

# Raport 2

Aleksander Milach

25 kwietnia 2019

## Zadanie 1

### Podpunkt a

```
##              1      numeracy      anxiety
## 1          46.2198633 -0.24830459 -3.06128454
## numeracy -0.2483046  0.06154567 -0.02381546
## anxiety  -3.0612845 -0.02381546  0.23078671

## [1] 0.2467438
## [1] 0.4751613
## [1] 0.0006631581
## [1] 0.005008419
```

Zgodnie z teorią, wartości na przekątnej macierzy informacji Fishera (czyli estymator macierzy kowariancji estymatorów), różnią się bardzo niewiele od kwadratów wartości odchyłeń standardowych dla zmiennych anxiety i numeracy, obliczonych przez funkcję glm.

### Podpunkt b

```
1-pchisq(M1$null.deviance-deviance(M1),2)
```

```
## [1] 2.343107e-09
```

Wymienioną w zadaniu hipotezę testujemy sprawdzając zgodność różnicy null deviance i residual deviance z rozkładem chi-kwadrat z dwoma stopniami swobody. Obliczona wartość jest dużo mniejsza od 0,05, toteż odrzucamy hipotezę o braku wpływu którejkolwiek ze zmiennych na zmienną odpowiedzi.

### Podpunkt c

```
1-pchisq(M1$deviance,df.residual(M1))
```

```
## [1] 0.9860377
```

Wymienioną w zadaniu hipotezę testujemy sprawdzając zgodność wartości residual deviance z rozkładem chi-kwadrat z n-p stopniami swobody. Obliczona wartość jest dużo większa od 0,05, toteż przyjmujemy hipotezę o rozkładzie danych związanym z modelem.

### Podpunkt d

	10 <sup>-1</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-06</sup>	10 <sup>-8</sup>
Ilosc iteracji	3.0000	4.0000	5.0000	6.0000	6.0000
Est. intercept	12.8901	14.0925	14.2368	14.2386	14.2386
Est. numeracy	0.5376	0.5735	0.5773	0.5774	0.5774

	$10^{-1}$	$10^{-2}$	$10^{-3}$	$10^{-6}$	$10^{-8}$
Est. anxiety	-1.2640	-1.3713	-1.3839	-1.3841	-1.3841

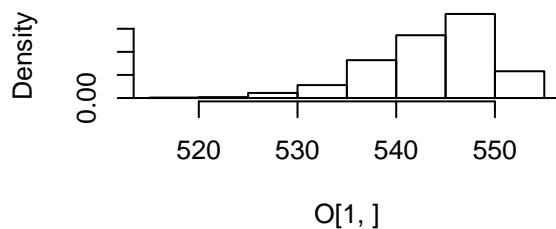
glm.control(epsilon = 1e-8, maxit = 25, trace = FALSE) epsilon - positive convergence tolerance eps; the iterations converge when  $|\text{dev} - \text{dev\_old}| / (|\text{dev}| + 0.1) < \text{eps}$ .

Parametr epsilon ustala warunek, przy którym funkcja glm przestaje przeprowadzać interakcje Fishera. Wartość domyślna to  $\epsilon = 10^{-8}$ .

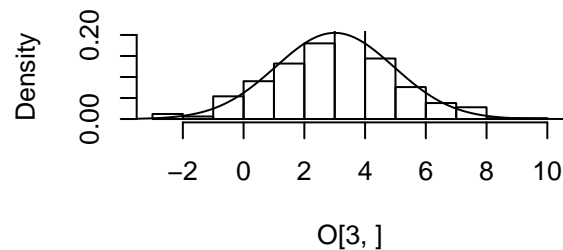
Od wartości epsilon równej  $\epsilon = 10^{-3}$  wartości estymatorów obliczone przez funkcję glm różnią się na trzecim miejscu po przecinku. Pierwsze dwie wartości epsilon dają już zauważalne różnice w wartościach estymatorów, ponieważ przeprowadzają za mało interakcji Fishera. Skoro jednak wartość  $\epsilon = 10^{-3}$  wystarcza do bardzo dobrej estymacji parametrów, to możemy uważać, że aż tak mała wartość domyślna jest nie potrzebna.

