

Projekt zaliczeniowy

Zofia Siwiec, Radosław Różyński

2023-06-07

```
library(dplyr)

##
## Dołączanie pakietu: 'dplyr'
## Następujące obiekty zostały zakryte z 'package:stats':
##
##   filter, lag
## Następujące obiekty zostały zakryte z 'package:base':
##
##   intersect, setdiff, setequal, union
library(ggplot2)
library(tidyr)
library(ResourceSelection)

## ResourceSelection 0.3-5    2019-07-22
library(statmod)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
library("car")
```

Ładowanie wymaganego pakietu: carData

```
##
## Dołączanie pakietu: 'car'
## Następujący obiekt został zakryty z 'package:dplyr':
##
##   recode
```

I zbiór danych

Dane o suwach

Dane dotyczą zakupu suv'a.

```
suvy <- read.csv("suv_data.csv")
```

Nazwa zmiennej	Opis
User.ID	ID użytkownika
Gender	pleć

Nazwa zmiennej	Opis
Age	wiek
EstimatedSalary	przewidywana płaca
Purchased	czy zakupiony został suv (1 – tak, 0 – nie)

Zmiana zmiennej *pleć* na factor:

```
suvy <- suvy %>% mutate(Gender = factor(Gender))
```

Zobaczmy jak prezentują się dane:

```
summary(suvy)
```

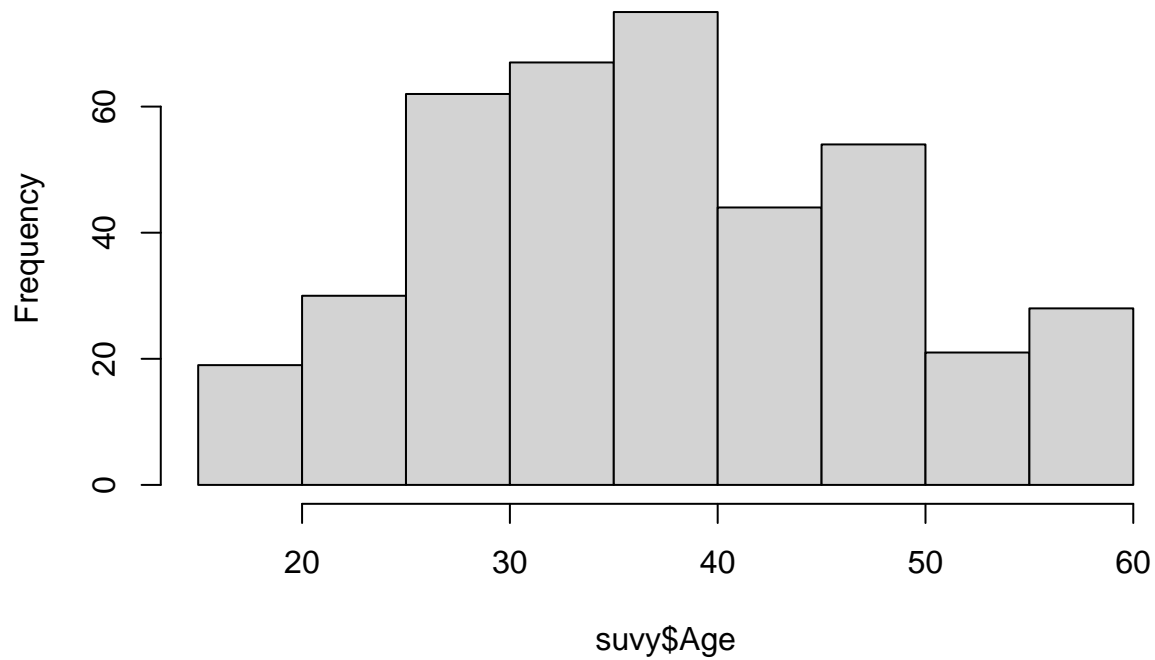
```
##      User.ID      Gender      Age      EstimatedSalary
##  Min.   :15566689  Female:204  Min.   :18.00  Min.   : 15000
## 1st Qu.:15626764  Male  :196  1st Qu.:29.75  1st Qu.: 43000
## Median :15694342           Median :37.00  Median : 70000
## Mean   :15691540           Mean   :37.66  Mean   : 69742
## 3rd Qu.:15750363           3rd Qu.:46.00  3rd Qu.: 88000
## Max.   :15815236           Max.   :60.00  Max.   :150000
##      Purchased
##  Min.   :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean   :0.3575
## 3rd Qu.:1.0000
## Max.   :1.0000
```

Widać, że dane są kompletne. Stosunek płci jest stosunkowo równy. Średnia wieku wygląda autentycznie – przed 18 rokiem życia ciężko inwestować w samochód.

Sprawdźmy jak nasze dane rozkładają się na histogramach:

```
hist(suvy$Age)
```

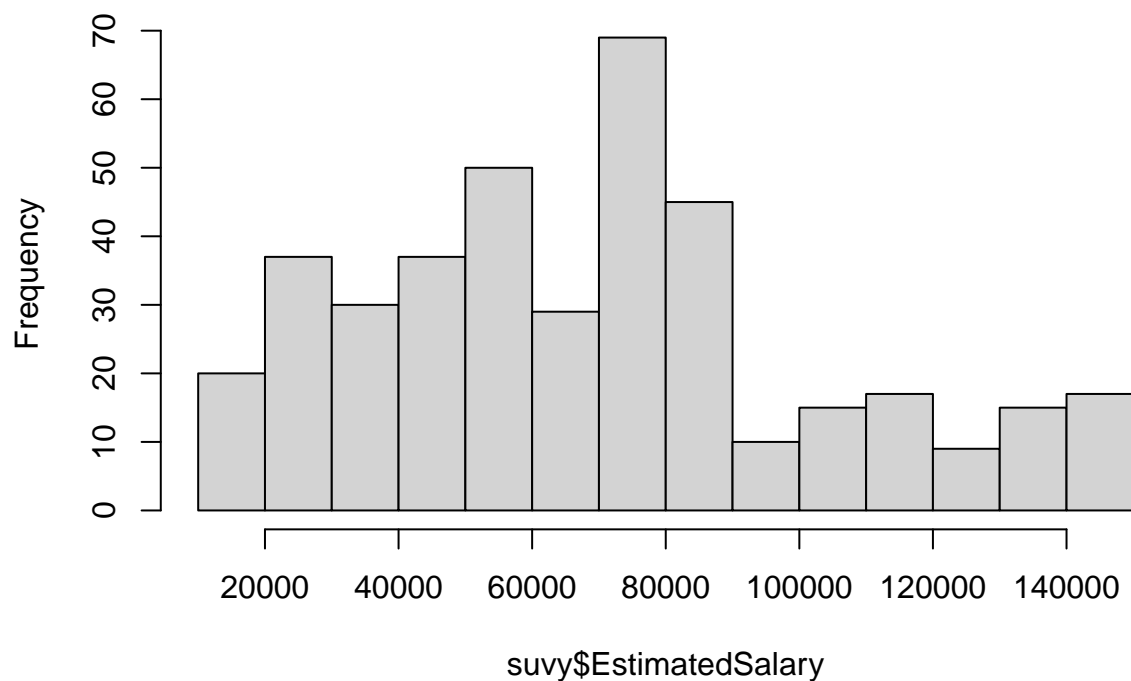
Histogram of suvy\$Age



Histogram zmiennej `age` przypomina lekko zdeformowany rozkład normalny. Zlogarytmowanie zmiennej niczego nie wniosło.

```
hist(suvy$EstimatedSalary)
```

