Kamil Marciniak

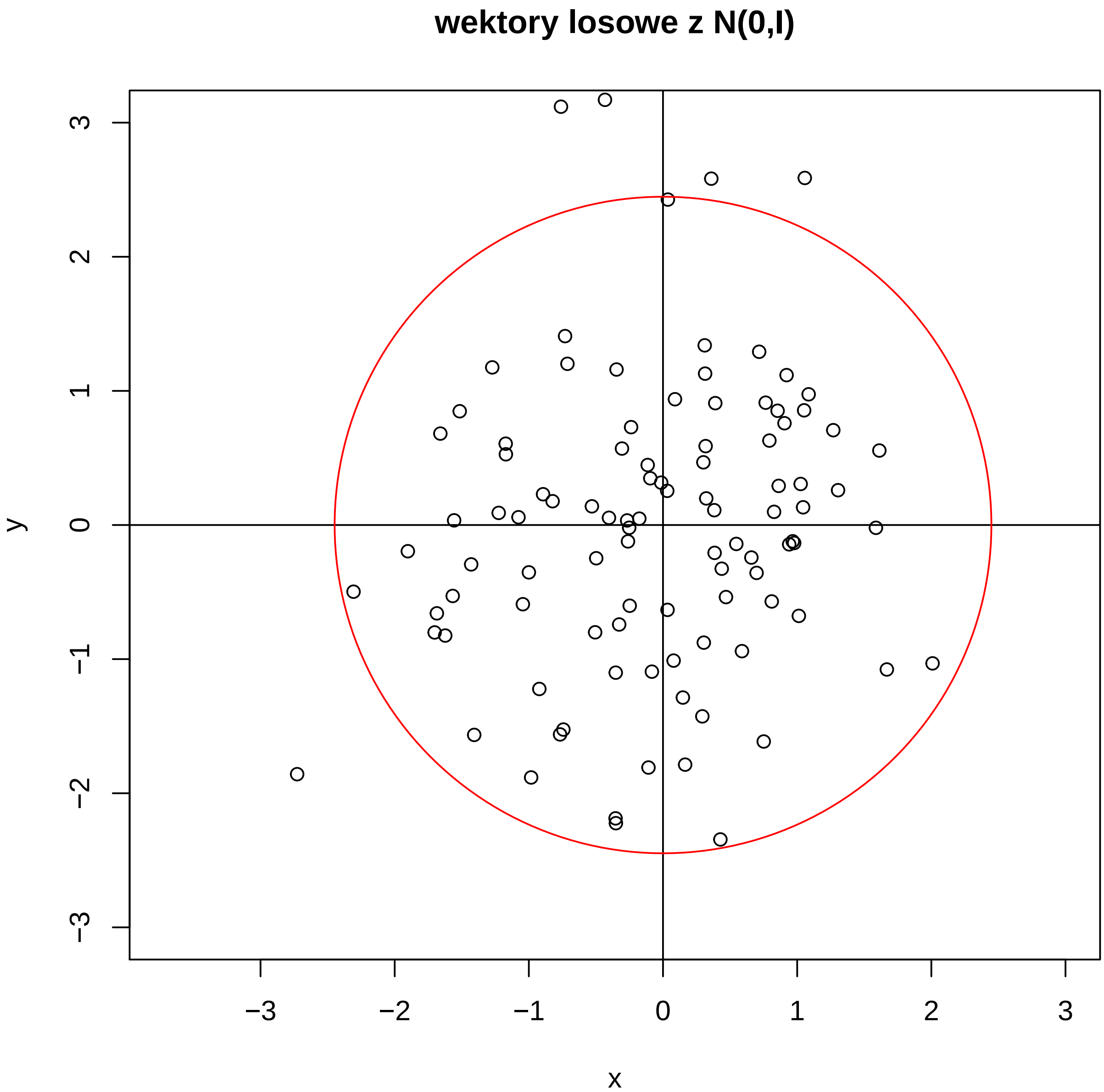
# Raport 1 Modele liniowe

23 października 2021

# Spis treści

# 1. Zadanie 1

Rozpocznijmy od narysowania dwuwymiarowego rozkładu normalnego o parametrach $N(0,I)$




