Projekt ***„Nowa oferta edukacyjna Uniwersytetu Wrocławskiego odpowiedzią na współczesne potrzeby rynku pracy i gospodarki opartej na wiedzy”***

UOGÓLNIONY MODEL LINIOWY (GLM)

```
moto <- read.table("motorins.txt",header=T)
str(moto)
```

```
'data.frame':  2182 obs. of  7 variables:
 $ Kilometres: int  1 1 1 1 1 1 1 1 1 1 ...
 $ Zone      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Bonus     : int  1 1 1 1 1 1 1 1 1 2 ...
 $ Make      : int  1 2 3 4 5 6 7 8 9 1 ...
 $ Insured   : num  455.1 69.2 72.9 1292.4 191 ...
 $ Claims    : int  108 19 13 124 40 57 23 14 1704 45 ...
 $ Payment   : int  392491 46221 15694 422201 119373 170913 56940 77487 6805992
214011 ...
```

```
fkilo=as.factor(Kilometres)
levels(fkilo) <- c("1","15","20","25","25+")#Kilometres travelled per year (tys)
motor <- transform(moto,fkilo=fkilo)
fzone <- as.factor(Zone)
levels(fzone) <- c("StoGoMa","large cities","south cities","south rural",
                   "north cities","north rural","Gotland")
motor <- transform(motor,fzone=fzone)
fmake <- as.factor(Make)
motor <- transform(motor,fmake=fmake)
```

```
attach(motor)
```

```
str(motor)
```

```
'data.frame':  2182 obs. of  10 variables:
 $ Kilometres: int  1 1 1 1 1 1 1 1 1 1 ...
 $ Zone      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Bonus     : int  1 1 1 1 1 1 1 1 1 2 ...
 $ Make      : int  1 2 3 4 5 6 7 8 9 1 ...
 $ Insured   : num  455.1 69.2 72.9 1292.4 191 ...
 $ Claims    : int  108 19 13 124 40 57 23 14 1704 45 ...
 $ Payment   : int  392491 46221 15694 422201 119373 170913 56940 77487 6805992
214011 ...
 $ fkilo     : Factor w/ 5 levels "1","15","20",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ fzone     : Factor w/ 7 levels "StoGoMa","large cities",..: 1 1 1 1 1 1 1 1 1
1 ...
 $ fmake     : Factor w/ 9 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 1 ...
```

```
fivenum(Payment)
```

```
[1]        0.0     2985.0    27403.5   112252.0 18245026.0
```

```
pay30 <- ifelse(Payment<30000,0,1)
pay30 <- as.factor(pay30)
```

```
pay30.lgst1 <- glm(pay30 ~ Insured, family=binomial)
summary(pay30.lgst1)
```

```
Call:
glm(formula = pay30 ~ Insured, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.7484  -0.6076  -0.5205   0.2407   2.0186

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.9325947  0.0890266  -21.71   <2e-16 ***
