e-04  9.501461e-04  1.650894e-01  6.344384e-02 -9.900413e-02 -1.258803e-03
    -0.0004248317  0.0004040097  1.578119e-06 -1.181270e-05 -3.325017e-05  1.770234e-04 -3.031147e-05 -9.903871e-07
    -0.0046664195  0.0009501461 -1.181270e-05  4.227534e-04 -1.180526e-03  5.096833e-03  2.533051e-03 -3.170968e-06
    -0.0878833366  0.1650894436 -3.325017e-05 -1.180526e-03  2.163809e-01 -1.053748e-01 -7.333121e-02 -1.894900e-04
W1  -0.4386517470  0.0634438395  1.770234e-04  5.096833e-03 -1.053748e-01  5.127126e-01  2.823527e-01 -3.819651e-05
W2  -0.2277801076 -0.0990041260 -3.031147e-05  2.533051e-03 -7.333121e-02  2.823527e-01  3.086164e-01  1.993740e-04
     0.0005023519 -0.0012588034 -9.903871e-07 -3.170968e-06 -1.894900e-04 -3.819651e-05  1.993740e-04  2.780059e-06
> diag((solve(t(X)%*%X))*sigma^2)->D
> D
                                                                            W1           W2
7.081148e-01 7.638387e-01 1.578119e-06 4.227534e-04 2.163809e-01 5.127126e-01 3.086164e-01 2.780059e-06
> sqrt(D)
                                                                      W1          W2
0.841495549 0.873978633 0.001256232 0.020560967 0.465167649 0.716039496 0.555532577 0.001667351
> |
```

Los valores de las desviaciones estándar coeficientes son iguales a aquellas que fueran obtenidas usando la función "lm" de R Ejercicio: Calcular $R^2$, utilizando la fórmula, el vector $Y$, la mztriz $X$ y $\hat{\beta}$

# Modelo de Regresión Lineal: Intervalos de confianza

Utilizamos el resultado:

$$\hat{\beta} \sim N(\beta, (X^t X)^{-1}\sigma^2)$$

Entonces para un coeficiente específico $\beta_k$ obtenemos:

$$\hat{\beta}_k \sim N(\beta_k, \{(X^t X)^{-1}\sigma^2\}_{kk})$$

y

$$\frac{(\hat{\beta}_k - \beta_k)}{\{(X^t X)^{-1}\sigma^2\}_{kk}} \sim N(0,1)$$

Intervalo de confianza para $\beta_k$ del nivel $1-\alpha$

$$-z_{1-\frac{\alpha}{2}} \leq \frac{(\hat{\beta}_k - \beta_k)}{\{(X^t X)^{-1}\sigma^2\}_{kk}} \leq z_{1-\frac{\alpha}{2}}$$

# Modelo de Regresión Lineal: Intervalos de confianza

Intervalo de confianza para $\beta_k$ del nivel $1 - \alpha$

$$\hat{\beta}_k - z_{1-\frac{\alpha}{2}}\{(X^tX)^{-1}\sigma^2\}_{kk} \le \beta \le \hat{\beta}_k + z_{1-\frac{\alpha}{2}}\{(X^tX)^{-1}\sigma^2\}_{kk}$$

Puesto que el valor de $\sigma^2$ es desconocido, tenemos que sustituirlo por el estimados $\hat{\sigma}^2$ y utilizar la distribución $t_{(n-p)}$.

