Pearson Bayesiano

(Coeficiente de Correlación)

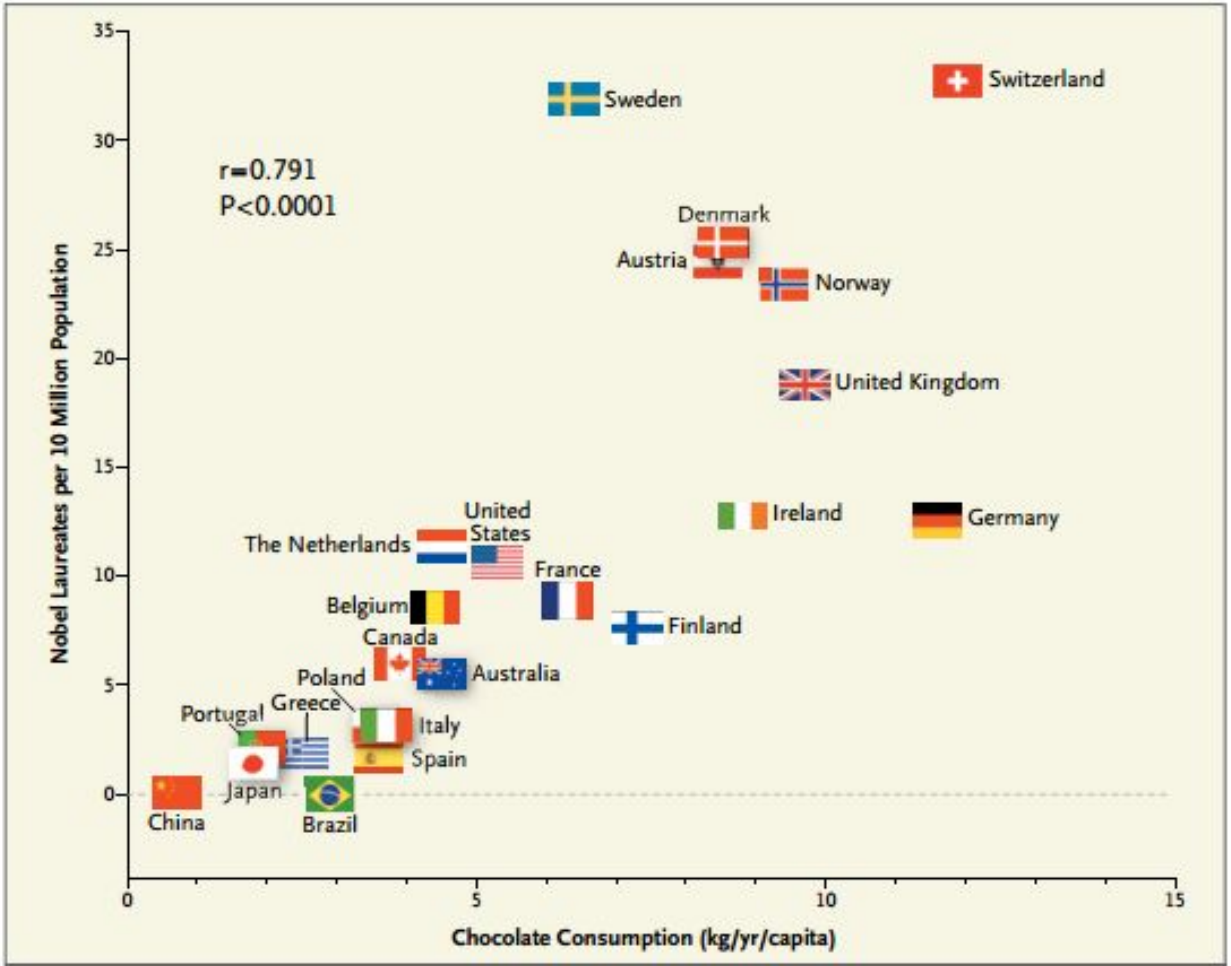


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

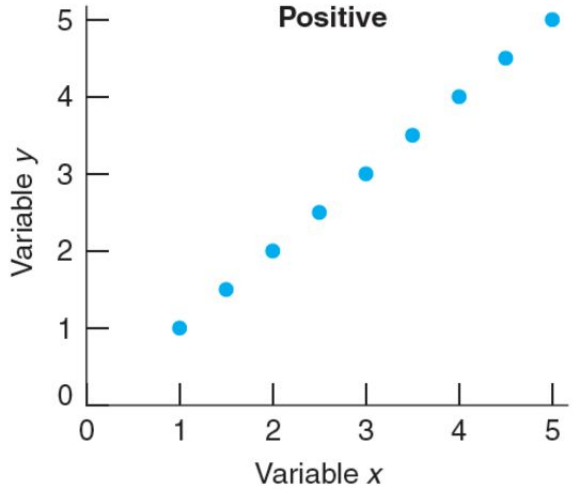
Pearson Product-Moment Correlation Coefficient

