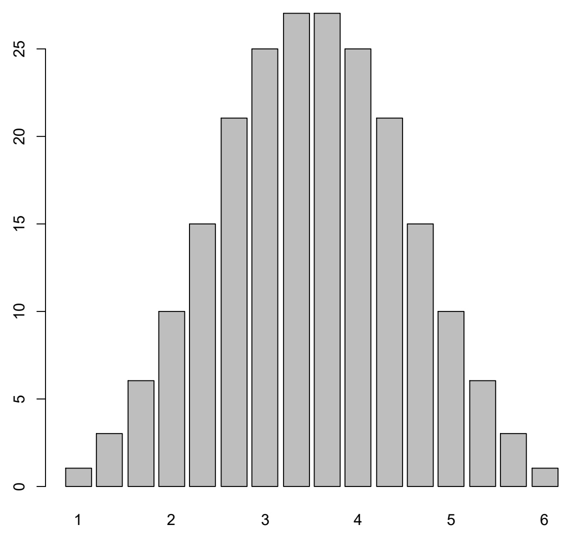
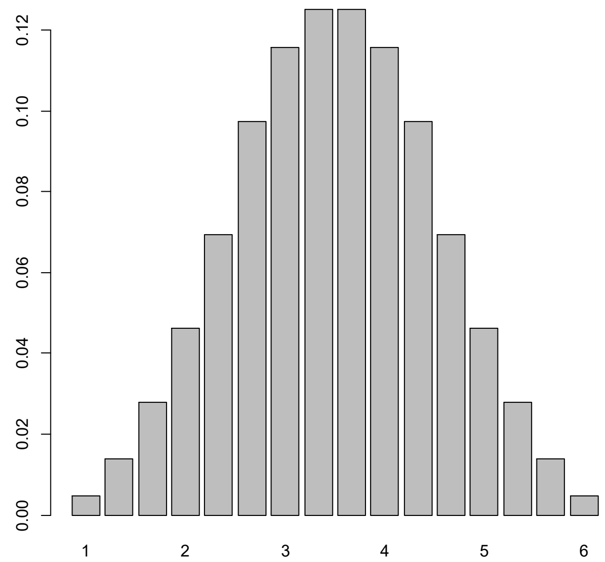
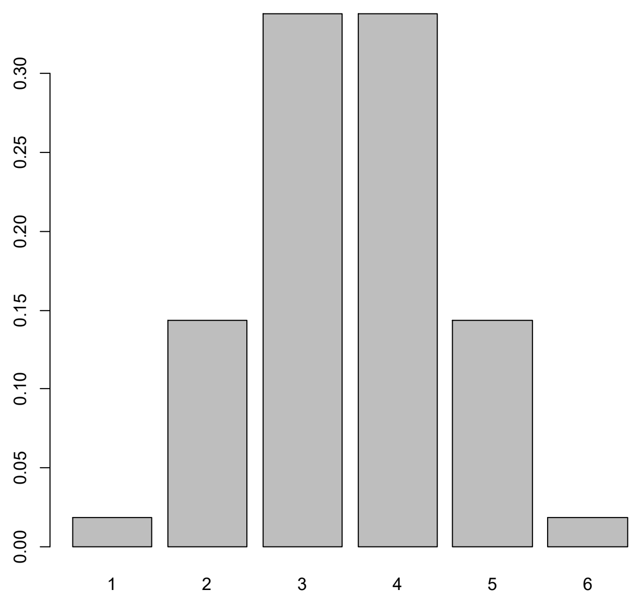
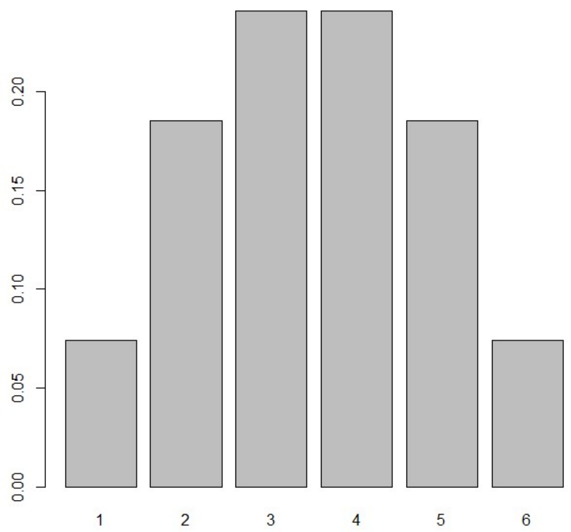
```
> D = 1:6
> # Tres dados, todas las posibilidades
> E = expand.grid(D,D,D)
> dim(E)
[1] 216   3
> E$media = apply(E[,1:3], 1, mean)
> table(E$media)

          1 1.3333333333 1.6666666667            2
          1            3            6           10
2.3333333333 2.6666666667            3 3.3333333333
         15           21           25           27
3.6666666667            4 4.3333333333 4.6666666667
         27           25           21           15
          5 5.3333333333 5.6666666667            6
         10            6            3            1
> barplot(table(E$media))
> barplot(table(E$media)/216)
```



```
> # media redondeada (ahora vale 1, 2, 3, 4, 5, 6)
> E$rmed = round(E$media)
> barplot(table(E$rmed)/216)
> cbind(table(E$rmed)/216)
        [,1]
1 0.01851852
2 0.14351852
3 0.33796296
4 0.33796296
5 0.14351852
6 0.01851852
> # mediana (valor central)
> E$mediana = apply(E[,1:3], 1, median)
> barplot(table(E$mediana)/216)
> cbind(table(E$mediana)/216)   # func probab
mediana tres dados
[,1]
1 0.07407407
2 0.18518519
3 0.24074074
4 0.24074074
5 0.18518519
6 0.07407407
```