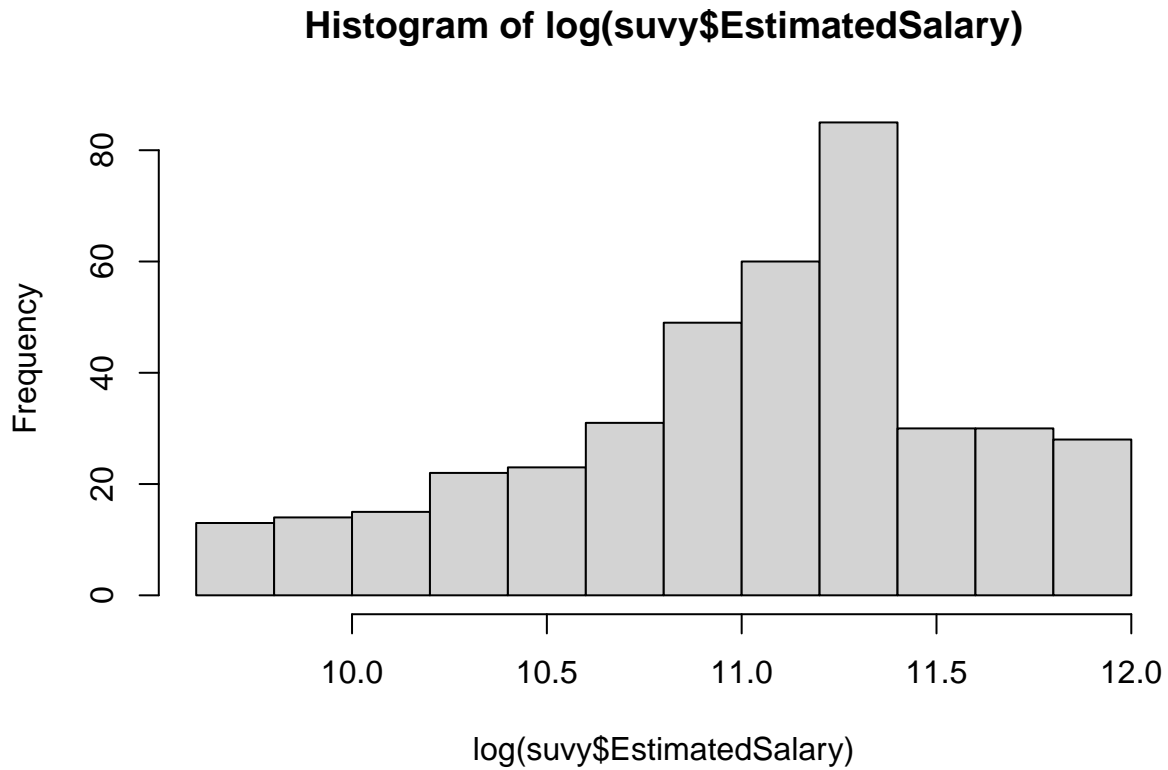
Histogram of suvy\$EstimatedSalary



Histogram średniej płacy nie przypomina żadnego rozkładu. Spróbujmy go zlogarytmować – jest szansa, że

dysproporcja między lewą a prawą częścią wykresu się zminimalizuje.