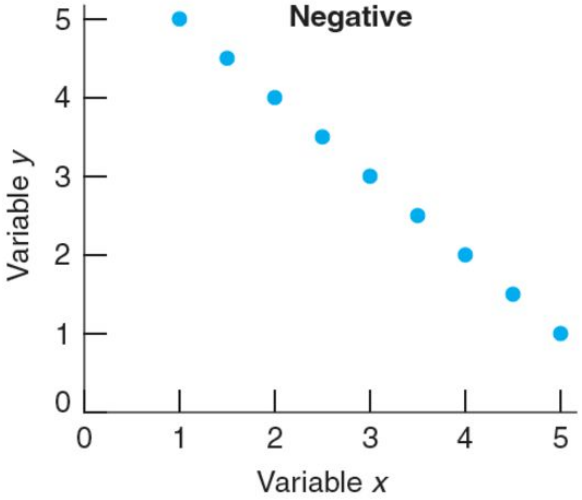
r

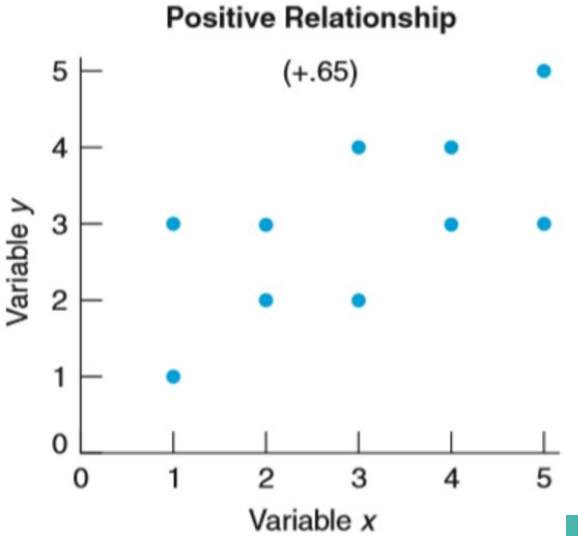
- Estadístico que evalúa la relación entre dos variables

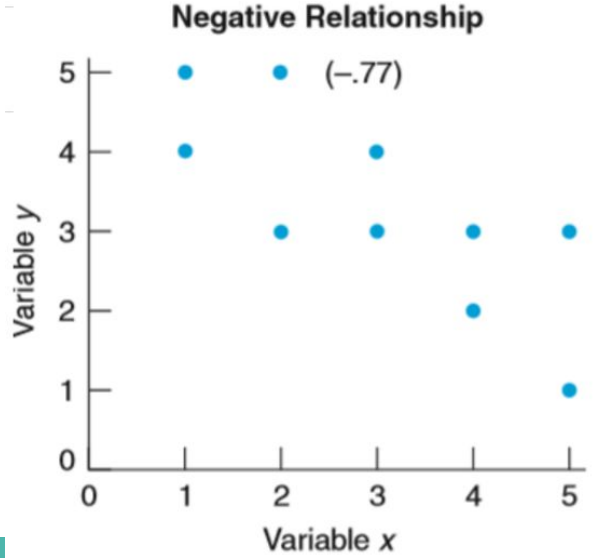
-1 a 1

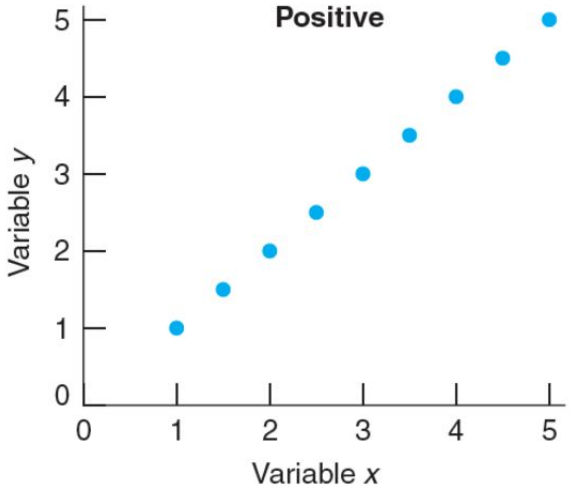
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