Chmura punktów przyjmuje wartości losowo rozrzucone wokół punktu (0,0) zgodnie z tym jak działa rozkład normalny. Korzystając z biblioteki mixtools czerwonym kolorem narysowałem róWnież elipsę która zakreśla kontur w którym zgodnie z teoretycznie wyliczonym prawdopodobieństwem powinno znajdować się 95% obserwacji. Widzimy że elipsa w tym wypadku jest okręgiem o środku w zerze ponieważ mamy do czynienia z rozkładem $N(0,I)$

# 2. Zadanie 2

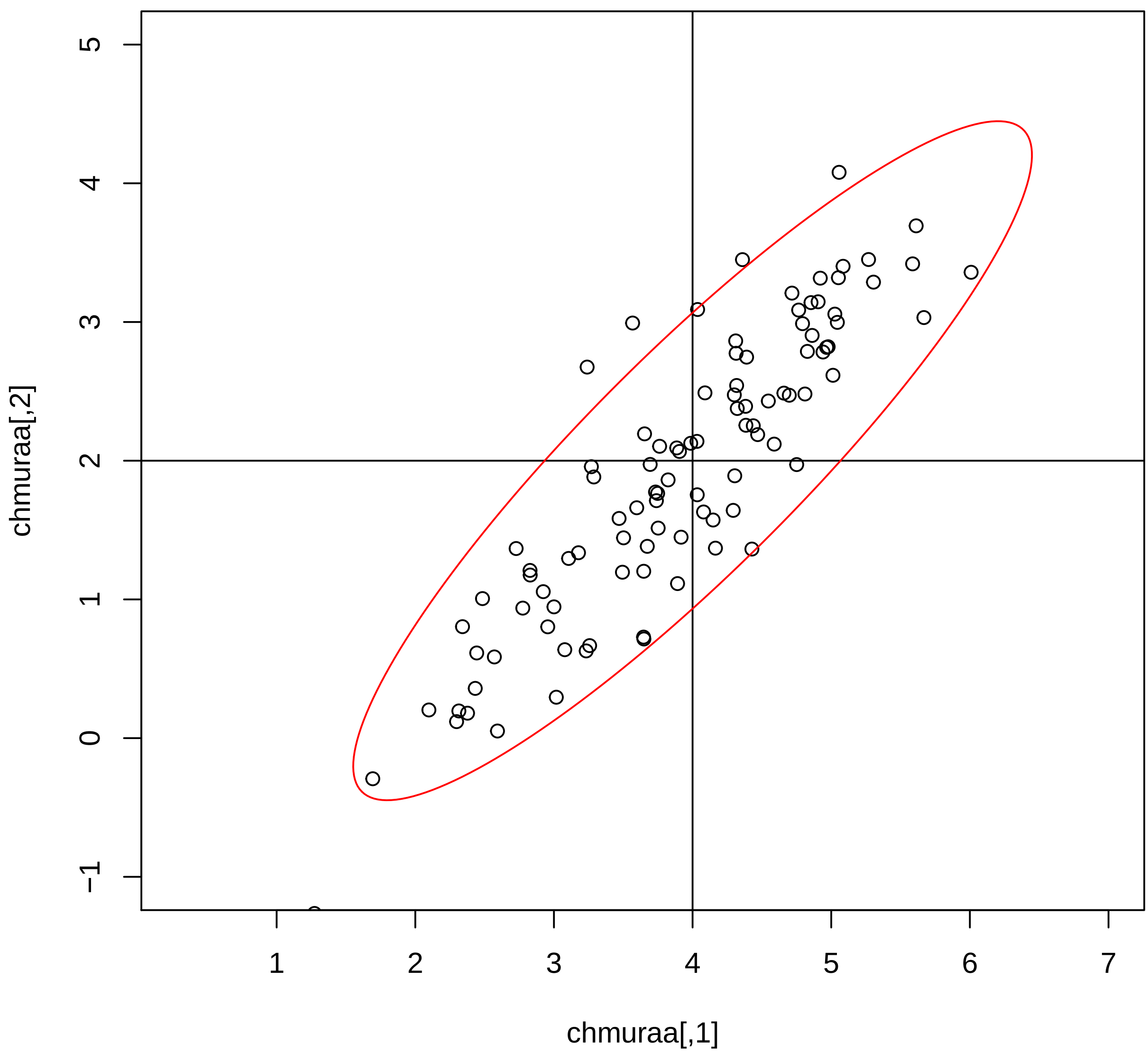
## 2.1. moja implementacja metody Choleskiego

```
choleski <- function (matrix){
  n <- nrow(matrix)
  matrixszukana <- matrix(0,nrow=n,ncol=n)
  matrixszukana[1,1] <- sqrt(matrix[1,1])
  for (j in 1:n){
    for (i in 1:n){
      if(j>=i){
        if(i==j){
          matrixszukana[i,i] <- sqrt(matrix[i,i]-sum(matrixszukana[i,1:i-1]^2))
        }else {
        matrixszukana[j,i] <- (matrix[j,i]-sum(matrixszukana[j,1:i-1]*matrixszukana[i,1:
        }
      }

    }
  }
  return(matrixszukana)
}
```