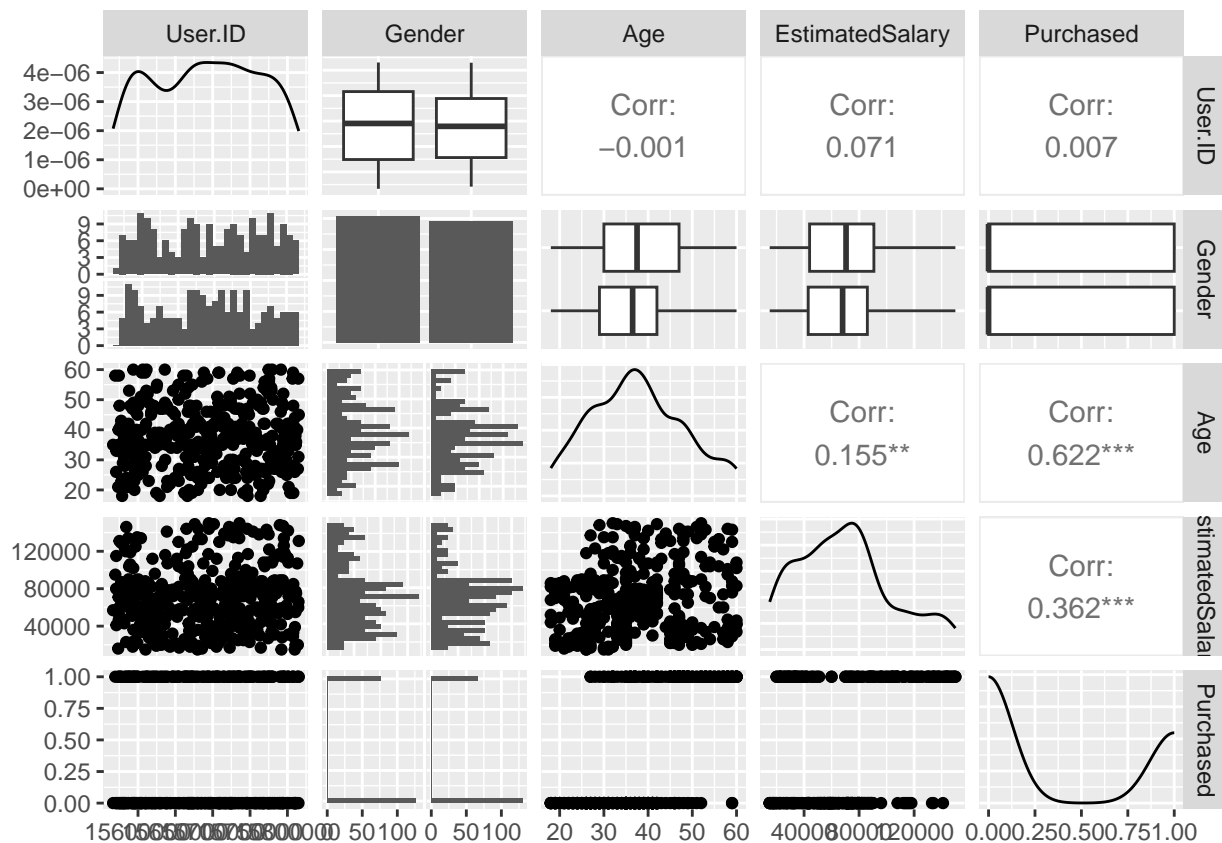
```
hist(log(suvy$EstimatedSalary))
```



Zauważamy większą proporcję. Ale wykres wcale nie jest lepszy.

```
ggpairs(suvy)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

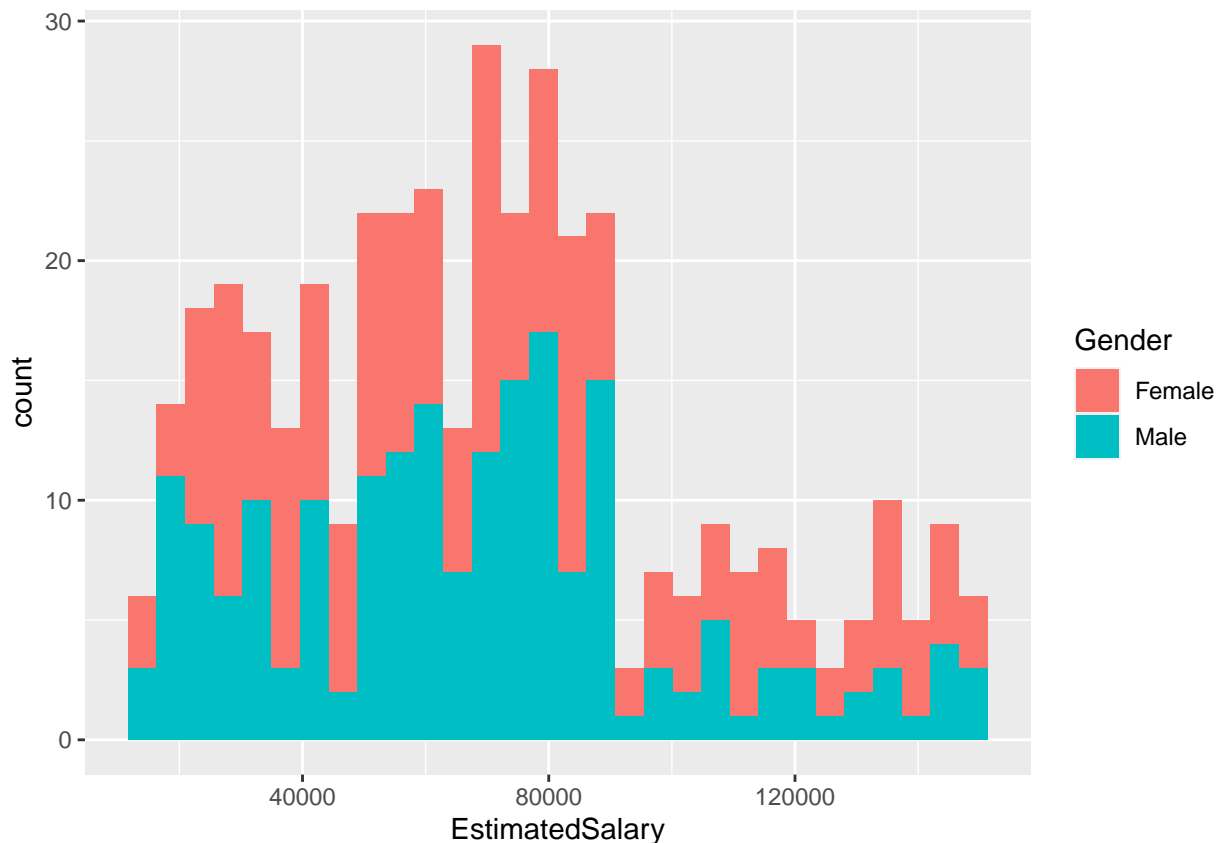


Z wykresu widzimy korelacje pomiędzy zmienną *wiek*, a *wysokość zarobków* – jest to logiczne, ponieważ wraz z wiekiem suma zarobków się powiększa. Widać też korelację pomiędzy *wiekiem* a *wynikiem zakupu* – jest to spowodowane korelacją zmiennej wieku i wysokością zarobków, ale także wraz z wiekiem zakładamy rodziny i jest potrzeba kupienia większego samochodu. Korelację między *wynikiem zakupu* i *wysokością zarobków* można też łatwo wytlumaczyć, im więcej zarabiamy, na tym więcej nas stać.

Średnia *wysokość płacy* w zależności od *płci* jest bardzo podobna – nie zauważamy zjawiska faworyzacji.

```
ggplot(suvy, aes(x = EstimatedSalary, fill = Gender)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Powyższe wykresy pokazały, że płeć nie ma wpływu na płacę. Sprawdźmy jak wygląda sytuacja w nabyciu suv'a.

Postawmy hipotezę badawczą:

Czy płeć ma wpływ na zakup suv'a?

```
modelg <- glm(Purchased ~ Gender, family = binomial, data = suvy)
summary(modelg)
```

```
##
## Call:
## glm(formula = Purchased ~ Gender, family = binomial, data = suvy)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9736  -0.9736  -0.9062   1.3959   1.4754
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.5004     0.1444  -3.464 0.000531 ***
## GenderMale   -0.1775     0.2091  -0.849 0.395858
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 521.57  on 399  degrees of freedom
## Residual deviance: 520.85  on 398  degrees of freedom
```

```
## AIC: 524.85
```

```
##
```

```
## Number of Fisher Scoring iterations: 4
```

Zmienna *pleć* nie jest istotna statystycznie. AIC jest dosyć wysokie, sprawdźmy co się stanie jak dodamy inne zmienne:

```
modelsuvy <- glm(Purchased ~ Gender + EstimatedSalary + Age, family = binomial, data = suvy)
summary(modelsuvy)
```

```
##
```

```
## Call:
```

```
## glm(formula = Purchased ~ Gender + EstimatedSalary + Age, family = binomial,
##      data = suvy)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.9109  -0.5218  -0.1406   0.3662   2.4254
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.278e+01  1.359e+00  -9.405  < 2e-16 ***
## GenderMale    3.338e-01  3.052e-01   1.094    0.274
## EstimatedSalary 3.644e-05  5.473e-06   6.659 2.77e-11 ***
## Age          2.370e-01  2.638e-02   8.984  < 2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
```

```
##      Null deviance: 521.57  on 399  degrees of freedom
```

```
## Residual deviance: 275.84  on 396  degrees of freedom
```

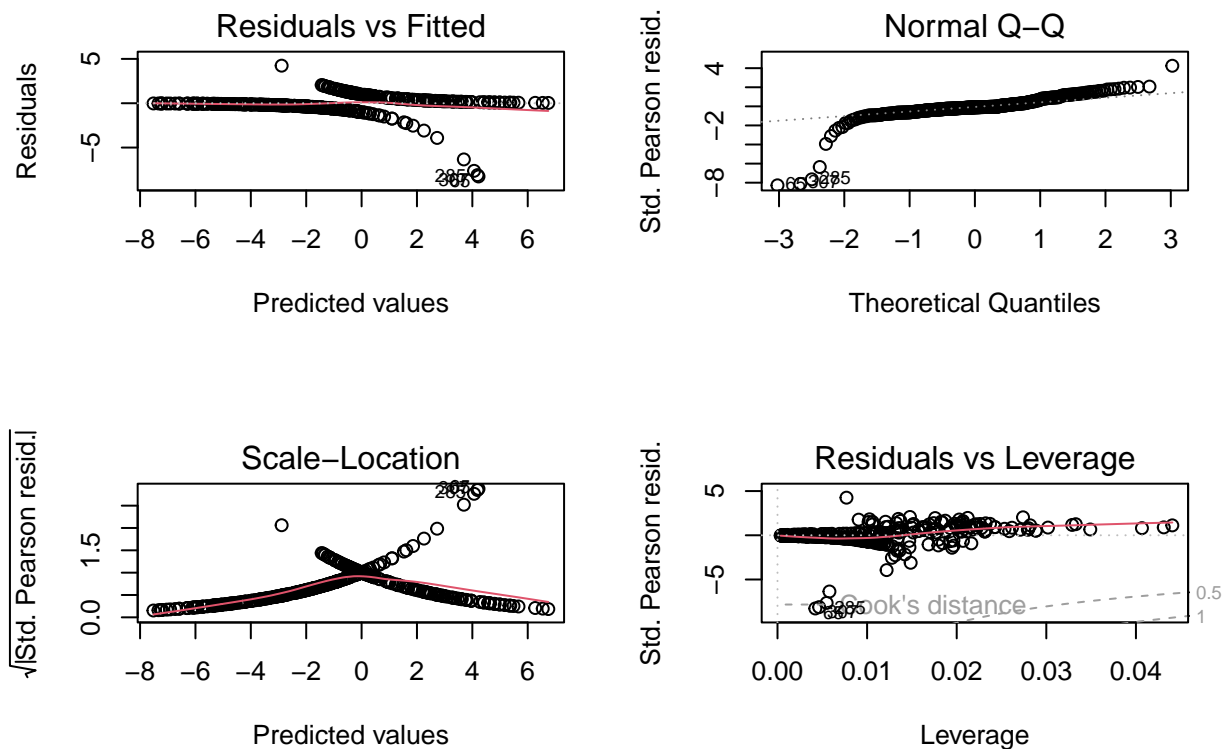
```
## AIC: 283.84
```

```
##
```

```
## Number of Fisher Scoring iterations: 6
```

Na samym początku przetestujmy model bez zlogarytmowanej zmiennej. Jak zauważamy, *pleć* wciąż nie jest zmienną istotną statystycznie. Jest to dosyć dziwne, ponieważ w życiu codziennym przeważnie to mężczyźni siadają za kierownicę większych pojazdów – widok kobiet jest o wiele mniej pospolity. Oczywiście zakładamy, że skoro osoba zakupiła pojazd to będzie nim jeździć. Istotność wieku i średniej płacy jest jasna. AIC, jednak jest mniejsze niż w powyższym modelu co wskazuje na lepsze dopasowanie niż model zerowy.

```
par(mfrow=c(2, 2))
plot(modelsuvy)
```



Wykresy pozostawiają wiele do życzenia, co może być spowodowane istnieniem innych zmiennych, które mają istotny wpływ na nabycie suv'a.