## Zadanie 2

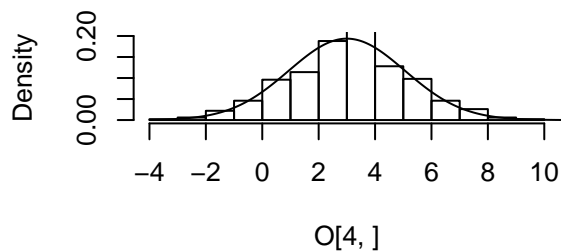
**Histogram of O[1, ]**



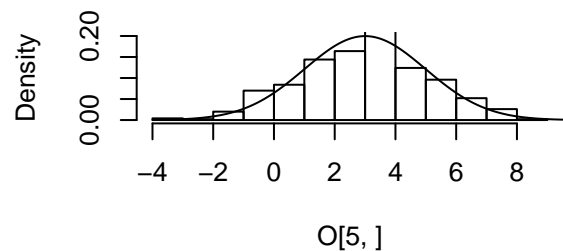
**Estymator beta1 vs N(3,estkwar(1,1))**



**Estymator beta2 vs N(3,estkwar(2,2))**



**Estymator beta3 vs N(3,estkwar(3,3))**

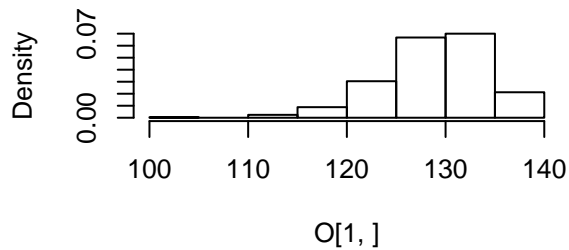


```
## $`Fish(beta)`
##           [,1]
## [1,] 0.6714276
## [2,] 0.6753692
## [3,] 0.7415401
##
## $`Est. cov`
##           [,1]      [,2]      [,3]
## [1,]  4.70772246 -0.1702877 -0.06188142
## [2,] -0.17028769  4.4511798  0.14583958
## [3,] -0.06188142  0.1458396  3.96883890
##
## $`Obc. 1`
## [1] 0.1477947
##
## $`Obc. 2`
```

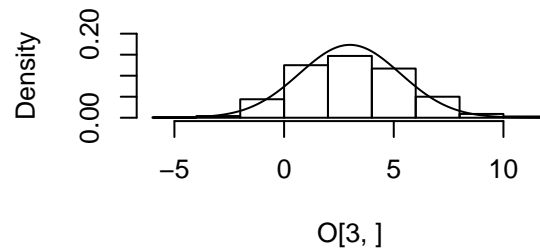
```
## [1] -0.02721017
##
## $`Obc. 3`
## [1] -0.02970444
##
## $`As. cov`
##           X1           X2           X3
## X1  4.81368746 -0.57181947 -0.07033558
## X2 -0.57181947  5.09474347  0.05364947
## X3 -0.07033558  0.05364947  4.59980303
##
## $Roznica
## [1] 0.4015318
```

### Zadanie 3

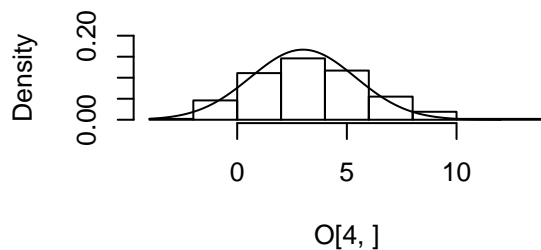
**Histogram of O[1, ]**



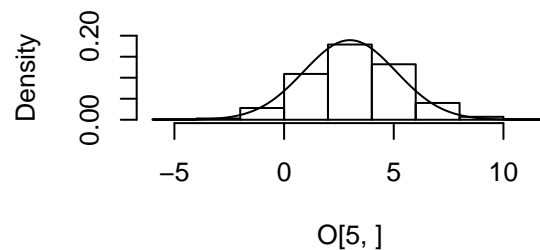
**Estymator beta1 vs N(3,estkowar(1,1))**