Chcemy wygenerować rozkład o parametrach $\mu = (4{,}2)$ oraz $\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. Przekształcenie liniowe którego potrzebujemy możemy uzyskać stosując metodę Choleskiego do naszej macierzy $\Sigma$

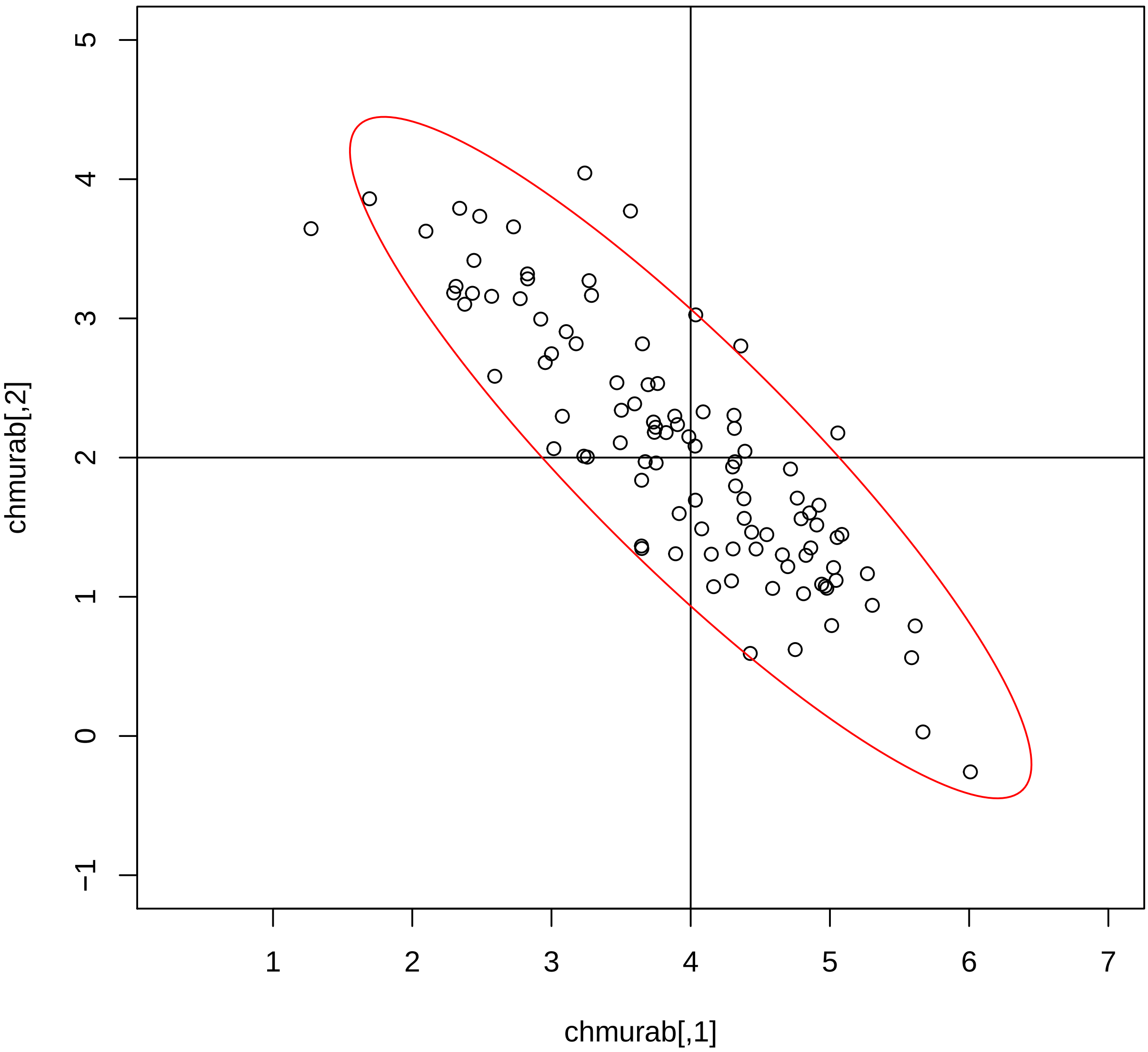
Tak więc ostatecznie okazuje się że szukanym przekształceniem jest
$X = \begin{bmatrix} 1 & 0 \\ 0.9 & 0.4358899 \end{bmatrix} \cdot \xi + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ gdzie $\xi$ ma rozkład $N(0{,}I)$
Teraz narysujmy chmurę punktów