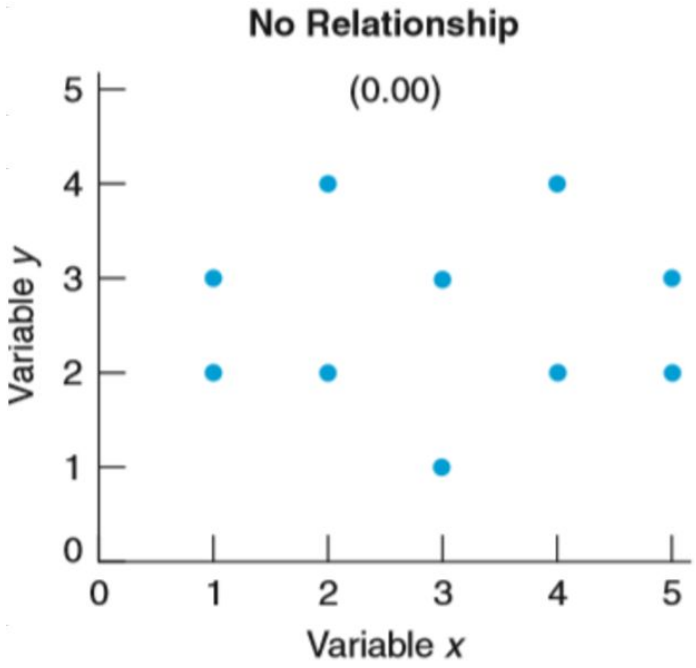

Positive
Variable y
Variable x
0
1
2
3
4
5

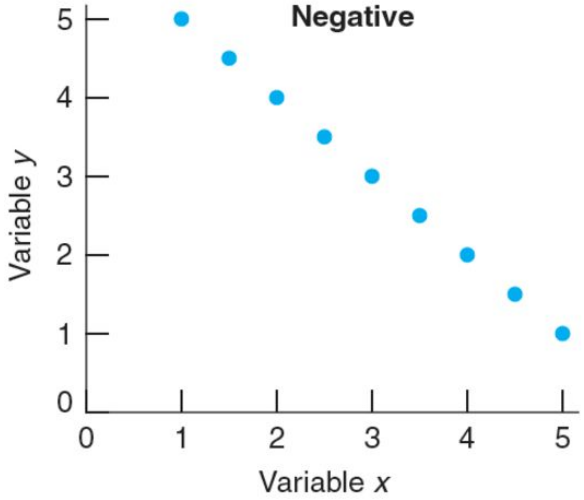

Negative
Variable y
Variable x
0
1
2
3
4
5

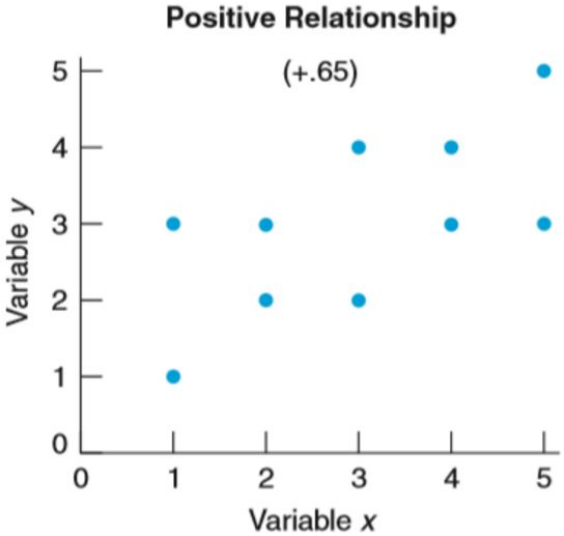

Positive Relationship
(+.65)
Variable y
Variable x
0
1
2
3
4
5

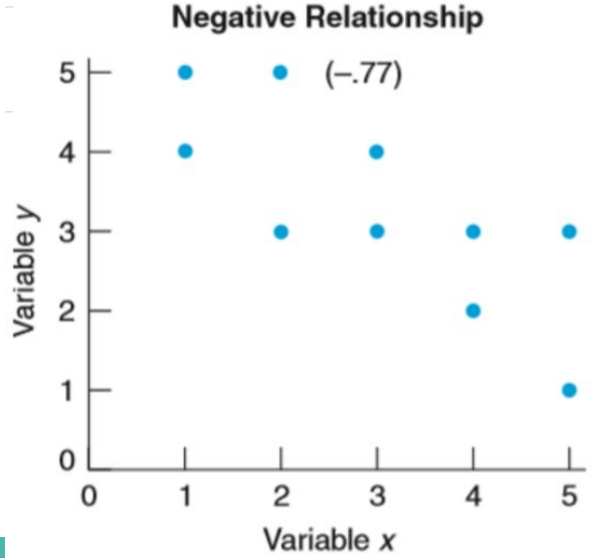

Negative Relationship
(–.77)
Variable y
Variable x
0
1
2
3
4
5


Positive
Variable y
Variable x
0
1
2
3
4
5


No Relationship
(0.00)
Variable y
Variable x
0
1
2
3
4
5


Negative
Variable y
Variable x
0
1
2
3
4
5


Positive Relationship
(+.65)
Variable y
Variable x
0
1
2
3
4
5


Negative Relationship
(–.77)
Variable y
Variable x
0
1
2
3
4
5

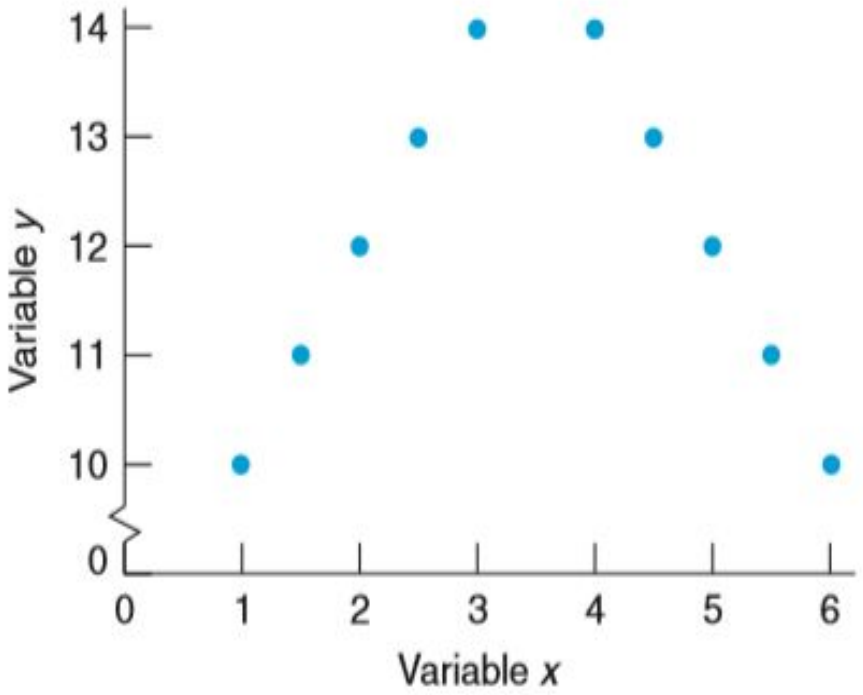
Supuestos

1. Asume que los datos X y Y se distribuyen normalmente

2. Asume que la relación entre X y Y es lineal

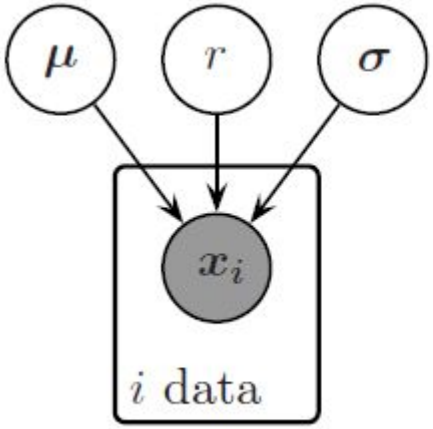
Limitaciones

- Puede subestimar la ‘significancia’ de las relaciones no lineales.
- Requiere recopilar datos de un rango amplio de valores tanto de X como de Y.



Pearson Bayesiana

Implica hacer un estimado sobre los posibles valores que la **r de Pearson** puede tener (distribución posterior).