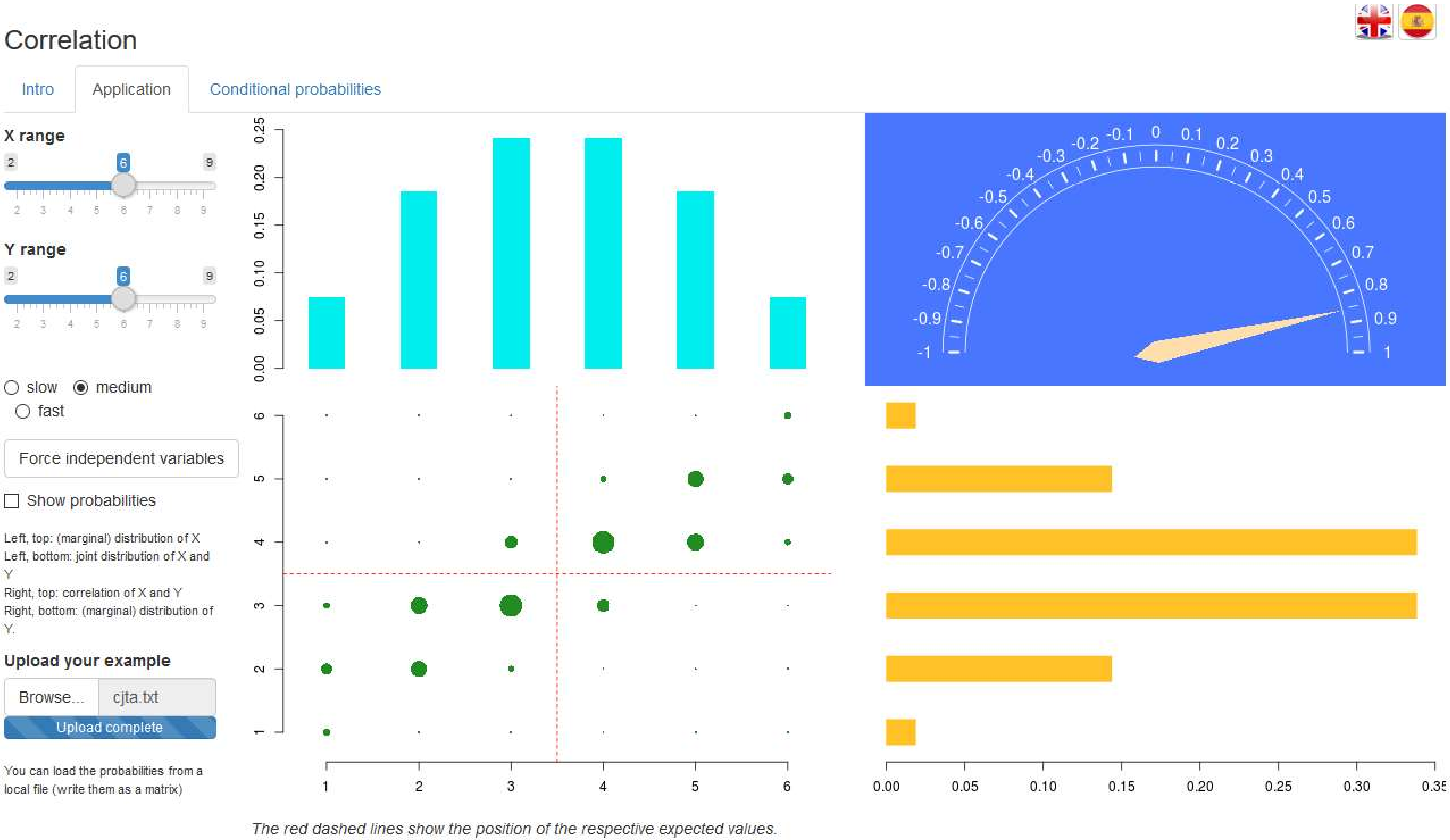



```
> p.X = table(E$media)/216
> p.Y = table(E$rmed)/216
> p.Z = table(E$mediana)/216
> sum( seq(1, 6, by=1/3) * p.X )
[1] 3.5
> sum( D * p.Y )
[1] 3.5
> sum( D * p.Z )
[1] 3.5
> sum( seq(1, 6, by=1/3)^2 * p.X )
[1] 13.22222
> sum( D^2 * p.Y )
[1] 13.2963
> sum( D^2 * p.Z )
[1] 14.12963
> 13.22222 - 3.5^2
[1] 0.97222
> 13.2963 - 3.5^2
[1] 1.0463
> 14.12963 - 3.5^2
[1] 1.87963
```

