Modele liniowe raport nr 1

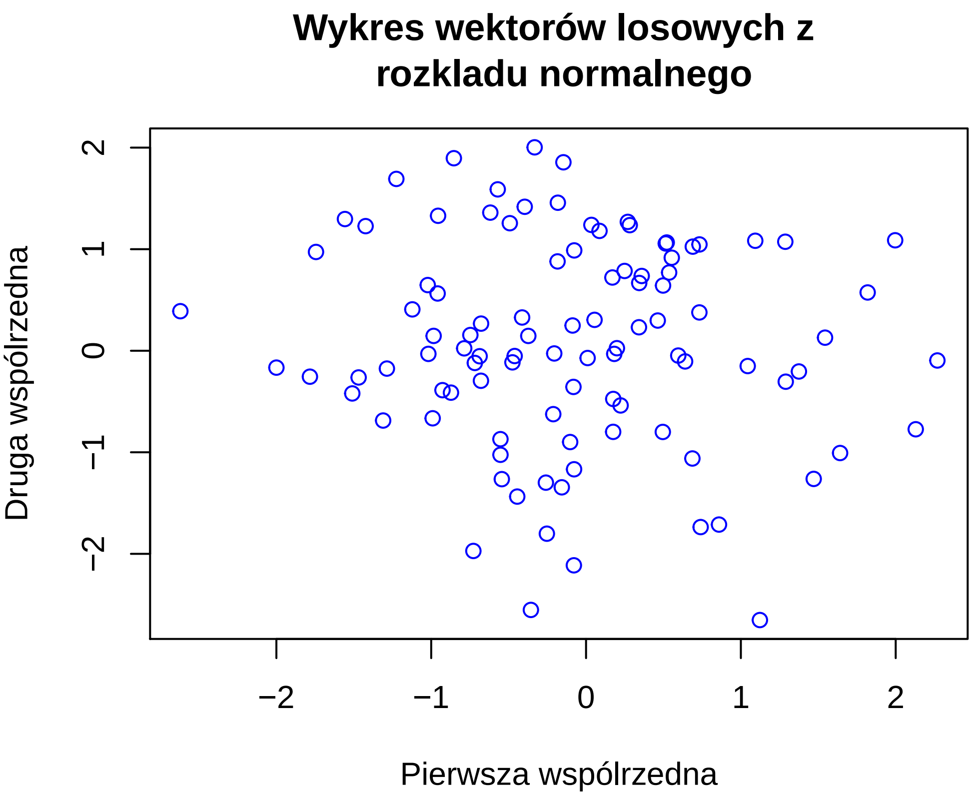
Dominik Mika

20 października 2020

1 Zadanie 1

W pierwszym zadaniu przy pomocy funkcji *rnorm* mamy wygenerować 100 wektorów losowych z rozkładu dwuwymiarowego normalnego $N(0, I)$

Aby wykonać to zadanie skorzystamy z podanej wyżej funkcji *rnorm* do wylosowania 200 wartości z rozkładu normalnego $N(0, 1)$. Następnie użyjemy funkcji *matrix* do stworzenia macierzy 100×2, gdzie każdy wiersz reprezentuje jeden wektor losowy z rozkładu normalnego dwuwymiarowego.



Widzimy, że chmura punktów jest rozrzucona co oznacza niezależność wylosowanych wektorów

2 Zadanie 2

W tym zadaniu chcemy znaleźć przekształcenie liniowe, które przekształca chmurę punktów z poprzedniego zadania w chmurę punktów z rozkładu $N(\mu, \Sigma)$, gdzie μ to wektor wartości oczekiwanej, Σ to macierz kowarancji podane w poniższych podpunktach.
Aby znaleźć takie przekształcenie skorzystamy z faktu:

Fakt 1
Dla dowolnej ustalonej macierzy $A_{k \times p}$ i wektora $B \in R^k$ definiujemy wektor losowy $Y = AX + B$. Zachodzi $\mu^Y = A\mu^X + B$ i $\Sigma^Y = A\Sigma^X A^T$.