Narysowanie elipsy okazuje się pomocne w tym aby stwierdzić że rzeczywiście wylosowaliśmy wektory losowe z odpowiedniego rokzładu ponieważ rzeczywiście zdecydowana większość punktów trafia do wnętrza elipsy

Chcemy wygenerować rozkład o parametrach $\mu = (4{,}2)$ oraz $\Sigma = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$. Przekształcenie liniowe którego potrzebujemy możemy uzyskać stosując metodę Choleskiego do naszej macierzy $\Sigma$

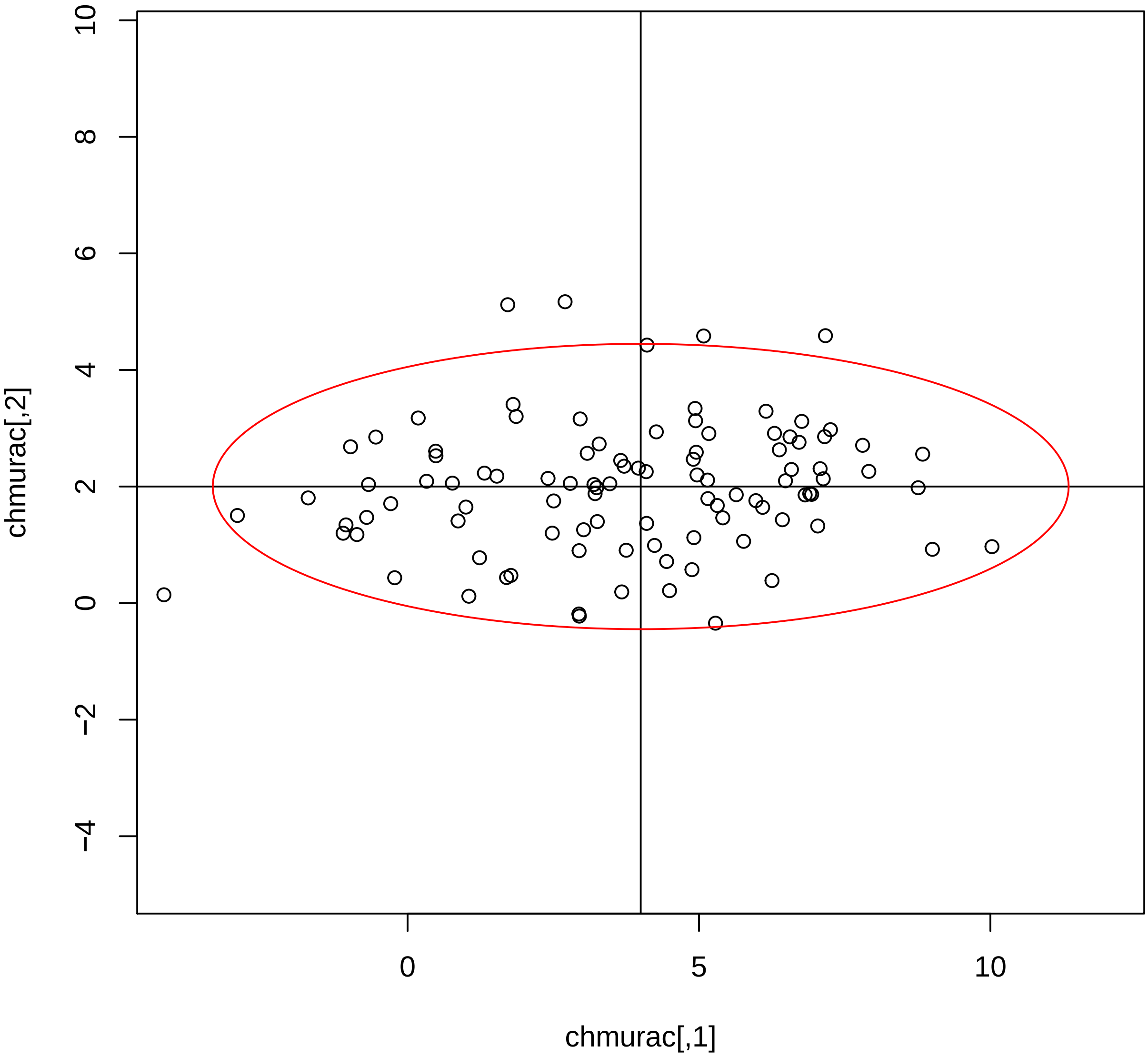
Tak więc ostatecznie okazuje się że szukanym przekształceniem jest
$X = \begin{bmatrix} 1 & 0 \\ -0.9 & 0.4358899 \end{bmatrix} \cdot \xi + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ gdzie $\xi$ ma rozkład $N(0,I)$
Teraz narysujmy chmurę punktów




Zdecydowana większość punktów trafia do wnętrza elipsy. Warto również odnotować, że zgodnie z macierzą kowariancji $\Sigma$ w tym przykładzie korelacja między zmiennymi $x$ i $y$ jest ujemna.

Chcemy wygenerować rozkład o parametrach $\mu = (4{,}2)$ oraz $\Sigma = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$. Przekształcenie liniowe którego potrzebujemy możemy uzyskać stosując metodę Choleskiego do naszej macierzy $\Sigma$