Sprawdźmy jak wygląda model, gdy zlogarytmujemy *średni zarobek*:

```
modelsuvy_log <- glm(Purchased ~ log(EstimatedSalary) + Gender + Age , family = binomial, data = suvy)
summary(modelsuvy_log)
```

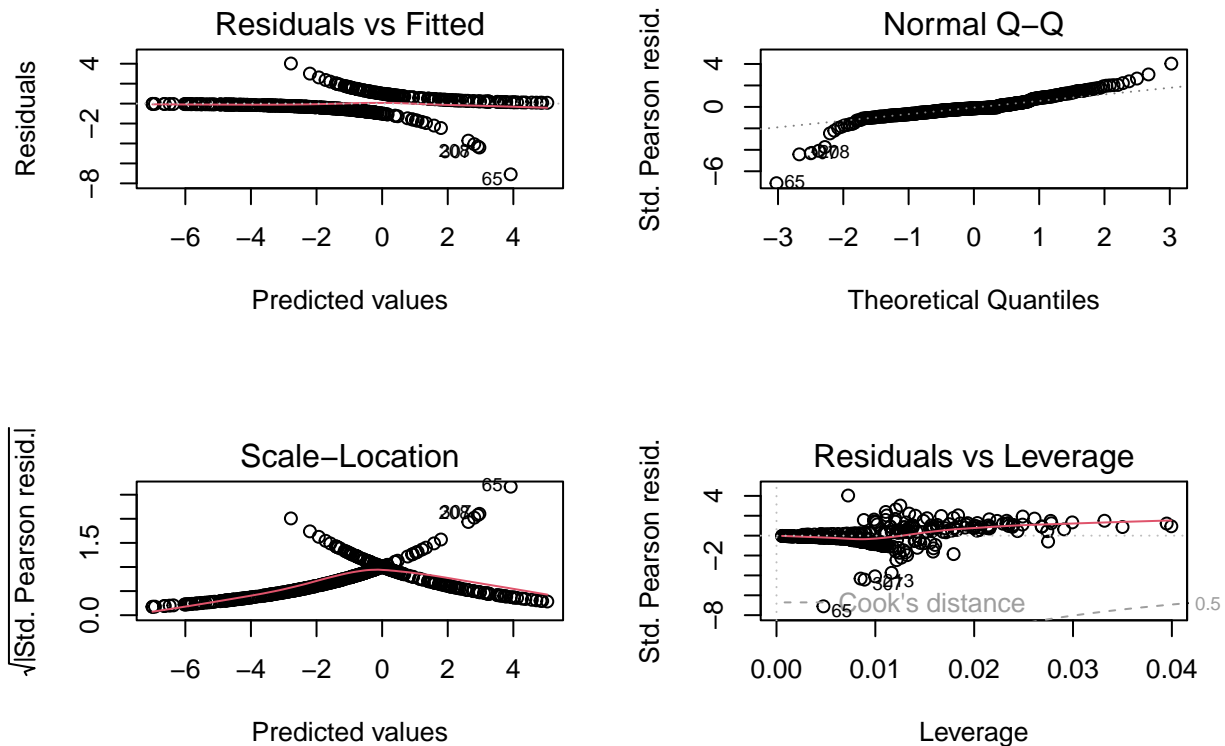
```
##
## Call:
## glm(formula = Purchased ~ log(EstimatedSalary) + Gender + Age,
##      family = binomial, data = suvy)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8064  -0.6339  -0.1800   0.4831   2.3840
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -27.33105     4.07244  -6.711 1.93e-11 ***
## log(EstimatedSalary)  1.62972     0.32052   5.085 3.68e-07 ***
## GenderMale         0.22976     0.28782   0.798  0.425
## Age                0.21678     0.02349   9.227 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 521.57  on 399  degrees of freedom
## Residual deviance: 305.25  on 396  degrees of freedom
## AIC: 313.25
```



```
##
## Number of Fisher Scoring iterations: 6
```

Wciąż rezultat bez zmian – płeć nie ma znaczenia w zakupie suv’a. AIC jest troszeczkę wyższe niż w modelu wyżej – zostaniemy raczej przy powyższym jako ostatecznym. Zobaczmy jeszcze wykresy:

```
par(mfrow=c(2, 2))
plot(modelsuvy_log)
```



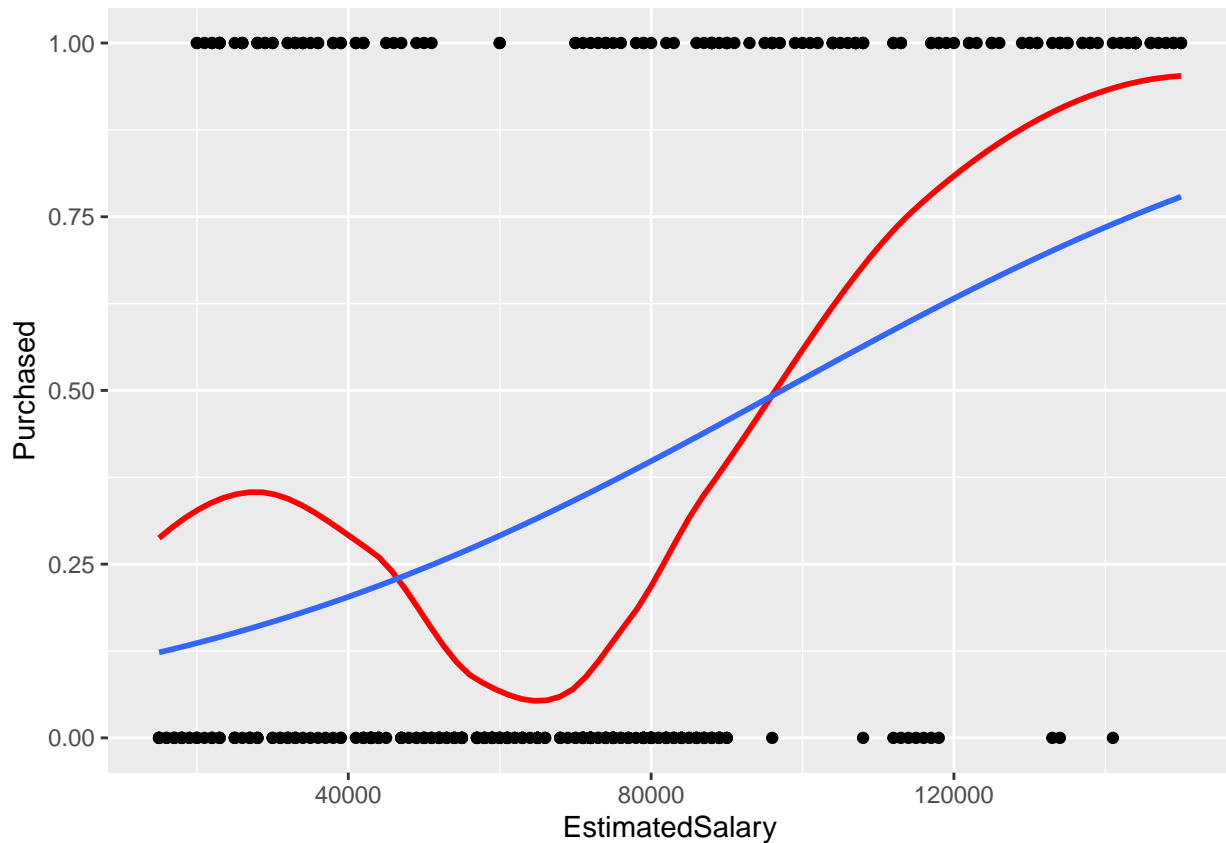
Prezentują się tak samo słabo.