**Estymator beta2 vs N(3,estkowar(2,2))**



**Estymator beta3 vs N(3,estkowar(3,3))**



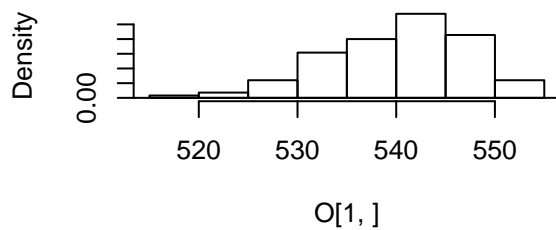
```
## $`Fish(beta)`
##           [,1]
## [1,] 0.6142900
## [2,] 0.6809506
## [3,] 0.8073739
##
## $`Est. cov`
##           [,1] [,2] [,3]
## [1,] 4.4529963 -0.4172750 0.6796264
## [2,] -0.4172750 5.4793862 -0.5881578
## [3,] 0.6796264 -0.5881578 3.6947178
##
## $`Obc. 1`
## [1] 0.2885405
##
## $`Obc. 2`
```

```
## [1] 0.1960707
##
## $`Obc. 3`
## [1] 0.04847063
##
## $`As. cov`
##           X1           X2           X3
## X1  5.2719743 -0.5607776  0.5811534
## X2 -0.5607776  6.3491327 -0.3288940
## X3  0.5811534 -0.3288940  3.7738652
##
## $Roznica
## [1] 0.1435026
```

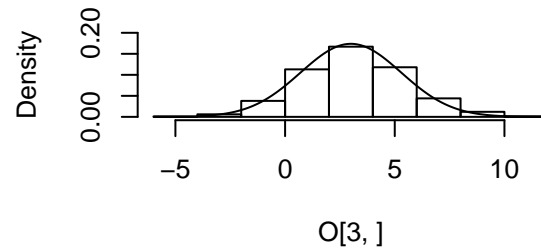
#### Zadanie 4

n=400

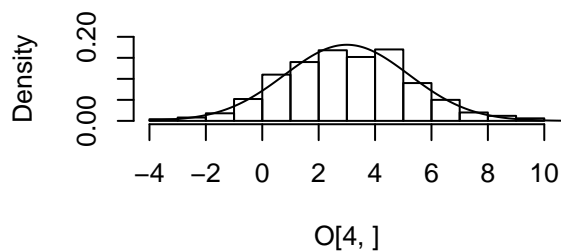
**Histogram of O[1, ]**



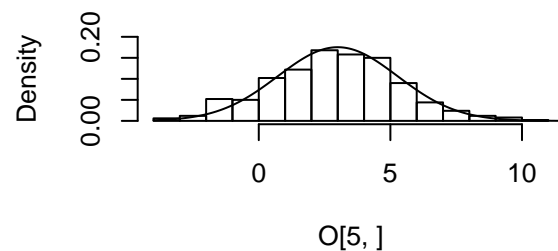
**Estymator beta1 vs N(3,estkowar(1,1))**



**Estymator beta2 vs N(3,estkowar(2,2))**



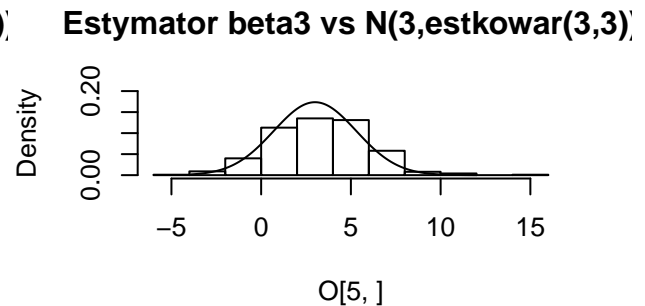
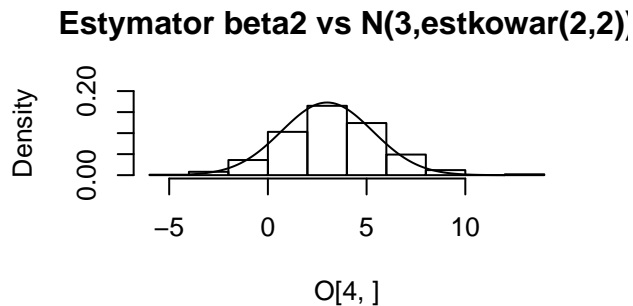
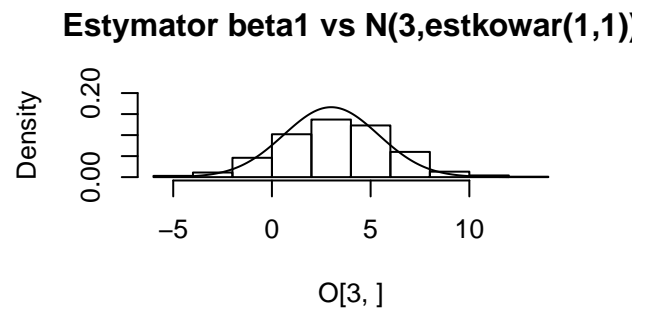
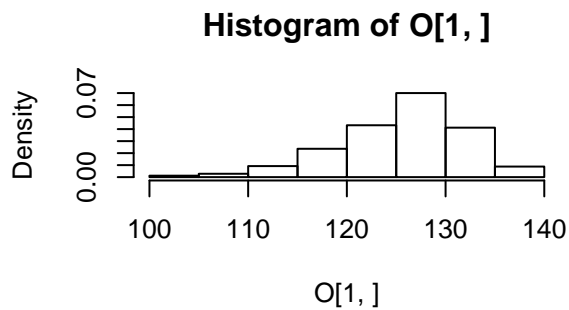
**Estymator beta3 vs N(3,estkowar(3,3))**



```
## $`Fish(beta)`
##           [,1]
## [1,] 1.145313
## [2,] 1.065366
## [3,] 1.089225
##
## $`Est. cov`
##           [,1]      [,2]      [,3]
## [1,] 4.5585243 -0.8753275 -1.182853
## [2,] -0.8753275  4.7882541 -1.008717
## [3,] -1.1828526 -1.0087168  4.984636
##
## $`Obc. 1`
## [1] -0.01376518
```

```
##
## $`Obc. 2`
## [1] 0.1210019
##
## $`Obc. 3`
## [1] -0.03207121
##
## $`As. cov`
##           X1           X2           X3
## X1  4.7566353 -0.6211065 -1.318403
## X2 -0.6211065  5.1512961 -1.294591
## X3 -1.3184025 -1.2945911  5.167226
##
## $Roznica
## [1] 0.2858743
```