Insured      0.0140032  0.0007641   18.33   <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: 3024.7  on 2181  degrees of freedom
Residual deviance: 1626.5  on 2180  degrees of freedom
AIC: 1630.5

Number of Fisher Scoring iterations: 11
```

```
fivenum(Insured)
```

```
[1]      0.010     21.610     81.525    389.820 127687.270
```

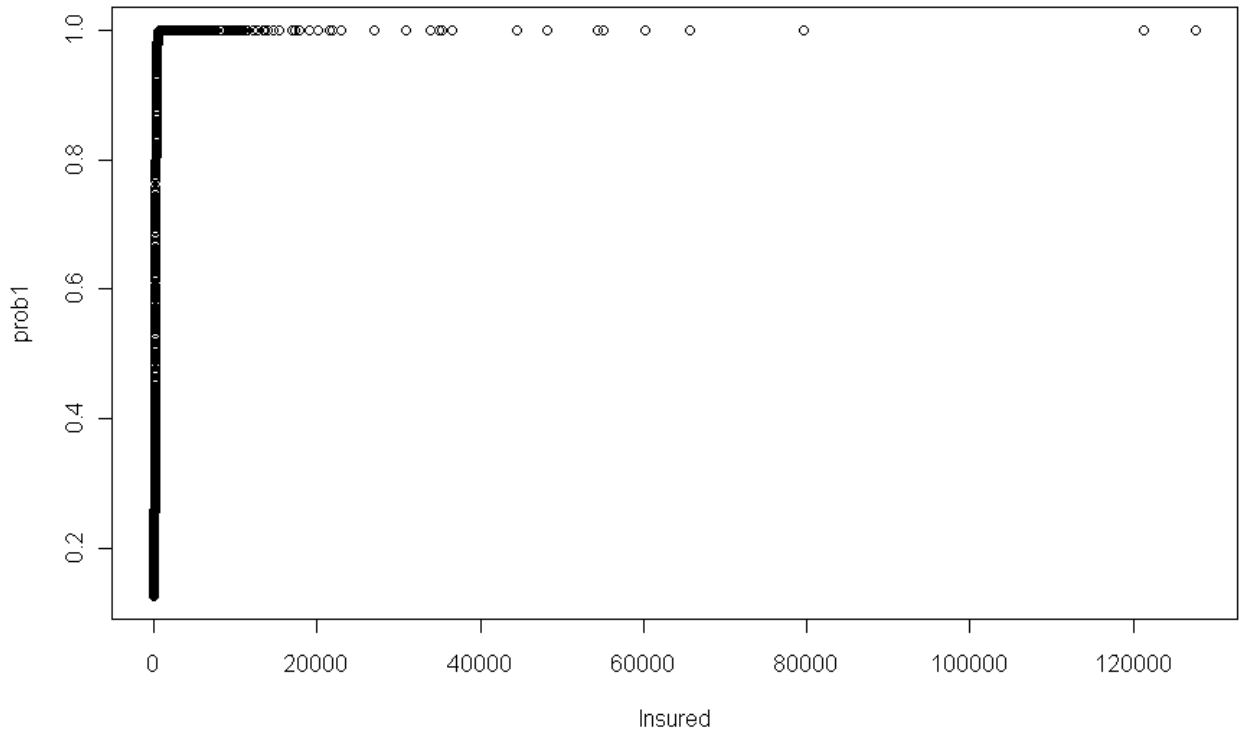
```
exp(10*coef(pay30.lgst1)["Insured"])
```

```
Insured
1.150311
```

```
prob1 <-predict(pay30.lgst1,type="response")
summary(prob1)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1265  0.1638  0.3120  0.4950  0.9714  1.0000
```

```
plot(prob1~Insured)
```




```
pay30.lgst11 <- glm(pay30 ~ log10(Insured), family=binomial)
summary(pay30.lgst11)
```

```
Call:
glm(formula = pay30 ~ log10(Insured), family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3860  -0.5444  -0.0143   0.4831   3.1735

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -6.3140     0.2777  -22.74   <2e-16 ***
log10(Insured)   3.2434     0.1404   23.11   <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: 3024.7  on 2181  degrees of freedom
Residual deviance: 1551.4  on 2180  degrees of freedom
AIC: 1555.4

Number of Fisher Scoring iterations: 6
```