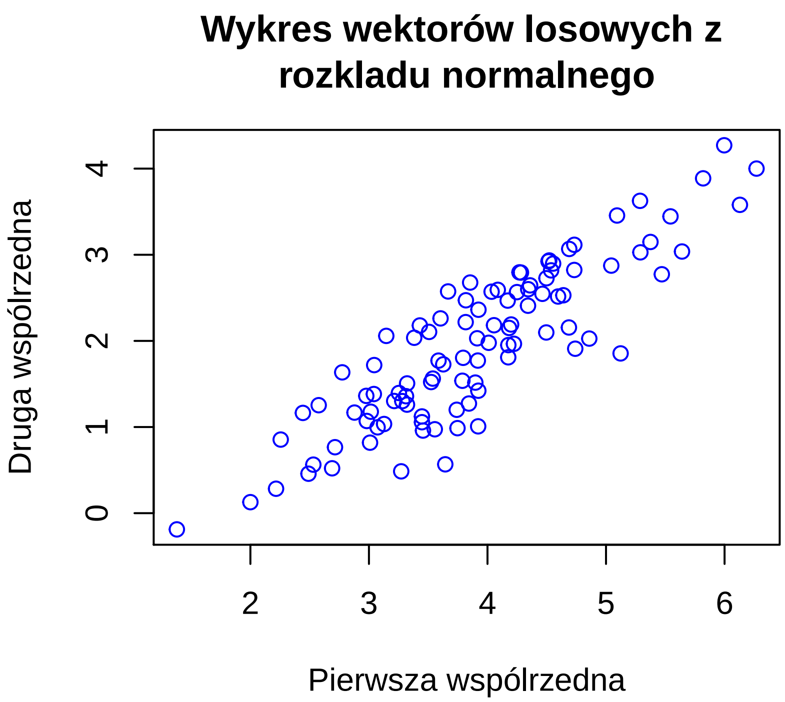
Dlatego do wyznaczenia tego przekształcenia potrzebna jest nam macierz A z powyższego faktu. Do znalezienia jej posłużymy się metodą Choleskiego. Skorzystamy z wbudowanej funkcji w R (*chol*), która wyznacza macierz A przy pomocy macierzy Σ. Przechodząc już konkretnie do naszego zadania w naszym przypadku we wzorze $Y = AX + B$: X-macierz wektorów losowych z zadania pierwszego, A-szukana macierz przekształcenia, B-wektor wartości oczekiwanej.

Do wszystkich podpunktów będziemy stosować tą samą powyższą metodę, będą zmieniać się tylko dane, dlatego nie będę niżej powtarzał jej opisu.

2.1 a)

W pierwszym podpunkcie mamy dane:

$$\mu = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$



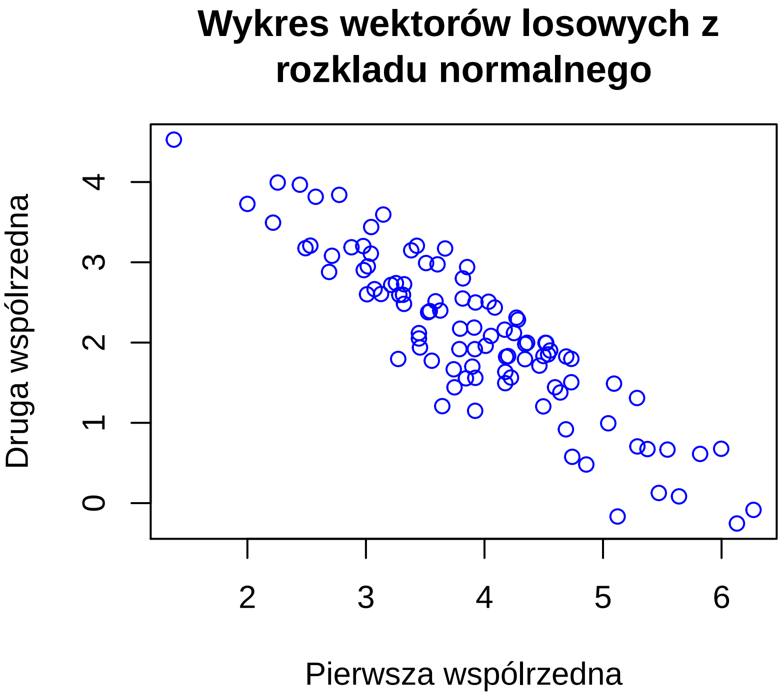
Widzimy, że punkty te są bardziej skupione co oznacza jakąś zależność między wektorami, co nie powinno nas dziwić, gdyż kowariancje wynoszą 0.9.