$\mu$

$\sigma^2$

```
> # ¿cómo es la distribución conjunta de media
(redondeada) y mediana?
> round(table(E$rmed, E$mediana)/216, 5)

          1       2       3       4       5       6
  1 0.01852 0.00000 0.00000 0.00000 0.00000 0.00000
  2 0.04167 0.08796 0.01389 0.00000 0.00000 0.00000
  3 0.01389 0.09722 0.17130 0.05556 0.00000 0.00000
  4 0.00000 0.00000 0.05556 0.17130 0.09722 0.01389
  5 0.00000 0.00000 0.00000 0.01389 0.08796 0.04167
  6 0.00000 0.00000 0.00000 0.00000 0.00000 0.01852
> write.table(table(E$rmed, E$mediana)/216,
"cjta.txt", row.names=FALSE, col.names=FALSE)
> # para leer con el aplicativo "corre"
> cjta = table(E$rmed, E$mediana)/216
```


Correlation
Intro
Application
Conditional probabilities
X range
Y range
slow
medium
fast
Force independent variables
Show probabilities
Left, top: (marginal) distribution of X
Left, bottom: joint distribution of X and Y
Right, top: correlation of X and Y
Right, bottom: (marginal) distribution of Y.
Upload your example
Browse...
cjta.txt
Upload complete
You can load the probabilities from a local file (write them as a matrix)
The red dashed lines show the position of the respective expected values.