**n=100**



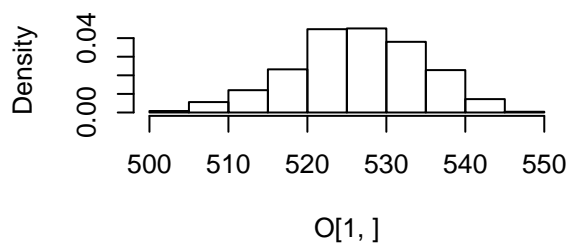
```
## $`Fish(beta)`
##           [,1]
## [1,] 0.7368779
## [2,] 0.7738326
## [3,] 0.8722937
##
## $`Est. cov`
##           [,1]      [,2]      [,3]
## [1,]  7.358477 -1.7163222 -1.2543397
## [2,] -1.716322  6.2922893 -0.6929536
## [3,] -1.254340 -0.6929536  5.1135589
##
## $`Obc. 1`
```

```
## [1] -0.0929682
##
## $`Obc. 2`
## [1] 0.01165608
##
## $`Obc. 3`
## [1] 0.3609026
##
## $`As. cov`
##          X1          X2          X3
## X1  8.540837 -2.2026290 -1.1141752
## X2 -2.202629  7.1379171 -0.8710285
## X3 -1.114175 -0.8710285  5.8207742
##
## $Roznica
## [1] 0.4863068
```

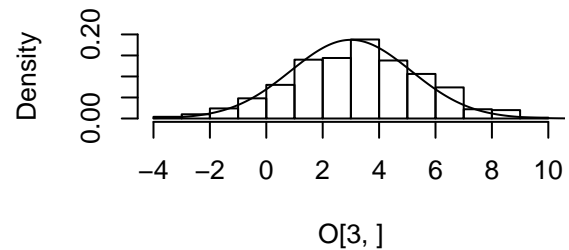
## Zadanie 5

$n=400$

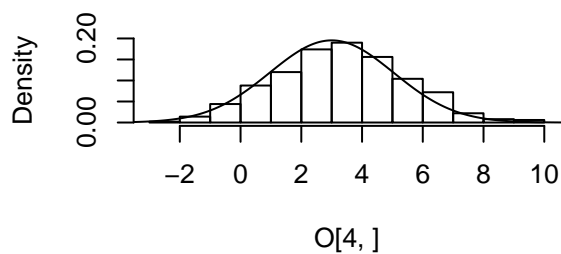
**RSS vs  $\chi^2(n-p)$**



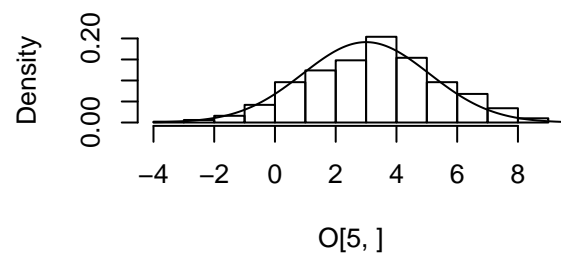
**Estymator beta1 vs  $N(3, \text{estkogar}(1,1))$**