Natomiast nasze przekształcenie liniowe przedstawia się wzorem $Y_1 = (A_1 X)^T + B$, gdzie:
X- macierz wektorów losowych z pierwszego zadania,

$$A_1 = \begin{pmatrix} 1 & 0.9 \\ 0 & 0.44 \end{pmatrix}, \quad B = \begin{pmatrix} \mu^T \\ \vdots \\ \mu^T \end{pmatrix}$$

2.2 b)

$$\mu = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$



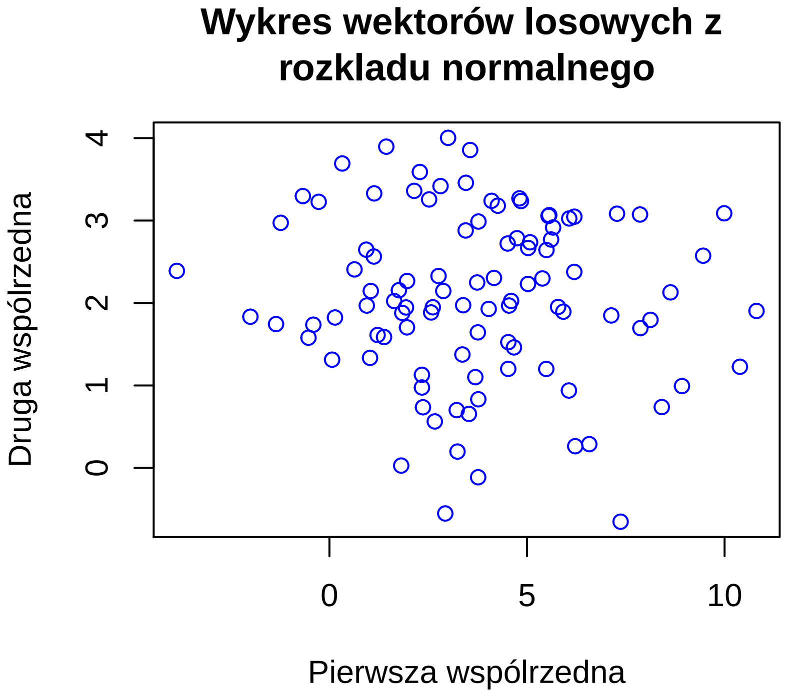
Punkty również są skupione jak w poprzednim podpunkcie, tylko blisko innej prostej, ponieważ w tym przypadku kowariancje są równe -0.9.

Przekształcenie afiniczne:
$Y_2 = (A_2 X)^T + B$, gdzie:
X- macierz wektorów losowych z pierwszego zadania,

$$A_2 = \begin{pmatrix} 1 & -0.9 \\ 0 & 0.44 \end{pmatrix}, \quad B = \begin{pmatrix} \mu^T \\ \vdots \\ \mu^T \end{pmatrix}$$

2.3 c)

$$\mu = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$$



Widzimy, że tutaj punkty są rozrzucone z powodu nieskorelowania.

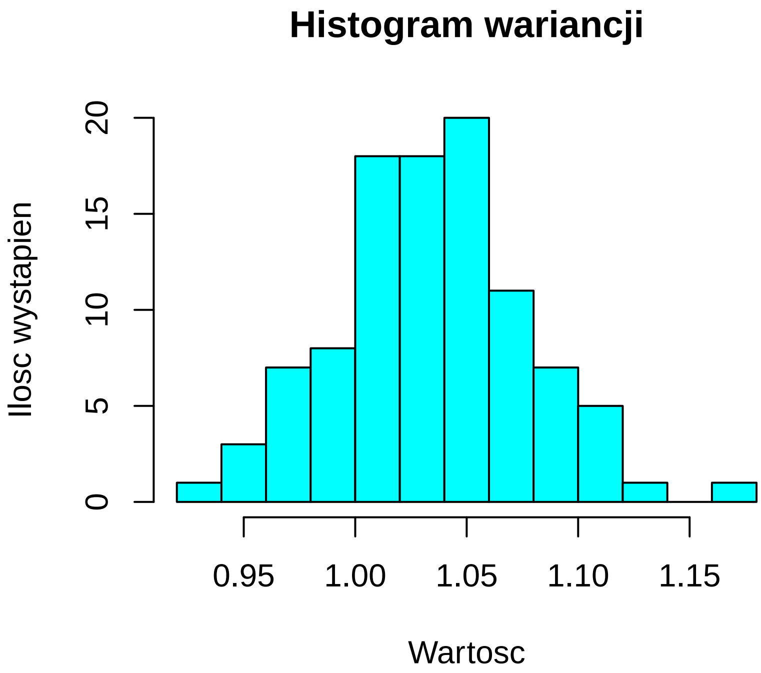
Przekształcenie afiniczne:
$Y_3 = (A_3X)^T + B$, gdzie:
X- macierz wektorów losowych z pierwszego zadania,