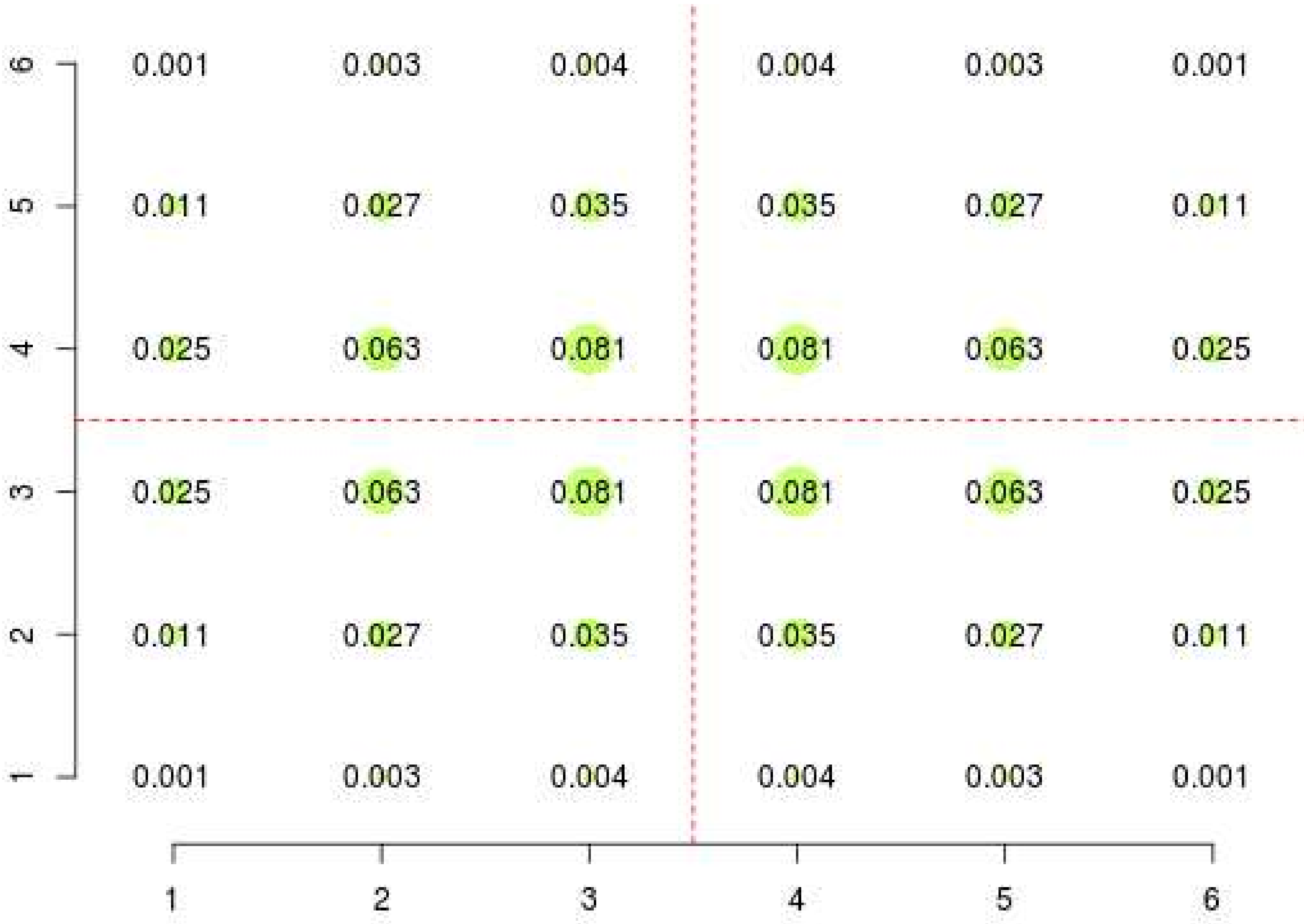
**¿Son media (redondeada) y mediana variables independientes?**

Obviamente no.

Cuanto mayor es una, mayor es la otra.