Obalamy naszą hipotezę, że płeć ma wpływ na zakup suv’a. Jesteśmy dosyć zdziwieni finałem, ponieważ jak wspomnieliśmy wcześniej jednak mężczyźni przeważnie jeżdżą większymi samochodami, a kobiety małymi aby nie mieć problemu z parkowaniem. Nawet jeśli oboje są w samochodzie to przeważnie za kółkiem częściej występuje mężczyzna.

Zobaczmy jeszcze wykresy dopasowania zmiennych:

```
ggplot(suvy, aes(x = EstimatedSalary, y = Purchased)) + geom_point() + geom_smooth(method = "loess", col = "red", se = FALSE)
geom_smooth(method = "glm", method.args = list(family = "binomial"), se = FALSE)

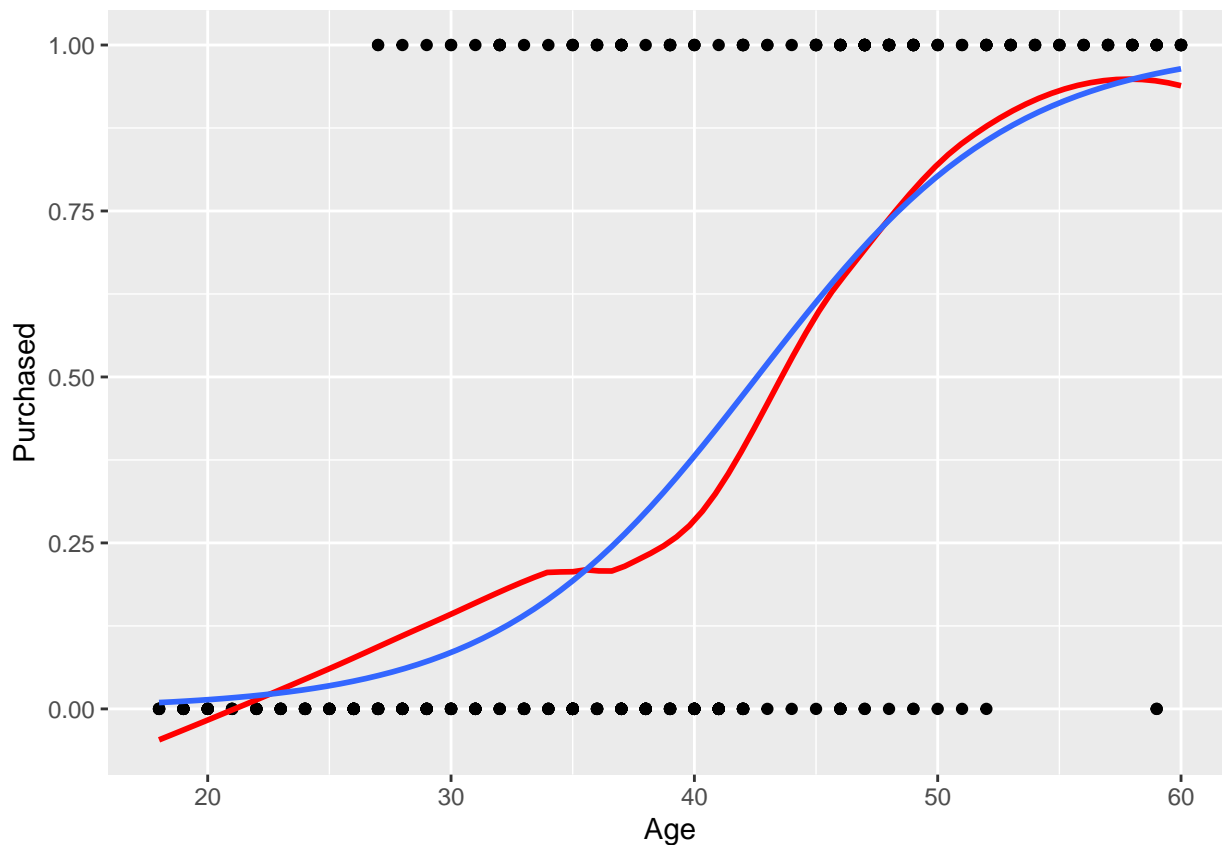
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



Zmienna *wysokość zarobku* sama nie objaśnia dobrze *prawdopodobieństwa zakupu*, z wykresu wynika że nawet przy zerowych zarobkach istnieje prawdopodobieństwo zakupu suv'a.

```
ggplot(suvy, aes(x = Age, y = Purchased)) + geom_point() + geom_smooth(method = "loess", color = "red",
  geom_smooth(method = "glm", method.args = list(family = "binomial"), se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



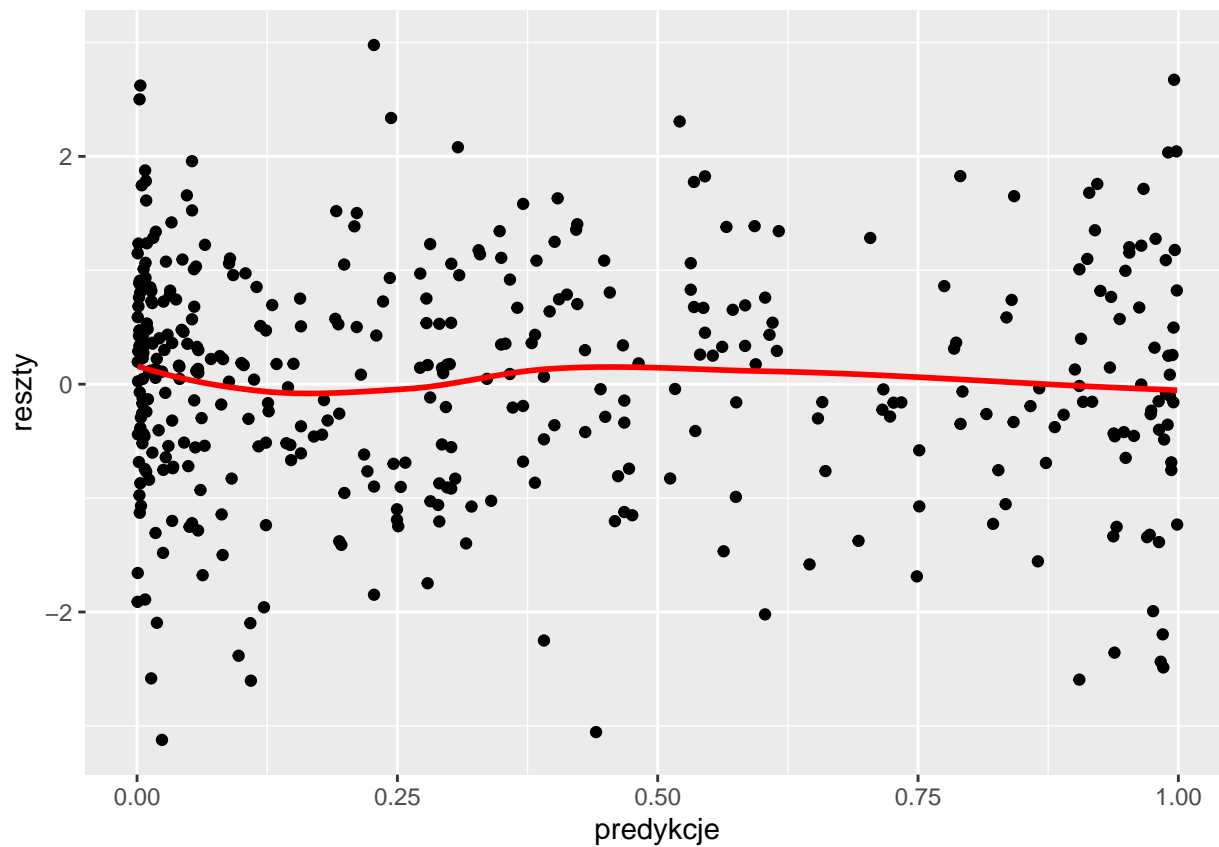
Przy zmiennej *wiek* zauważamy większe dopasowanie. I potwierdzamy, że starsze osoby częściej kupują suv'y (jest większe prawdopodobieństwo).