Entotnces, el intervalo de confianza es:

$$\hat{\beta}_k - t_{n-p}^{1-\frac{\alpha}{2}}\{(X^tX)^{-1}\hat{\sigma}^2\}_{kk} \le \beta \le \hat{\beta}_k + t_{n-p}^{1-\frac{\alpha}{2}}\{(X^tX)^{-1}\hat{\sigma}^2\}_{kk}$$

Por ejemplo, para el coeficiente de la variable $W_1$ el intervalo de confianza del nivel 0.95 es:

$$(2.29 - 1.99 * 0.716, 2.29 - 1.99 * 0.716) = (0.865, 3.715)$$

Ejercicio: calcular un intervalo de confianza del nivel 0.90 para el coeficiente de la variable loc host

# Modelo de Regresión Lineal: Prueba de Hipotesis

$$H_0 : \beta_k = 0$$

Utilizamos el mismo resultado:

$$\hat{\beta} \sim N(\beta, (X^t X)^{-1}\sigma^2)$$

Entonces, rechasamos $H_0$ si
$\frac{(\hat{\beta}_k - \beta_k)}{\{(X^t X)^{-1}\hat{\sigma}^2\}_{kk}} \leq -t_{n-p}^{1-\frac{\alpha}{2}}$ o $\frac{(\hat{\beta}_k - \beta_k)}{\{(X^t X)^{-1}\hat{\sigma}^2\}_{kk}} \geq t_{n-p}^{1-\frac{\alpha}{2}}$
Para $\alpha = 0.05$ los valores críticos son: -1.99 y 1.99
Por ejemplo para probar $H_0 : \beta_k = 0$ para el coeficiente de la variable $W_1$, el valor de estadística $\frac{(\hat{\beta}_k - \beta_k)}{\{(X^t X)^{-1}\hat{\sigma}^2\}_{kk}}$ es $\frac{2.29 - 0}{0.716} = 3.198$, y en este caso rechasamos $H_0$.
$P - valor = 2(1 - P(T_{(81)>3.198})) = 0.001973$
Ejercicio: Calcular el p-valor del coeficiente de la variable loc host.

# Modelo de Regresión Lineal

$$Y_i = X_i\beta + \epsilon_i,\ \ \epsilon_i \sim N(0, V)$$

Las varianzas de los errores no son iguales, además los errores no son necesariamente independientes.

En este caso:

- $\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y$
- $\hat{\beta} \sim N(\beta, (X^t V^{-1} X)^{-1})$

# Modelo de Regresión Lineal

$$Y_i = X_i\beta + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma^2)$$

Suponemos que:

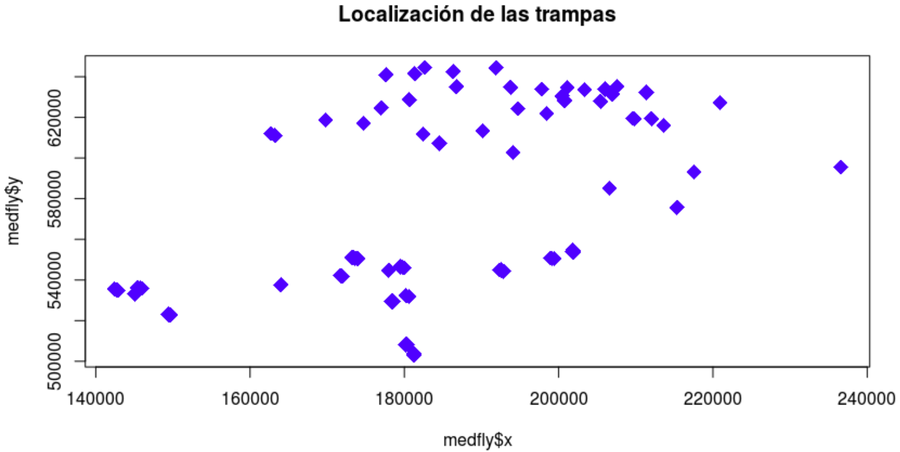
- los valores observados en i-ésimo y j-ésimo sitios son correlacionados.
- los errores $\epsilon_i$, $i = 1, ..., n$ son independientes.

Que ocurre si el valor observado en un sitio específico está correlacionado con el valor del sitio vecino?

# Datos Espaciales

- los valores observados en i-ésimo y j-ésimo sitios son correlacionados.
- los errores $\epsilon_i$, $i = 1, ..., n$ son independientes.

Que ocurre si el valor observado en un sitio específico está correlacionado con el valor del sitio vecino?