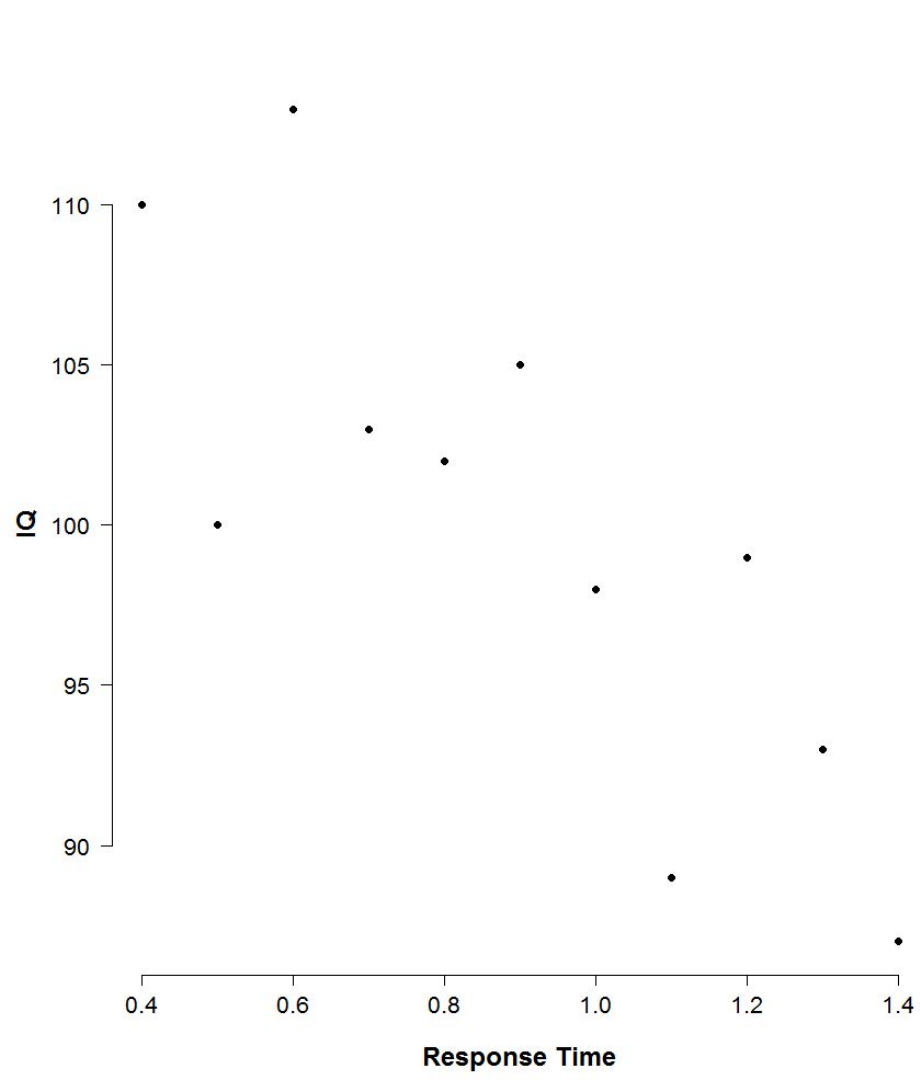


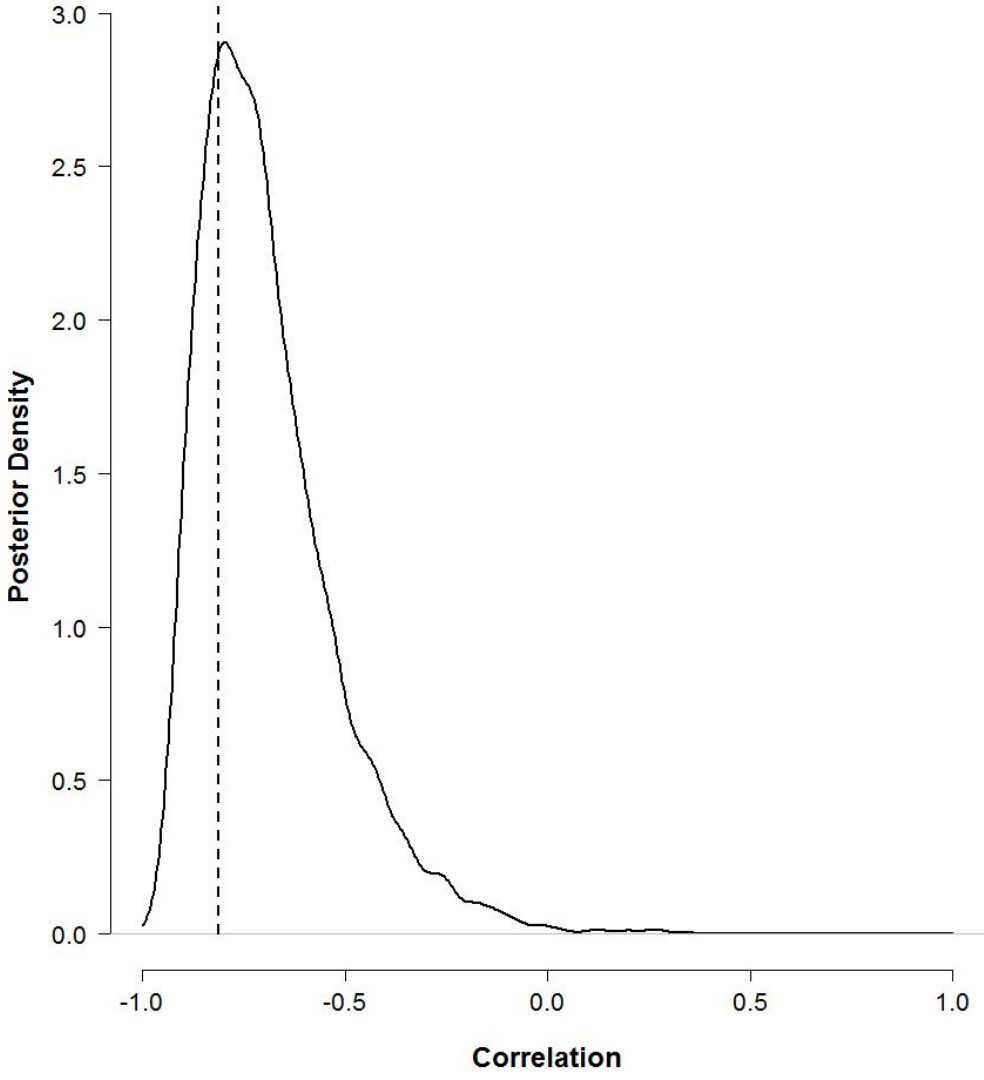
$$\mu_1, \mu_2 \sim \text{Gaussian}(0, 0.001)$$