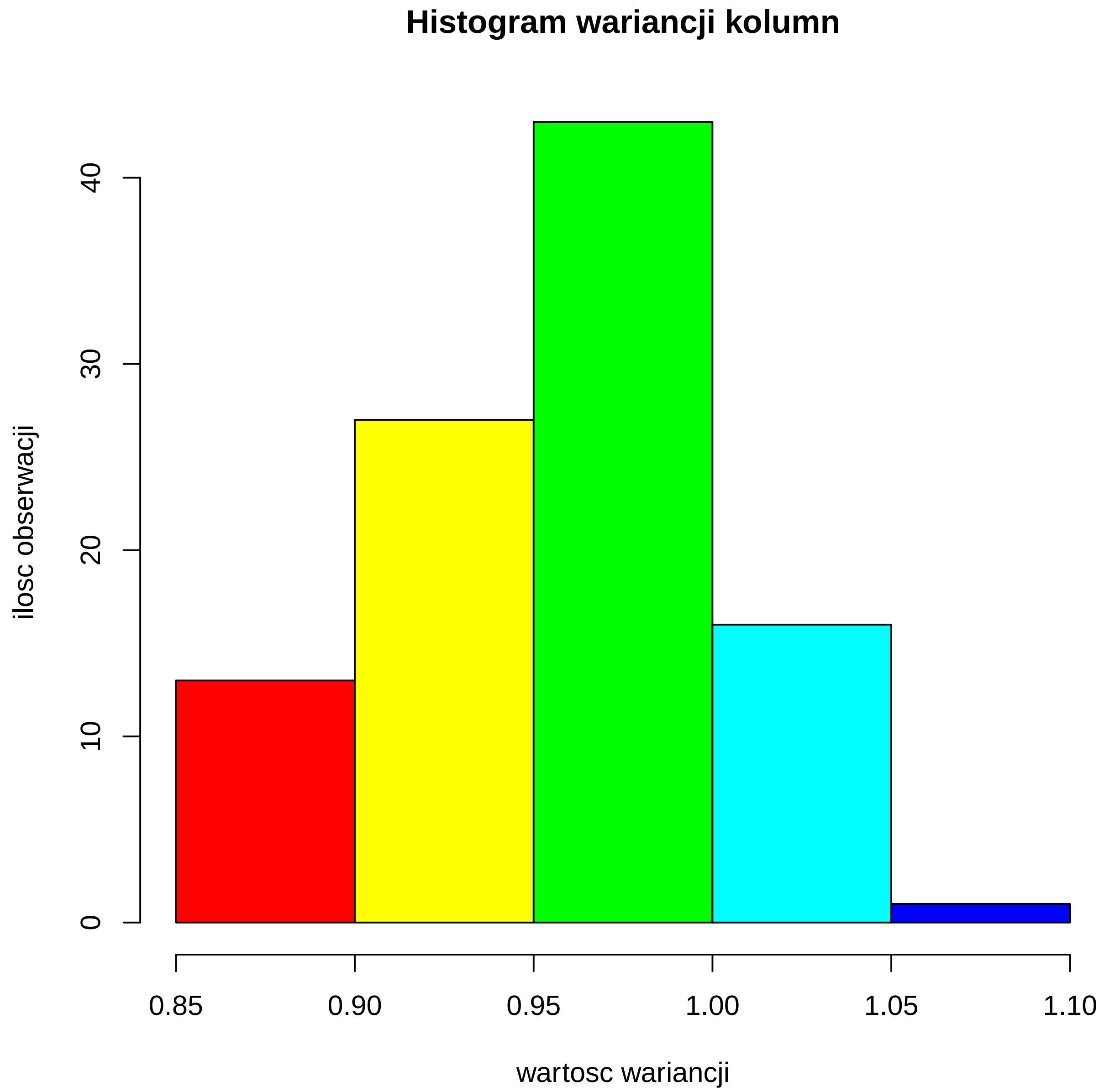
Tak więc ostatecznie okazuje się że szukanym przekształceniem jest
$X = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \cdot \xi + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ gdzie $\xi$ ma rozkład $N(0,I)$
Teraz narysujmy chmurę punktów