$$A_3 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} \mu^T \\ \vdots \\ \mu^T \end{pmatrix}$$

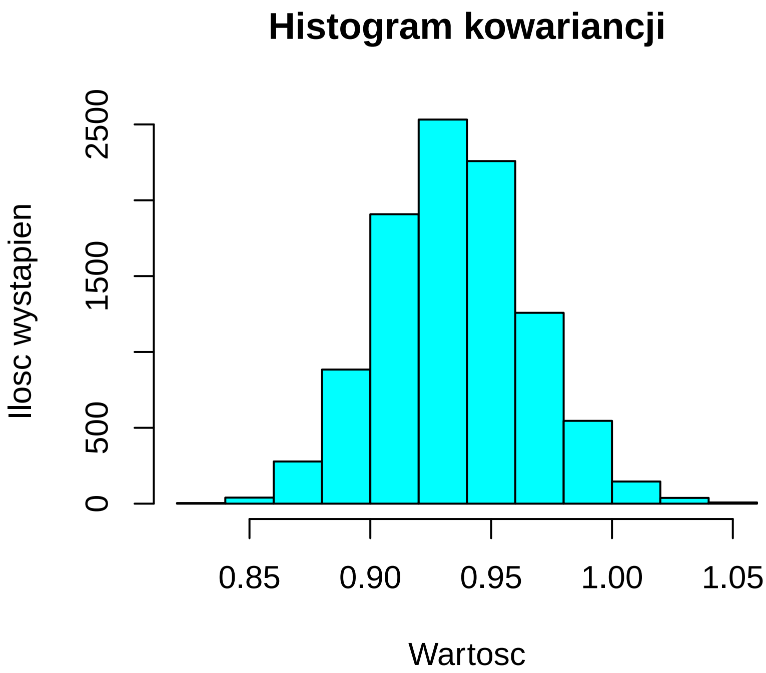
3 Zadanie 3

W tym zadaniu wykonamy podobny eksperyment jak w zadaniu pierwszym i drugim. Chcemy wylosować 200 wektorów losowych z dwuwymiarowego rozkładu normalnego- $N(0, I_{100\times100})$ i umieścić je w macierzy $X_{200\times100}$. Następnie chcemy znaleźć macierz A (przy pomocy funkcji *chol*), taką żeby $\tilde{X} = XA$ zawierała 200 wektorów z rozkładu wielowymiarowego normalnego $N(0, \Sigma_{100\times100})$, gdzie $\Sigma(i,i) = 1$ i $\Sigma(i,j) = 0.9$ dla $i \neq j$. Finalnie zweryfikujemy wyniki wyliczając średnią wariancji próbkowych oraz rysując histogramy wariancji i kowariancji.

-Otrzymaliśmy średnią wariancji próbkowych w zaokrągleniu równą 1.036.



Widzimy, że histogram tak jak oczekiwaliśmy jest prawie symetryczny względem 1.



W tym wypadku histogram jest symetryczny względem 0.9, czyli tak jak się spodziewaliśmy.

Podsumowując, dzięki kilku faktom oraz metodzie Choleskiego jesteśmy w stanie dość dobrze generować próby losowe z rozkładu normalnego o różnych parametrach. Po naszych testach, gdzie próby nie były bardzo duże możemy stwierdzić, że ta metoda działa.