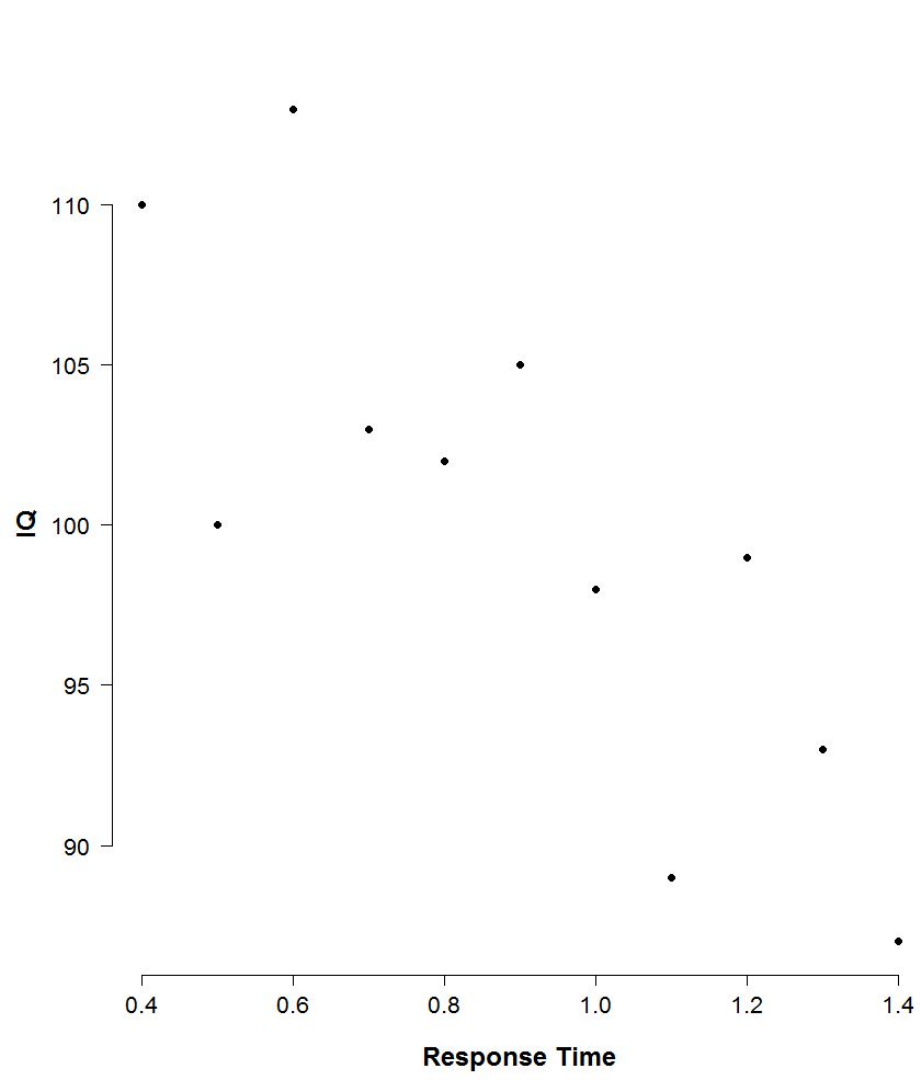
$$\sigma_1, \sigma_2 \sim \text{InvSqrtGamma}(0.001, 0.001)$$