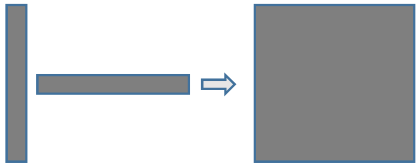
Así sería la función de probabilidad conjunta si fueran independientes:

Además, vemos que la correlación es 0.85 aprox., muy alta.



```
> # Cálculo manual de la correlación de dos VAD
> # paso 1: hacer productos cruzados
> (D-3.5) %*% t(D-3.5) # (asignamos a prod)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,]  6.25  3.75  1.25 -1.25 -3.75 -6.25
[2,]  3.75  2.25  0.75 -0.75 -2.25 -3.75
[3,]  1.25  0.75  0.25 -0.25 -0.75 -1.25
[4,] -1.25 -0.75 -0.25  0.25  0.75  1.25
[5,] -3.75 -2.25 -0.75  0.75  2.25  3.75
[6,] -6.25 -3.75 -1.25  1.25  3.75  6.25
> # paso 2: productos cruzados × probabilidades
> prod * cjta
          1        2        3        4        5        6
  1 0.11574  0.00000  0.00000  0.00000  0.00000  0.00000
  2 0.15625  0.19792  0.01042  0.00000  0.00000  0.00000
  3 0.01736  0.07292  0.04282 -0.01389  0.00000  0.00000
  4 0.00000  0.00000 -0.01389  0.04282  0.07292  0.01736
  5 0.00000  0.00000  0.00000  0.01042  0.19792  0.15625
  6 0.00000  0.00000  0.00000  0.00000  0.00000  0.11574
> # paso 3: covarianza (sumar todo)
> sum(prod * cjta)
[1] 1.199074
> sum(prod * cjta) / sqrt(1.0463*1.87963) # paso 4: correlación
[1] 0.8550307
```

$$Cov(X,Y)=\sum_{\forall x}\sum_{\forall y}\underbrace{(x-E(x))(y-E(y))}_{1}\cdot \underbrace{p_{XY}(x,y)}_{2}$$



La covariancia interviene en varias propiedades de los indicadores de variables no independientes:

$$V(X+Y) = V(X) + V(Y) + 2\ \mathrm{Cov}(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2\ \mathrm{Cov}(X,Y)$$

$$E(X\cdot Y) = E(X)\cdot E(Y) + \mathrm{Cov}(X,Y)$$

Podemos comprobar que realmente es así.

Suma de media (redondeada) y mediana:

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

```
> # ... needs some programming ☹
> x = D
> y = D
> z = c()
> for (i in y) z = c(z, x+i)
> Z = matrix(z, ncol=length(y))
> Z
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    2    3    4    5    6    7
[2,]    3    4    5    6    7    8
[3,]    4    5    6    7    8    9
[4,]    5    6    7    8    9   10
[5,]    6    7    8    9   10   11
[6,]    7    8    9   10   11   12
> sum(Z*cjta)
[1] 7   # OK: E(Suma)=E(X)+E(Y) = 3.5 + 3.5
> sum(Z^2*cjta)
[1] 54.32407
> # 54.32407 - 7^2 = 5.32407: V(Suma)
> # Hallamos V(Suma) usando la propiedad:
> # V(X)+V(Y)+2Cov(X,Y) =
1.0463+1.87963+2*1.199074 = 5.324078 (OK)
```

```
> # Por simulación
> # generamos miles de lanzamientos de
tres dados.
> M = 10000
> d1 = sample(D, M, replace=TRUE)
> d2 = sample(D, M, replace=TRUE)
> d3 = sample(D, M, replace=TRUE)
> A = data.frame(d1, d2, d3)
> A$rmed= round(apply(A[,1:3], 1, mean))
> A$mediana = apply(A[,1:3], 1, median)
> var(A$rmed)
[1] 1.042468
> var(A$mediana)
[1] 1.860979
> cor(A$rmed, A$mediana)
[1] 0.8519287
> var(A$rmed + A$mediana)
[1] 5.276654
> jitter2 = function(u)
{u+rnorm(length(u),0,0.15)}
> plot(jitter2(A$rmed),
jitter2(A$mediana), pch='.', cex=1.5)
```

nota: sin "jitter" (ruido), los puntos se superponen y no se aprecia la diferencia de densidad.