Gdy zmienna obserwowana jest binarna, nie możemy oczekiwać równomiernego rozłożenia reszt - zalecane jest stosowanie reszt kwantylowych.

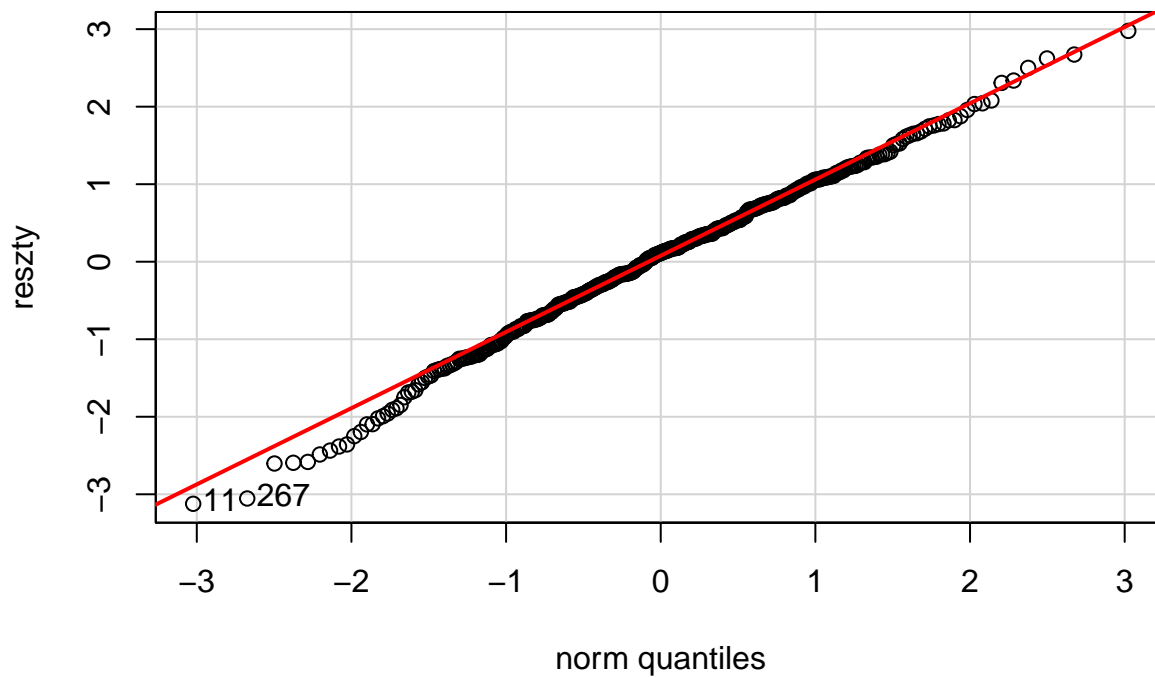
Reszty kwantylowe względem predykcji:

```
predykcje <- predict(modelsuvy, type = "response")
reszty <- qresid(modelsuvy, dispersion = NULL)
df = data.frame(predykcje, reszty)

ggplot(df, aes(x = predykcje, y = reszty)) + geom_point() + geom_smooth(method = "loess", color = "red",
## `geom_smooth()` using formula = 'y ~ x'
```



```
qqPlot(reszty, envelope = FALSE, col.lines = 'red')
```



```
## [1] 11 267
```

Reszty kwantylowe wskazują na w miarę dobrze dopasowany model.

II zbiór danych

Kobiety

Dane dotyczą kobiet.

```
woman <- read.csv("data.csv.xls")
```

Nazwa zmiennej	Opis
mnthborn	month woman born
yearborn	year woman born
age	age in years
electric	1 if has electricity
radio	1 if has radio
tv	1 if has tv
bicycle	1 if has bicycle
educ	years of education
ceb	children ever born
agefbrth	age at first birth
children	number of living children
knowmeth	1 if know about birth control
usemeth	1 if ever use birth control
monthfm	month of first marriage
yearfm	year of first marriage
agefm	age at first marriage
idlnchld	'ideal' number of children
heduc	husband's years of education
agesq	age^2
urban	1 if live in urban area
urb_educ	urban*educ
spirit	1 if religion == spirit
protest	1 if religion == protestant
catholic	1 if religion == catholic
frsthalf	1 if mnthborn <= 6
educ0	1 if educ == 0
evermarr	1 if ever married

Zmieniamy zmienne, których będziemy używać na factory:

```
woman$educ <- as.factor(woman$educ)
woman$usemeth <- as.factor(woman$usemeth)
woman$urban <- as.factor(woman$urban)
```

Zobaczmy jak się prezentują dane:

```
summary(woman)
```

```
##          X          mnthborn          yearborn          age
## Min.      : 1      Min.      : 1.000      Min.      :38.00      Min.      :15.00
## 1st Qu.:1091      1st Qu.: 3.000      1st Qu.:55.00      1st Qu.:20.00
## Median :2181      Median : 6.000      Median :62.00      Median :26.00
## Mean      :2181      Mean      : 6.331      Mean      :60.43      Mean      :27.41
## 3rd Qu.:3271      3rd Qu.: 9.000      3rd Qu.:68.00      3rd Qu.:33.00
## Max.      :4361      Max.      :12.000      Max.      :73.00      Max.      :49.00
##
##      electric          radio          tv          bicycle
## Min.      :0.0000      Min.      :0.0000      Min.      :0.00000      Min.      :0.0000
```

```

## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000
## Median :0.0000 Median :1.0000 Median :0.00000 Median :0.0000
## Mean :0.1402 Mean :0.7018 Mean :0.09291 Mean :0.2758
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.00000 Max. :1.0000
## NA's :3 NA's :2 NA's :2 NA's :3
## educ ceb agefbrth children
## 7 :1162 Min. : 0.000 Min. :10.00 Min. : 0.000
## 0 : 906 1st Qu.: 1.000 1st Qu.:17.00 1st Qu.: 0.000
## 10 : 527 Median : 2.000 Median :19.00 Median : 2.000
## 6 : 298 Mean : 2.442 Mean :19.01 Mean : 2.268
## 5 : 234 3rd Qu.: 4.000 3rd Qu.:20.00 3rd Qu.: 4.000
## 9 : 232 Max. :13.000 Max. :38.00 Max. :13.000
## (Other):1002 NA's :1088
## knowmeth usemeth monthfm yearfm agefm
## Min. :0.0000 0 :1812 Min. : 1.00 Min. :50.00 Min. :10.00
## 1st Qu.:1.0000 1 :2478 1st Qu.: 3.00 1st Qu.:72.00 1st Qu.:17.00
## Median :1.0000 NA's: 71 Median : 6.00 Median :78.00 Median :20.00
## Mean :0.9633 Mean : 6.27 Mean :76.91 Mean :20.69
## 3rd Qu.:1.0000 3rd Qu.: 9.00 3rd Qu.:83.00 3rd Qu.:23.00
## Max. :1.0000 Max. :12.00 Max. :88.00 Max. :46.00
## NA's :7 NA's :2282 NA's :2282 NA's :2282
## idlnchld heduc agesq urban urb_educ
## Min. : 0.000 Min. : 0.000 Min. : 225.0 0:2108 Min. : 0.000
## 1st Qu.: 3.000 1st Qu.: 0.000 1st Qu.: 400.0 1:2253 1st Qu.: 0.000
## Median : 4.000 Median : 6.000 Median : 676.0 Median : 0.000
## Mean : 4.616 Mean : 5.145 Mean : 826.5 Mean : 3.469
## 3rd Qu.: 6.000 3rd Qu.: 8.000 3rd Qu.:1089.0 3rd Qu.: 7.000
## Max. :20.000 Max. :20.000 Max. :2401.0 Max. :20.000
## NA's :120 NA's :2405
## spirit protest catholic frsthalf
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000 Median :0.0000 Median :1.0000
## Mean :0.4222 Mean :0.2277 Mean :0.1025 Mean :0.5405
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
##
## educ0 evermarr
## Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000
## Mean :0.2078 Mean :0.4767
## 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000
##

```

Na pierwszy rzut oka widać, że są małe braki w danych np. zmienne *electric*, *radio*, *tv* oraz większe braki w *monthfm* – jest to zmienna mówiąca o miesiącu wzięcia pierwszego ślubu, braki są wyjaśnione ponieważ nie każdy ma ślubne dzieci. Tak samo *agefbrth*, braki są wytłumaczalne, ponieważ jeśli nie urodziła dziecka to występuje brak.