Localización de las trampas
medfly$y
500000
540000
580000
620000
140000
160000
180000
200000
220000
240000
medfly$x

# Dependencia Espacial

$$
\begin{aligned}
y_i =& \alpha_j y_j + X_i\beta + \epsilon_i \\
y_j =& \alpha_i y_i + X_j\beta + \epsilon_j \\
\epsilon_i \sim& N(0, \sigma^2), \quad i = 1 \\
\epsilon_j \sim& N(0, \sigma^2), \quad j = 2
\end{aligned}
$$

- Aquí suponemos que las observaciones en dos sitios vecinos están correlacionadas.
- El proceso que genera las observaciones, las genera de forma simultanea.

# Dependencia Espacial

$$y_1 = \alpha_{12} y_2 + \alpha_{13} y_3 + ... + \alpha_{1n} y_n + X_1 \beta + \epsilon_1$$

$$y_2 = \alpha_{21} y_1 + \alpha_{23} y_3 + ... + \alpha_{2n} y_n + X_2 \beta + \epsilon_2$$

$$\vdots$$

$$\vdots$$

$$y_n = \alpha_{n1} y_1 + \alpha_{n2} y_2 + ... + \alpha_{n,n-1} y_n + X_n \beta + \epsilon_n$$

- Aquí tenemos una variable explicativa adicional que es una combinación lineal de los valores en los sitios vecinos.
- Los errores $\epsilon_1, ..., \epsilon_n$ son independientes.

# Modelo de regresión espacial

Sean $A(x_i, y_i)$ y $B(x_j, y_j)$ dos puntos en un espacio bidimensional. Entonces la distancia entre $A$ y $B$ es definida como

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Definamos $W_{ij} = 1/d_{ij}$, $i, j = 1, ..., n$ $si$ $i \neq j$; $W_{ii} = 0$
Entonces,

$$y_i = \rho \sum_{j=1}^{n} W_{ij} y_j + X_i \beta + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n$$

# Modelo de regresión espacial

$$W = \begin{pmatrix} 0 & W_{12} & ... & W_{1n} \\ W_{21} & 0 & ... & W_{2n} \\ ... & ... & ... & ... \\ W_{n1} & W_{n2} & ... & 0 \end{pmatrix}$$

Ahora, sea $\tilde{y} = Wy$, $\tilde{\beta} = (\rho, \beta_0, \beta_1, ..., \beta_p)^t$, $\tilde{X} = [\tilde{y} \ X]$

$$\tilde{X} = \begin{pmatrix} \tilde{y}_1 & 1 & ... & X_{1p} \\ \tilde{y}_2 & 1 & ... & X_{2p} \\ ... & ... & ... & ... \\ \tilde{y}_n & 1 & ... & X_{np} \end{pmatrix}$$

Entonces, el modelo puede ser escrito de la siguiente forma:

$$\begin{aligned} \mathbf{y} =& \rho \mathbf{W} \mathbf{y} + \mathbf{X}\beta + \epsilon \\ \epsilon \sim& N(0, \sigma^2 \mathbf{I}_n) \end{aligned}$$

# Modelo de regresión espacial

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X}\beta + \epsilon$$
$$\epsilon \sim N(0, \sigma^2 \mathbf{I}_n)$$

- $\rho$ es el coeficiente de autoregressión espacial.
  - Si $\rho = 0$, no existe autocorrelacion espacial. Información sobre el valor observado en un punto específico no proporciona información sobre otros sitios (independencia espacial).
  - Si $\rho > 0$, autocorrelación espacial positiva. Valores observados en sitios vecinos tienen una tendencia de ser similares (agrupamiento.)
  - Si $\rho < 0$, autocorrelación espacial negativa. Valores observados en sitios vecinos tienen una tendencia de ser diferentes (segregación.)