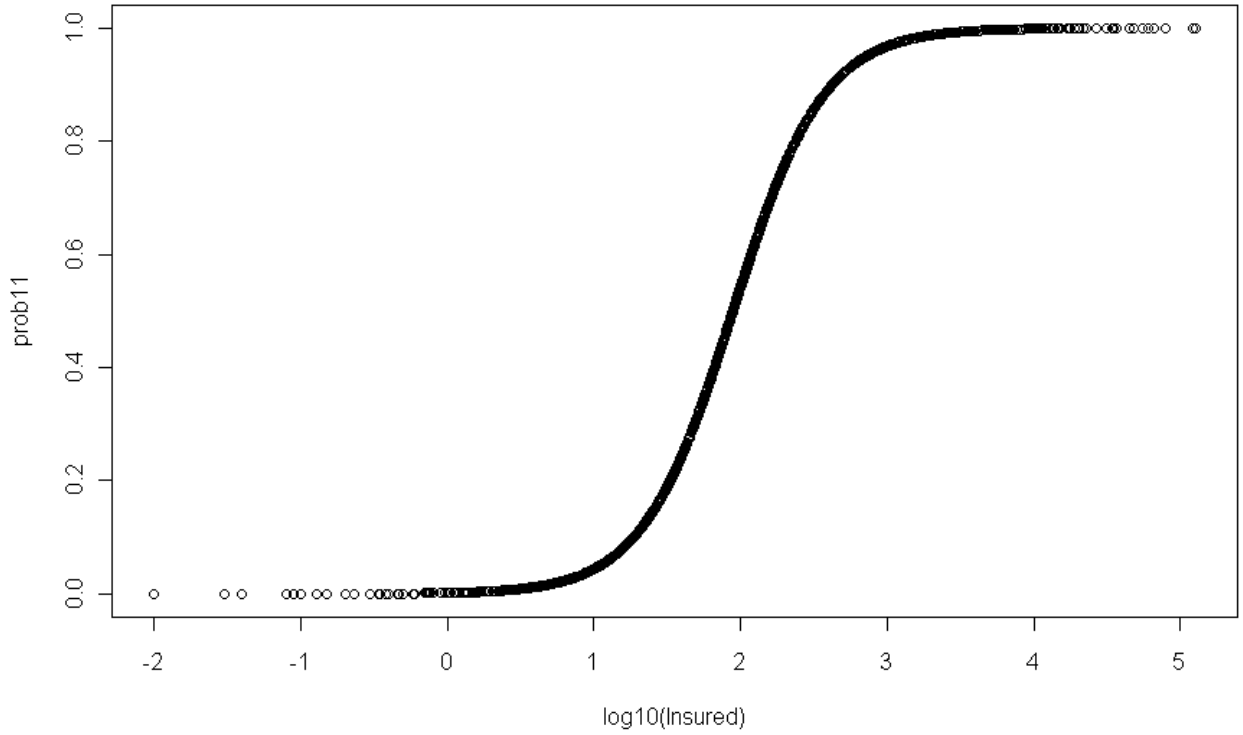
```
exp(coef(log10(2)*pay30.lgst11)["log10(Insured)"])
```

```
log10(Insured)
      2.654598
```

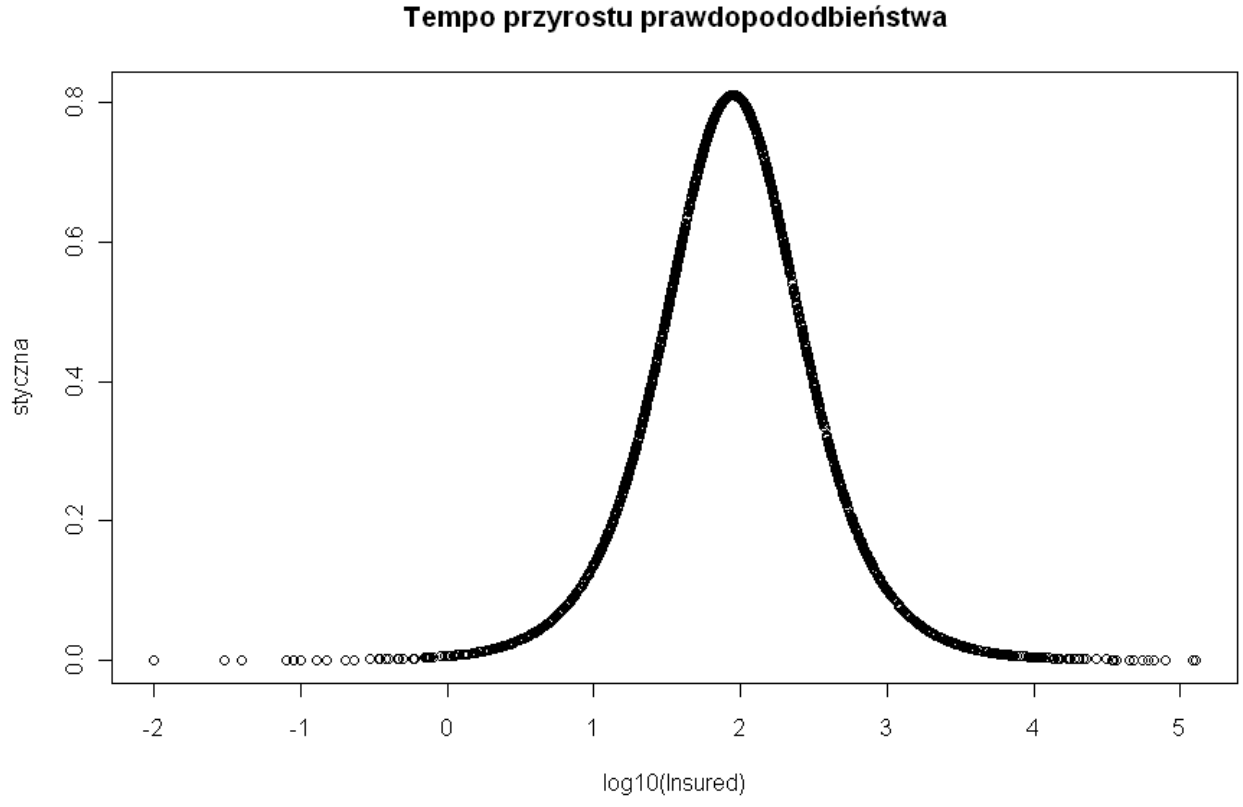
```
prob11 <-predict(pay30.lgst11,type="response")
summary(prob11)
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0000028 0.1208000 0.4713000 0.4950000 0.8898000 1.0000000
```

```
plot(prob11~log10(Insured))
```



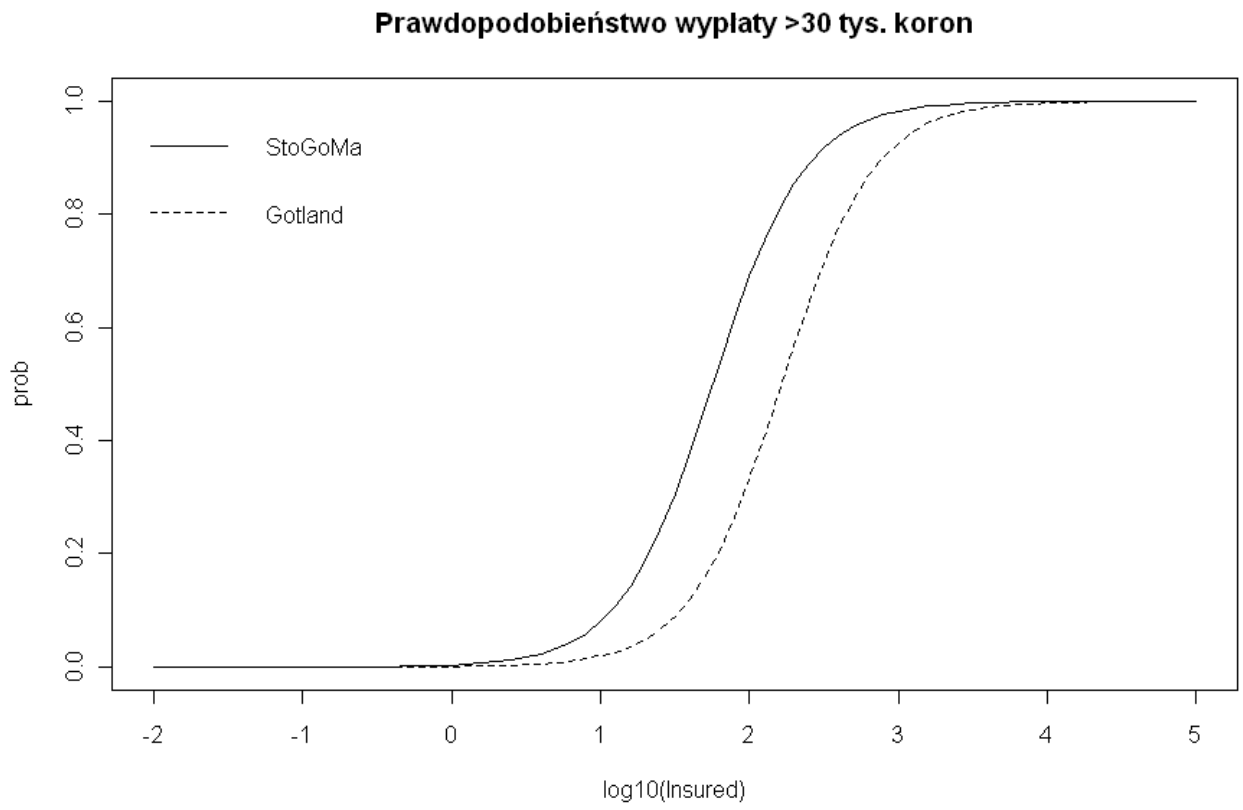

```
styczna <- coef(pay30.lgst11)["log10(Insured)"]*prob11*(1-prob11)
plot(styczna~log10(Insured),main="Tempo przyrostu prawdopododbieństwa")
```



```
lins<- log10(Insured)
pay30.lgst2 <- glm(pay30 ~ lins*fzone, family=binomial)
summary(pay30.lgst2)
```

```
Call:
glm(formula = pay30 ~ lins * fzone, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5713  -0.5301  -0.0228   0.4542   3.0901

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -5.86646    0.73007  -8.036 9.32e-16 ***
lins                       3.34276    0.39223   8.522  < 2e-16 ***
fzonelarge cities         -1.11013    1.07969  -1.028   0.3039
fzonesouth cities         -0.40748    1.04145  -0.391   0.6956
fzonesouth rural          -2.18059    1.25611  -1.736   0.0826 .