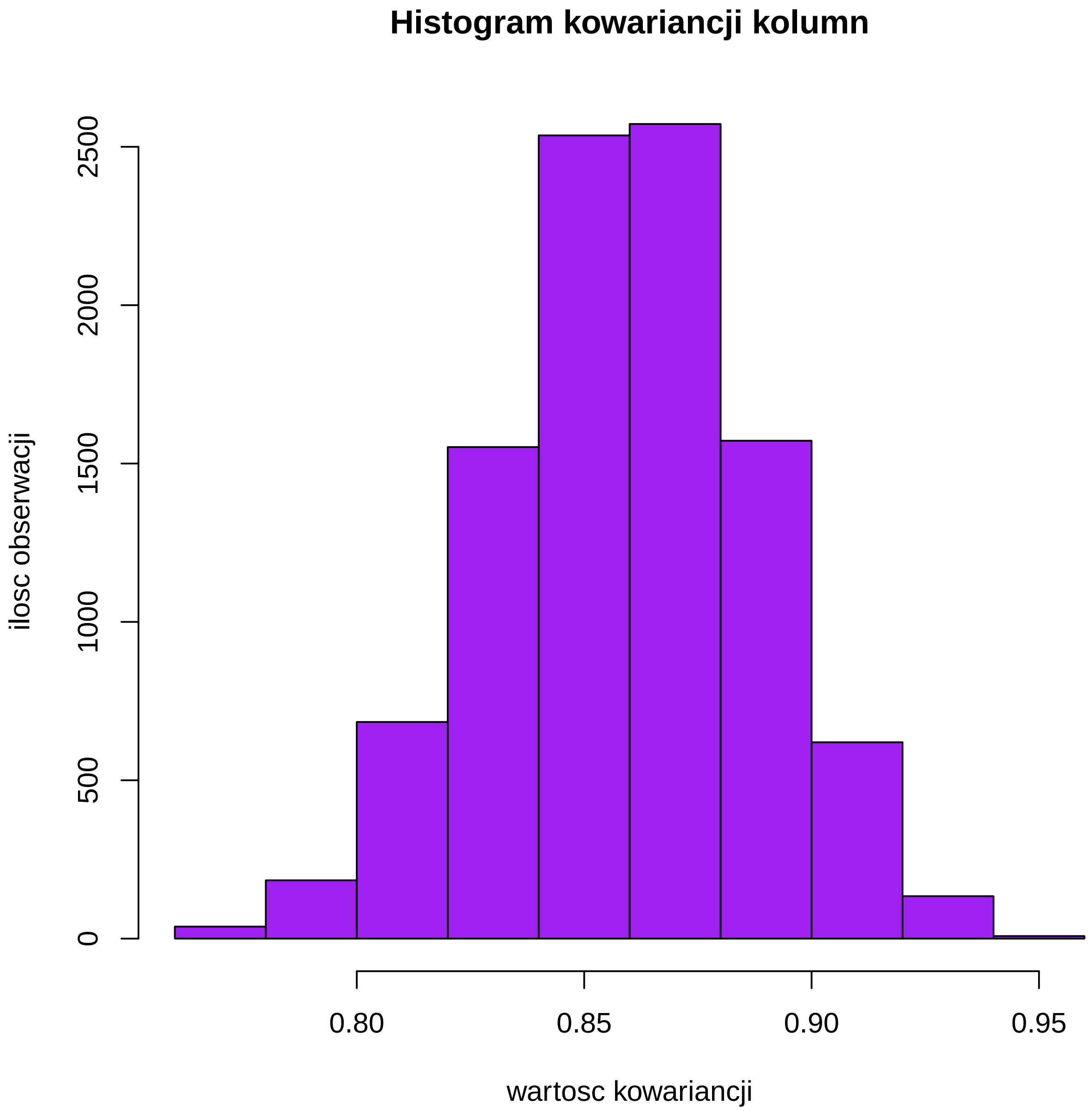

Tym razem zgodnie z macierzą kowariancji $\Sigma$ nie ma żadnej korelacji między zmiennymi $x$ i $y$ oraz wariancja zmiennej $x$ jest wyraźnie wyższa niż zmiennej $y$

# 3. Zadanie 3

Tym razem chcemy wygenerować rozkład 100-wymiarowy o parametrach $N(0,\Sigma)$ gdzie 0 to wektor który ma sto zer a $\Sigma$ ma jedynki na przekątnej a w pozostałych miejscach ma 0.9. Przy tak wielu wymiarach nie można już prosto graficznie przedstawić tego rozkładu tak jak zrobiłem to w przypadku dwuwymiarowym. Więc tym razem jako sprawdzenie posłużą nam histogramy próbkowej wariancji i kowariancji.




Na histogramie widać, że najwięcej obserwacji znajduje się wokół 1 choć przy kolejnych wywołaniach programu wartości trochę się od siebie różnią zwykle w granicach +- 0.1 Warto również policzyć średnią arytmetyczną wariancji poszczególnych kolumn. Wynosi ona 0.9582436 czyli jest zbliżona do jedynki.




Na histogramie widzimy, że najwięcej wartości jest skupionych wokół 0.9 co zgadza się z naszymi wiadomościami teoretycznymi ale tutaj również przy kolejnych wywołaniach mamy do czynienia z pewną niewielką losowością. Średnia arytmetyczna kowariancji między kolumnami wynosi 0.859203 i jest zbliżona do 0.9