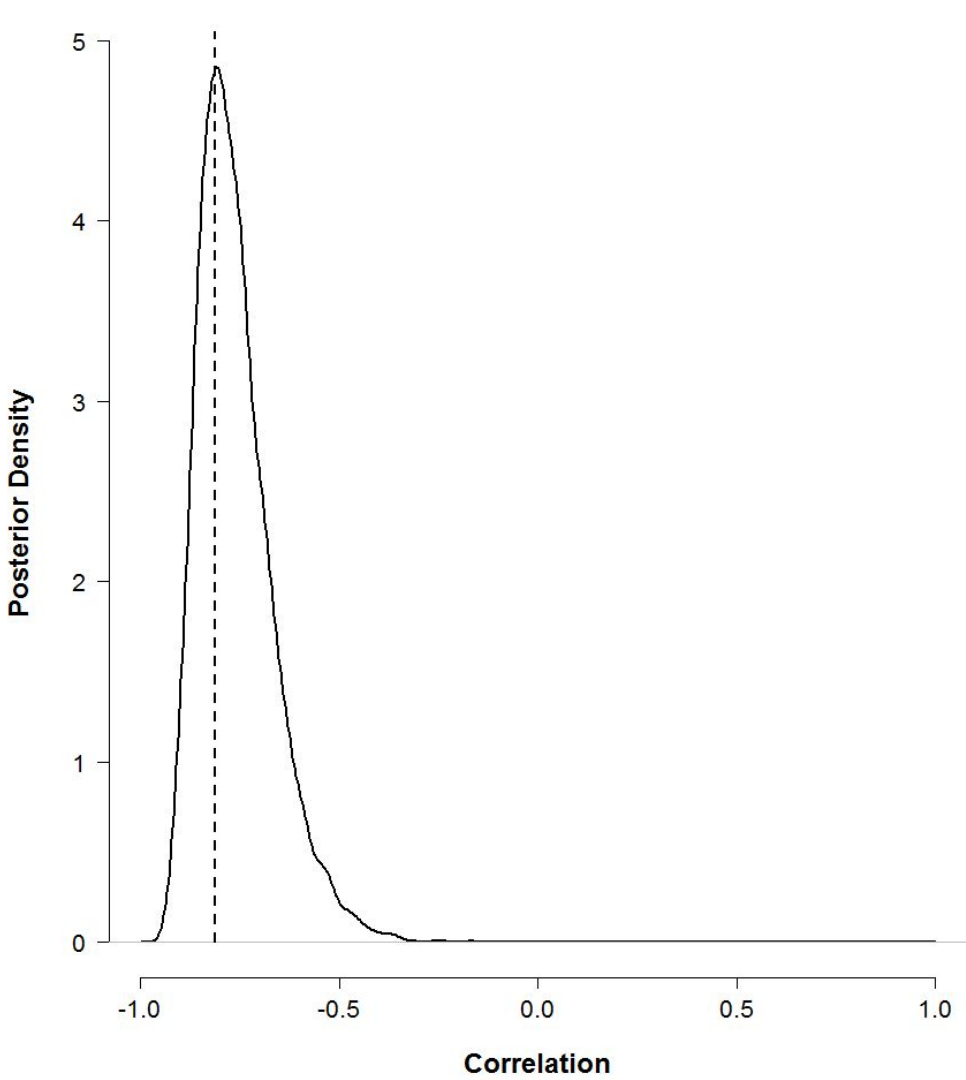
$$r \sim \text{Uniform}(-1, 1)$$

$$\boldsymbol{x}_i \sim \text{MvGaussian}\left((\mu_1, \mu_2), \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \right)$$


IQ
110
105
100
95
90
0.4
0.6
0.8
1.0
1.2
1.4
Response Time


Posterior Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-1.0
-0.5
0.0
0.5
1.0
Correlation


IQ
110
105
100
95
90
0.4
0.6
0.8
1.0
1.2
1.4
Response Time