**Estymator beta2 vs  $N(3, \text{estkogar}(2,2))$**



**Estymator beta3 vs  $N(3, \text{estkogar}(3,3))$**



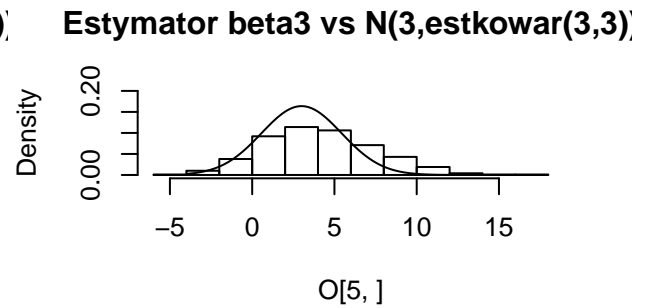
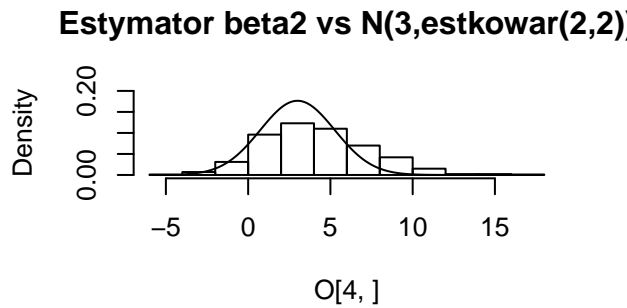
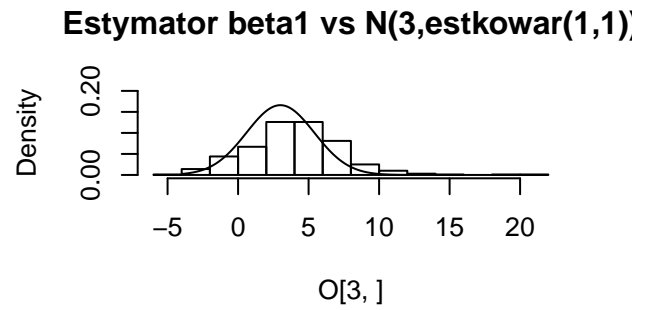
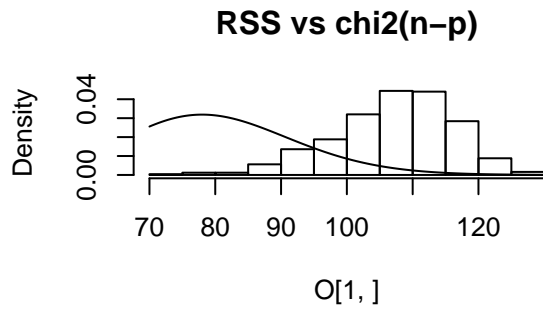
```
## $`Fish(beta)`
##          [,1]
## [1,] 0.625746755
## [2,] 0.792088469
## [3,] 0.803267161
## [4,] -0.042975782
## [5,] -0.013403724
## [6,] 0.023134903
## [7,] 0.006266601
## [8,] 0.126375944
## [9,] 0.110183692
```

```

## [10,] 0.014168242
## [11,] 0.054185882
## [12,] -0.103923361
## [13,] -0.039094773
## [14,] -0.051880884
## [15,] 0.007225017
## [16,] -0.034348256
## [17,] 0.075820767
## [18,] 0.001234694
## [19,] 0.071684942
## [20,] -0.073402619
##
## $`Est. cov`
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 4.54247654 0.3547111 -0.04021037 0.050622816 -0.292752807
## [2,] 0.35471110 4.1651764 -0.45121362 0.205182915 0.156221701
## [3,] -0.04021037 -0.4512136 4.34912608 -0.012429582 0.174775491
## [4,] 0.05062282 0.2051829 -0.01242958 4.573109823 -0.007135652
## [5,] -0.29275281 0.1562217 0.17477549 -0.007135652 4.349328752
##
## $`Obc. 1`
## [1] 0.2104616
##
## $`Obc. 2`
## [1] 0.2884784
##
## $`Obc. 3`
## [1] 0.2499649
##
## $`As. cov`
##           X1      X2      X3      X4      X5
## X1 5.43668945 0.1786541 0.266690910 -0.0727107 -0.052833609
## X2 0.17865405 4.4930684 -0.376647659 0.2957038 0.223437359
## X3 0.26669091 -0.3766477 4.782849600 -0.1236351 0.008266002
## X4 -0.07271070 0.2957038 -0.123635058 5.3102977 0.200613675
## X5 -0.05283361 0.2234374 0.008266002 0.2006137 4.704422658
##
## $Roznica
## [1] 0.9081223