Sprawdźmy, czy edukacja i miejsce zamieszkania ma wpływ na liczbę dzieci

Stworzymy model poissonowski, którym będziemy przewidywać liczbę dzieci w zależności od innych zmiennych.

```

poisson.model1 <- glm(children ~ educ + age + agesq +
                      agefbrth + urban +
                      usemeth,
                      data = woman, family = poisson())

summary(poisson.model1)

##
## Call:
## glm(formula = children ~ educ + age + agesq + agefbrth + urban +
##      usemeth, family = poisson(), data = woman)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7761  -0.4413  -0.0274   0.3999   3.1070
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.9751514  0.1876937 -10.523  < 2e-16 ***
## educ1        0.0574635  0.0692067   0.830  0.40636
## educ2        0.0260581  0.0531053   0.491  0.62365
## educ3       -0.0003254  0.0493524  -0.007  0.99474
## educ4       -0.0263256  0.0427196  -0.616  0.53774
## educ5       -0.0609914  0.0446180  -1.367  0.17164
## educ6       -0.0666214  0.0432387  -1.541  0.12337
## educ7       -0.0989587  0.0316636  -3.125  0.00178 **
## educ8       -0.1119873  0.0735882  -1.522  0.12806
## educ9       -0.1469922  0.0681974  -2.155  0.03113 *
## educ10      -0.2245303  0.0450166  -4.988 6.11e-07 ***
## educ11      -0.2625023  0.1686717  -1.556  0.11964
## educ12      -0.3480960  0.0669179  -5.202 1.97e-07 ***
## educ13      -0.3770216  0.1619621  -2.328  0.01992 *
## educ14      -0.3443654  0.1207268  -2.852  0.00434 **
## educ15      -0.3958123  0.1442296  -2.744  0.00606 **
## educ16      -0.4177031  0.2023985  -2.064  0.03904 *
## educ17      -0.5402623  0.2440344  -2.214  0.02684 *
## educ18      -0.4006539  0.4096417  -0.978  0.32804
## educ19      -0.2936270  0.4102578  -0.716  0.47417
## educ20      -0.2479171  0.7088112  -0.350  0.72652
## age          0.2090824  0.0108904  19.199  < 2e-16 ***
## agesq       -0.0022564  0.0001593 -14.163  < 2e-16 ***
## agefbrth    -0.0586068  0.0036300 -16.145  < 2e-16 ***
## urban1      -0.1151231  0.0212467  -5.418 6.01e-08 ***
## usemeth1     0.1424939  0.0239159   5.958 2.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 4281.0  on 3212  degrees of freedom
## Residual deviance: 1670.2  on 3187  degrees of freedom
##      (1148 obserwacji zostało skasowanych z uwagi na braki w nich zawarte)
## AIC: 10628
##

```

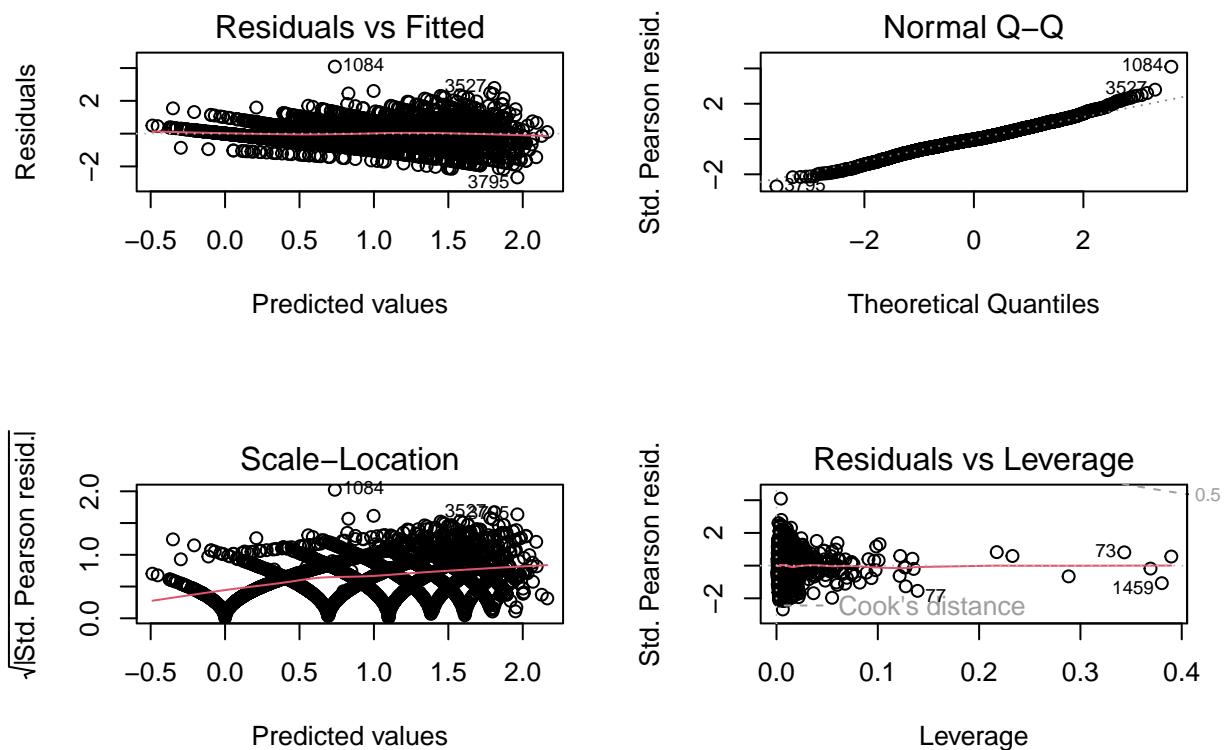
```
## Number of Fisher Scoring iterations: 4
```

Dużą istotność wykazują zmienne *educ10*, *educ12*, *age*, *agesq*, *agefbrth*, *urban* oraz *usemeth*. Zmienna **educ__** mówi nam ile kobieta spędziła lat na edukacji – jak widać z modelu są to istotne dane dopiero w dziesiątkach, czyli mamy doczynienia z młodymi mamami. Współczynniki są ujemne, co sugeruje że im kobieta się dłużej uczyła, tym mniej ma dzieci. Wiek analogicznie też gra rolę. Miejsce zamieszkania, też jest istotne statystycznie – jak mieszka w mieście to ma mniej dzieci niż kobiety żyjące na obrzeżach. Największym zaskoczeniem jest zmienna *usemeth*, z naszego modelu wynika, że jeśli kobieta stosuje antykoncepcję to ma więcej dzieci (dodatni współczynnik). Widzimy też, że powyższy model ma dewiancję resztową niższą niż liczba stopni swobody.

```
par(mfrow=c(2, 2))  
plot(poisson.model1)
```

```
## Warning: brak wykreślalnych obserwacji z jednej transmisji:
```