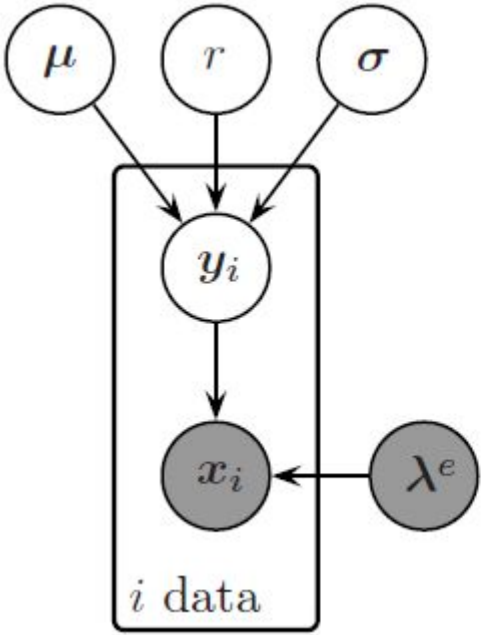

Posterior Density
5
4
3
2
1
0
-1.0
-0.5
0.0
0.5
1.0
Correlation

Pearson Bayesiana con incertidumbre

La incertidumbre en la medición de las variables a comparar debe ser tomada en cuenta.

Por ejemplo:

¿El IQ es un valor fijo o debería computarse como un valor con incertidumbre?



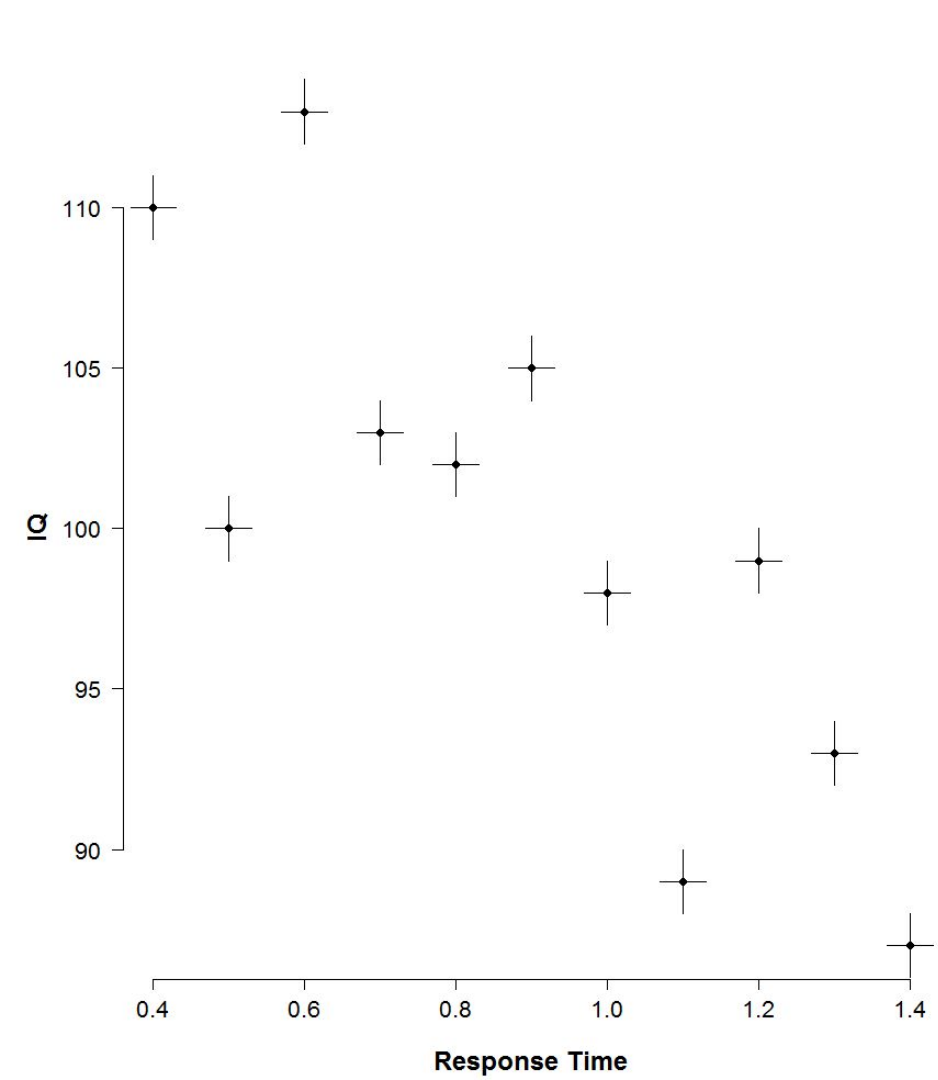
$$\mu_1, \mu_2 \sim \text{Gaussian}(0, 0.001)$$