# Cálculo de la matriz W

```r
W=matrix(0,89,89)
for (i in 1:88)
{
        for (j in (i+1):89)
        {
                W[i,j]=sqrt((medfly$x[i]-medfly$x[j])^2+(medfly$y[i]-medfly$y[j])^2)
                W[j,i]=W[i,j]
        }
}
```

```
> W[1:8,1:8]
         [,1]        [,2]        [,3]        [,4]        [,5]        [,6]        [,7]        [,8]
[1,]      0.0 118757.3852 118929.8468 118982.4102 119026.9316 116586.0564 116529.7207 116544.3590
[2,] 118757.4      0.0000    534.3489    644.4243    762.6448   3057.3033   3428.8124   3546.7483
[3,] 118929.8    534.3489      0.0000    111.2654    229.1929   2867.9320   3194.0897   3294.7692
[4,] 118982.4    644.4243    111.2654      0.0000    118.2550   2857.2378   3171.6618   3267.9388
[5,] 119026.9    762.6448    229.1929    118.2550      0.0000   2838.0073   3139.5819   3231.1095
[6,] 116586.1   3057.3033   2867.9320   2857.2378   2838.0073      0.0000    454.1457    615.4626
[7,] 116529.7   3428.8124   3194.0897   3171.6618   3139.5819    454.1457      0.0000    163.9976
[8,] 116544.4   3546.7483   3294.7692   3267.9388   3231.1095    615.4626    163.9976      0.0000
>
```

# Correlación Espacial: Intuición

```
> range(W[57,c(1:56,58:89)])
[1]   172.289 95203.150
>
> which(W[57,]>90000)
[1] 17 18 19 20 27
> which(W[57,]<300)
[1] 53 54 55 57
>
> medfly$A[57]
[1] 97
>
> medfly$A[53:55]
[1] 15 23 13
>
> medfly$A[c(17:20,27)]
[1] 1081  611 2684   70  497
```

```
> range(W[3,c(1,2,4:89)])
[1]    111.2654 120706.5498
> which(W[3,]>115000)
 [1]  1 16 17 19 20 27 32 37 38 50
> which(W[3,]<500)
[1]  3  4  5 14
>
> medfly$A[3]
[1] 169
>
> medfly$A[c(4,5,14)]
[1]  32   1 769
>
> medfly$A[c(1, 16, 17, 19, 20, 27, 32, 37, 38, 50)]
 [1] 2696  563 1081 2684   70  497  812  203    6  616
>
```

# Correlación Espacial: Geary's C

## Medida de Autocorrelación Espacial

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(x_i - x_j)^2}{2W \sum_i (x_i - \bar{x})^2}$$

where $N$ is the number of spatial units indexed by $i$ and $j$; $x$ is the variable of interest; $\bar{x}$ is the mean of $x$; $w_{ij}$ is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$); and $W$ is the sum of all $w_{ij}$.

- $C > 0$
- Valores de $C$ que son significativamente menores que 1 corresponden a autocorrelación espacial positiva.
- Valores de $C$ que son significativamente mayores que 1 corresponden a autocorrelación espacial negativa.

El valor del Geary's C de la variable explicativa ($log(1 + A)$) es igual a 0.0677 $\Rightarrow$ autocorrelación espacial positiva.

Ejercicio: Calcular el valor del Geary's C de los residuos del modelo lineal, ajustado a los datos.

# Correlación Espacial: Indice de Moran

## Medida de Autocorrelación Espacial

Moran's $I$ is defined as

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where $N$ is the number of spatial units indexed by $i$ and $j$; $x$ is the variable of interest; $\bar{x}$ is the mean of $x$; $w_{ij}$ is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$); and $W$ is the sum of all $w_{ij}$.

- $I \in (-1, 1)$
- $E(I) = \dfrac{-1}{N+1}$
- Valores de $I$ que son significativamente menores que $\dfrac{-1}{N+1}$ corresponden a autocorrelación espacial negativa.
- Valores de $I$ que son significativamente mayores que $\dfrac{-1}{N+1}$ corresponden a autocorrelación espacial positiva.

Ejercicio: Calcular l valor de $I$ de los residuos del modelo lineal, ajustado a los datos.