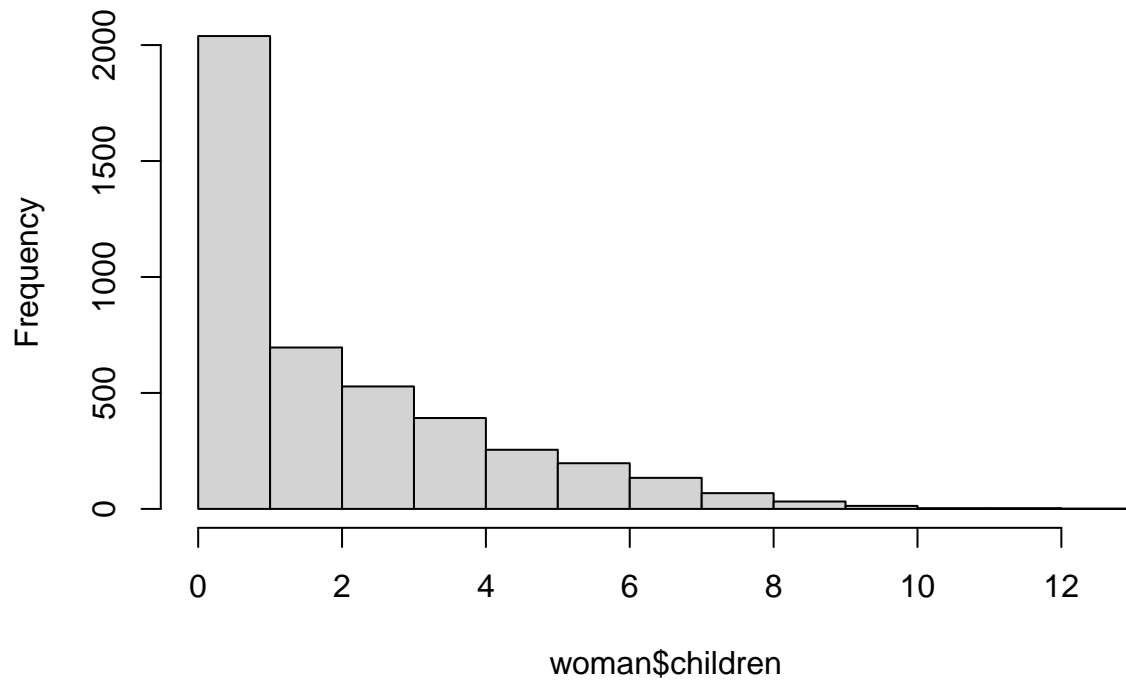
```
## 84
```



Wykresy prócz Scale-Location wyglądają dobrze.

```
hist(woman$children)
```


Histogram of woman\$children



Wykres objaśnianej zmiennej wygląda na rozkład Poissona.

```
mean(woman$children)
```

```
## [1] 2.267828
```

```
var(woman$children)
```

```
## [1] 4.937426
```

Natomiast wariancja jest za duża w porównaniu do średniej. Zastosujemy więc model quasipoissonowski.

```
quasipoisson.model <- glm(children ~ educ + age + agesq +  
  agefbrth + urban +  
  usemeth,  
  data = woman, family = quasipoisson())
```

```
summary(quasipoisson.model)
```

```
##  
## Call:  
## glm(formula = children ~ educ + age + agesq + agefbrth + urban +  
##   usemeth, family = quasipoisson(), data = woman)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -3.7761  -0.4413  -0.0274   0.3999   3.1070   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -1.9751514  0.1299685 -15.197  < 2e-16 ***  
## educ1       0.0574635  0.0479221   1.199  0.230578
```

```

## educ2      0.0260581  0.0367728   0.709 0.478609
## educ3     -0.0003254  0.0341741  -0.010 0.992403
## educ4     -0.0263256  0.0295812  -0.890 0.373563
## educ5     -0.0609914  0.0308957  -1.974 0.048456 *
## educ6     -0.0666214  0.0299406  -2.225 0.026143 *
## educ7     -0.0989587  0.0219255  -4.513 6.61e-06 ***
## educ8     -0.1119873  0.0509561  -2.198 0.028041 *
## educ9     -0.1469922  0.0472232  -3.113 0.001870 **
## educ10    -0.2245303  0.0311718  -7.203 7.32e-13 ***
## educ11    -0.2625023  0.1167967  -2.248 0.024675 *
## educ12    -0.3480960  0.0463373  -7.512 7.51e-14 ***
## educ13    -0.3770216  0.1121506  -3.362 0.000784 ***
## educ14    -0.3443654  0.0835973  -4.119 3.90e-05 ***
## educ15    -0.3958123  0.0998717  -3.963 7.56e-05 ***
## educ16    -0.4177031  0.1401508  -2.980 0.002901 **
## educ17    -0.5402623  0.1689816  -3.197 0.001401 **
## educ18    -0.4006539  0.2836563  -1.412 0.157911
## educ19    -0.2936270  0.2840829  -1.034 0.301403
## educ20    -0.2479171  0.4908161  -0.505 0.613515
## age       0.2090824  0.0075411  27.726 < 2e-16 ***
## agesq     -0.0022564  0.0001103  -20.453 < 2e-16 ***
## agefbrth  -0.0586068  0.0025136  -23.316 < 2e-16 ***
## urban1    -0.1151231  0.0147123  -7.825 6.85e-15 ***
## usemeth1   0.1424939  0.0165606   8.604 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.4794866)
##
##      Null deviance: 4281.0  on 3212  degrees of freedom
## Residual deviance: 1670.2  on 3187  degrees of freedom
##      (1148 obserwacji zostało skasowanych z uwagi na braki w nich zawarte)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4

```

Po zmianie modelu kilka czynników zyskało na istotności.

```

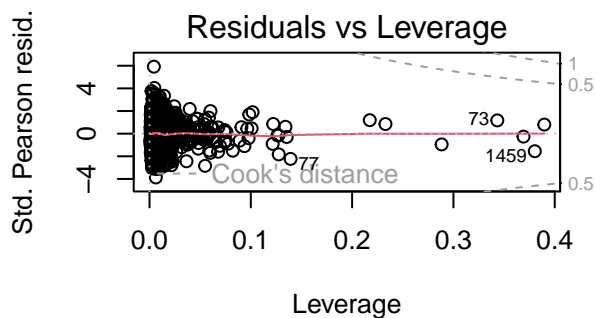
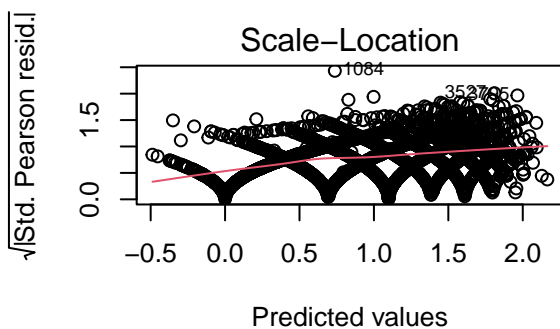
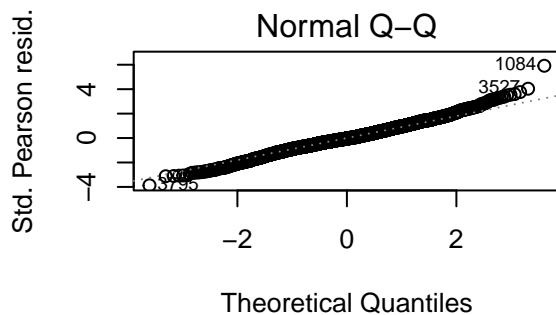
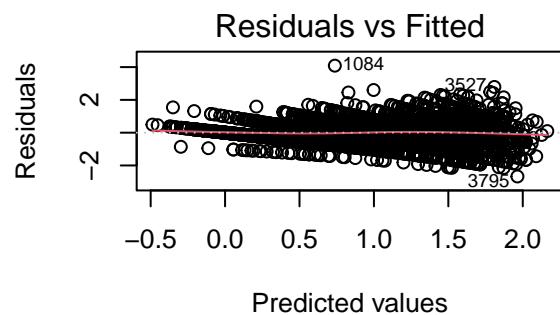
par(mfrow=c(2, 2))
plot(quasipoisson.model)

```

```

## Warning: brak wykreślalnych obserwacji z jednej transmisji:
##      84

```



Wykresy pozostają bez zmian.

Hipoteza została potwierdzona: miejsce zamieszkania oraz długość edukacji mają istotny wpływ na liczbę dzieci. Wpływ jest ujemny, czyli im dłużej się uczy, tym ma mniej dzieci. Identycznie jest w przypadku miejsca zamieszkania, kobiety mieszkające w miastach mają statystycznie mniej dzieci niż kobiety żyjące na wsi.