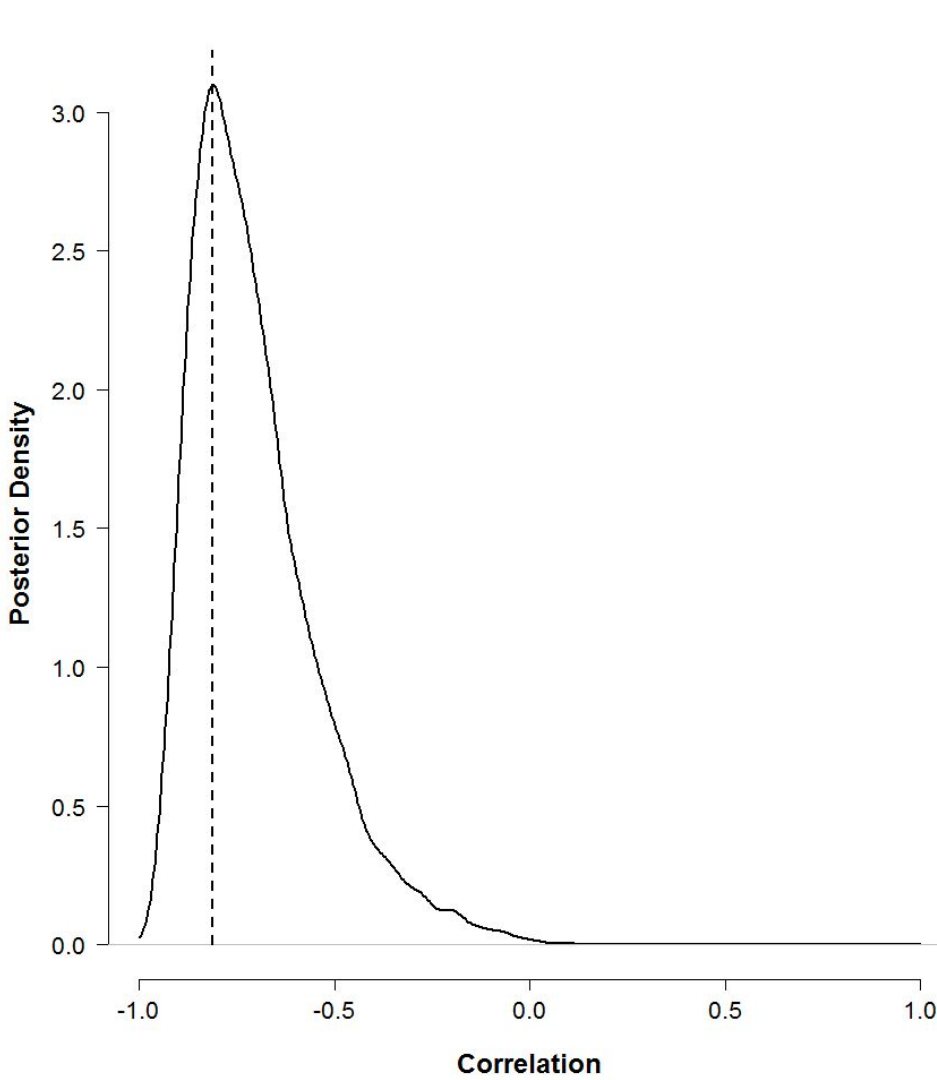
$$\sigma_1, \sigma_2 \sim \text{InvSqrtGamma}(0.001, 0.001)$$

$$r \sim \text{Uniform}(-1, 1)$$

$$\boldsymbol{y}_i \sim \text{MvGaussian}\left((\mu_1, \mu_2), \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1}\right)$$

$$x_{ij} \sim \text{Gaussian}(y_{ij}, \lambda_j^e)$$


IQ
110
105
100
95
90
0.4
0.6
0.8
1.0
1.2
1.4
Response Time


Posterior Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-1.0
-0.5
0.0
0.5
1.0
Correlation

Enlaces recomendados:

1. Gamma distribution: https://www.youtube.com/watch?v=8BLbzT9M0Ec
2. Square Root of the inverse gamma distribution https://stats.stackexchange.com/questions/5119/square-root-of-inverse-gamma-distribution