```

n=100



```
## $`Fish(beta)`
##           [,1]
## [1,]  0.548848008
## [2,]  0.890311151
## [3,]  0.707871497
## [4,] -0.195872129
## [5,] -0.171917724
## [6,]  0.032658024
## [7,] -0.030264108
## [8,] -0.098172041
## [9,] -0.003693524
## [10,] -0.039695949
## [11,] -0.030628158
## [12,]  0.209386274
## [13,] -0.116055465
## [14,] -0.107711279
## [15,]  0.007064733
## [16,] -0.237357107
## [17,] -0.033341261
## [18,]  0.220600604
## [19,] -0.020184299
## [20,]  0.060364377
##
## $`Est. cov`
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,]  5.77424114  0.3645738 -0.2118294 -0.09083152 -0.0662732
## [2,]  0.36457378  5.1080335 -0.8629612  0.88919729  0.4391979
## [3,] -0.21182941 -0.8629612  5.9304896  0.21786655  0.4923518
## [4,] -0.09083152  0.8891973  0.2178666  4.01754364  0.4530369
## [5,] -0.06627320  0.4391979  0.4923518  0.45303685  4.4320475
```



```
##
## $`Obc. 1`
## [1] 0.9362041
##
## $`Obc. 2`
## [1] 1.095836
##
## $`Obc. 3`
## [1] 1.158301
##
## $`As. cov`
##           X1           X2           X3           X4           X5
## X1 10.5592904  0.7164264 -0.3346170 -0.2989471  0.2732347
## X2  0.7164264 10.3251518 -0.6695715  2.5564000  1.2708750
## X3 -0.3346170 -0.6695715 11.1296561  0.7236261  1.3369247
## X4 -0.2989471  2.5564000  0.7236261  8.4684785  1.0239530
## X5  0.2732347  1.2708750  1.3369247  1.0239530  8.8075969
##
## $Roznica
## [1] 6.963282
```

W zadaniach 2-5 przeprowadzaliśmy ten sam eksperyment ze zmienionymi danymi. W zadaniach 2-4 wyniki są podobne i dobre. Mamy bardzo dobre przybliżenia histogramów wartości parametrów rozkładami normalnymi, małe wyestymowane obciążenia (co najwyżej 0,4), małe różnice między asymptotyczną a estymowaną macierzą kowariancji estymatorów (co najwyżej 0,9).

Wyniki pogarszają się w zadaniu 5, gdzie wprowadzamy 17 nieistotnych zmiennych, bardziej w przypadku, gdy mamy tylko 100 obserwacji. Obciążenia wynoszą około 1, przez co zauważalne jest gorsze przybliżenie rozkładem normalnym ze średnią 3. Również wariancja jest większa niż wyestymowaliśmy. Różnica między asymptotyczną a estymowaną macierzą kowariancji estymatorów wynosi już prawie 7.

Na żadnym z wykresów RSS nie widać rozkładu asymptotycznego -  $\chi^2(n-p)$ . Wynika to z tego, że jest on asymptotyczny dla  $p$  zbiegającego do  $n$ , a nasze wartości tego parametru, czyli 3 i 20 są dużo mniejsze od wartości  $n - 100$  i  $400$ . Na histogramie RSS dla  $p=20$  i  $n=100$  widać gęstości asymptotycznego rozkładu, jednak jak wcześniej wspomniałem, nie pasuje on do histogramu.

Macierze kowariancji estymatorów, asymptotyczna i estymowana mają rozmiar 20 na 20, przedstawianie ich całych w raporcie jest niepotrzebne, stąd w zadaniu 5 umieściłem tylko jej lewą górną część rozmiaru 5 na 5.