# Correlación Espacial: Indice de Moran

$$I \sim N(E(I), Var(I))$$

$$E(I) = \frac{-1}{N+1}$$

$$\mathrm{Var}(I) = \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)W^2} - (E(I))^2$$

$$S_1 = \frac{1}{2}\sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_i \left( \sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

$$S_3 = \frac{N^{-1} \sum_i (x_i - \bar{x})^4}{(N^{-1} \sum_i (x_i - \bar{x})^2)^2}$$

$$S_4 = (N^2 - 3N + 3)S_1 - NS_2 + 3W^2$$

$$S_5 = (N^2 - N)S_1 - 2NS_2 + 6W^2$$

# Vectores Aleatorios

Sean $U = (U_1, U_2, ..., U_n)^t$ y $V = (V_1, V_2, ..., V_m)^t$ dos vectores aleatorios, tal que

$EU_i = \mu_i,\ i = 1, ..., n$ y $EV_j = \nu_j,\ j = 1, ..., m$

$Var(U) = \Sigma_u,\ Var(V) = \Sigma_v$ y $Cov(U, V) = C$

Entonces para las matrices $A$ de dimención $p \times n$ y $B$ de dimención $q \times m$:

1. $E(AU) = AE(U) = A\mu$ .
2. $Var(AU) = AVar(U)A^t = A\Sigma_u A^t$
3. $Cov(AU, BV) = ACov(U, V)B^t = ACB^t$

Ejercicio. Sean $U = (U_1, U_2, U3)^t$ y $V = (V_1, V_2)^t$, tal que $E(U) = (2, 2, 3)^t,\ E(V) = (4, -2)^t$,

$$\Sigma_u = \begin{pmatrix} 3 & 1.5 & 1.2 \\ 1.5 & 1 & -0.4 \\ 1.2 & -0.4 & 1 \end{pmatrix}$$

$$\Sigma_v = \begin{pmatrix} 2 & 1.8 \\ 1.8 & 4 \end{pmatrix}$$

1.

# Ejercicio

$$C = \begin{pmatrix} -2 & 2.5 \\ 1 & 1.5 \\ 1.2 & 1.2 \end{pmatrix}$$

Definamos: $U^* = (U_1 + 2U_2, U_3 - 0.5U_1, U_1, U_3)^t$ y $V^* = (-V_1, V_1 - V_2)^t$. Calcular:

1. $E(U^*)$ y $E(V^*)$
2. $Var(U^*)$ y $Var(V^*)$
3. $Cov(U^*, V^*)$
4. Las correlaciones entre las componentes de $U^*$ y $V^*$.

# Modelo de regresión espacial

$$\begin{aligned} \mathbf{y} =& \rho \mathbf{W} \mathbf{y} + \mathbf{X}\beta + \epsilon \\ \epsilon \sim& N(0, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Entonces,
$y - \rho W y = X\beta + \epsilon \Rightarrow (I_n - \rho W)y = X\beta + \epsilon \Rightarrow$
$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\epsilon$
Sea $\Omega = (I_n - \rho W)^{-1}$, $X^* = \Omega X$ y $\epsilon^* = \Omega\epsilon$,
entonces se puede escribir:

$$y = X^*\beta + \epsilon^*,$$

donde

- La matriz $X^*$ es una matriz $n \times p$ que depende del parámetro desconocido $\rho$.
- $\epsilon^* \sim N(0, V)$, donde $V = \Omega\Omega^t\sigma^2$ (porque?)

# Modelo de regresión espacial

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X}\beta + \epsilon$$
$$\epsilon \sim N(0, \sigma^2 \mathbf{I}_n)$$

Entonces,
$y - \rho W y = X\beta + \epsilon \Rightarrow (I_n - \rho W) y = X\beta + \epsilon \Rightarrow$
$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\epsilon$
Sea $\Omega = (I_n - \rho W)^{-1}$, $X^* = \Omega X$ y $\epsilon^* = \Omega\epsilon$,
entonces se puede escribir:

$$y = X^*\beta + \epsilon^*,$$

donde

- La matriz $X^*$ es una matriz $n \times p$ que depende del parámetro desconocido $\rho$.
- $\epsilon^* \sim N(0, V)$, donde $V = \Omega\Omega^t\sigma^2$ (porque?)