fzonenorth cities          0.32571    0.94950   0.343   0.7316
fzonenorth rural          -0.49802    1.01424  -0.491   0.6234
fzoneGotland              -0.15573    1.06841  -0.146   0.8841
lins:fzonelarge cities     0.17881    0.55617   0.322   0.7478
lins:fzonesouth cities    -0.07877    0.54096  -0.146   0.8842
lins:fzonesouth rural      0.68999    0.62756   1.099   0.2716
lins:fzonenorth cities    -0.48643    0.50649  -0.960   0.3369
lins:fzonenorth rural     -0.19521    0.52462  -0.372   0.7098
lins:fzoneGotland         -0.72941    0.56536  -1.290   0.1970
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3024.7  on 2181  degrees of freedom
Residual deviance: 1513.8  on 2168  degrees of freedom
AIC: 1541.8
```

```
plot(c(-2,5), c(0,1), type = "n", xlab = "log10(Insured)",
     ylab = "prob",main="Prawdopodobieństwo wypłaty >30 tys. koron")
ll <- seq(-2,5,0.1)
lines(ll, predict(pay30.lgst2,
                      data.frame(lins=ll,
                      fzone=factor(rep("StoGoMa", length(ll)),
                                   levels=levels(fzone))),
                    type = "response"))
lines(ll, predict(pay30.lgst2,
                  data.frame(lins=ll,
                             fzone=factor(rep("Gotland", length(ll)),
                                          levels=levels(fzone))),
                  type = "response"),lty=2)
legend("topleft",legend=c("StoGoMa","Gotland"),lty=1:2,bty="n")
```



```
summary(step(pay30.lgst2,trace=0))
```

```
Call:
glm(formula = pay30 ~ lins + fzone, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3648  -0.5365  -0.0113   0.4684   3.2221

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -5.6757     0.3072 -18.473  < 2e-16 ***
lins                3.2378     0.1449  22.351  < 2e-16 ***
fzonelarge cities  -0.7400     0.2255  -3.282 0.001031 **
fzonesouth cities  -0.5472     0.2254  -2.427 0.015215 *
fzonesouth rural   -0.7306     0.2303  -3.173 0.001509 **
fzonenorth cities  -0.5631     0.2319  -2.428 0.015182 *
fzonenorth rural   -0.8663     0.2276  -3.806 0.000141 ***
fzoneGotland       -1.5008     0.3174  -4.729 2.26e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3024.7  on 2181  degrees of freedom
Residual deviance: 1520.9  on 2174  degrees of freedom
AIC: 1536.9

Number of Fisher Scoring iterations: 6
```

```
summary(step(pay30.lgst21,trace=0,k=log(length(pay30))))
```

```
Call:
glm(formula = pay30 ~ lins, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3860  -0.5444  -0.0143   0.4831   3.1735

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.3140     0.2777  -22.74   <2e-16 ***
lins          3.2434     0.1404   23.11   <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: 3024.7  on 2181  degrees of freedom
Residual deviance: 1551.4  on 2180  degrees of freedom
AIC: 1555.4

Number of Fisher Scoring iterations: 6
```

```
pay30.lgst <- glm(pay30 ~ log10(Insured) *fkilo, family=binomial)
summary(pay30.lgst)
```

```
Call:
glm(formula = pay30 ~ log10(Insured) * fkilo, family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.61306  -0.54743  -0.01457   0.47011   3.11361

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             -6.9309     0.6915 -10.024   <2e-16 ***
log10(Insured)           3.3686     0.3250  10.364   <2e-16 ***
fkilo15                  0.6853     0.9747   0.703    0.482
fkilo20                 -0.2154     1.0230  -0.211    0.833
fkilo25                  0.4924     0.9184   0.536    0.592
fkilo25+                 1.1297     0.8828   1.280    0.201
log10(Insured):fkilo15  -0.1518     0.4608  -0.329    0.742
log10(Insured):fkilo20   0.3830     0.5049   0.759    0.448
log10(Insured):fkilo25  -0.1090     0.4564  -0.239    0.811
log10(Insured):fkilo25+ -0.2502     0.4496  -0.556    0.578
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3024.7  on 2181  degrees of freedom
Residual deviance: 1537.2  on 2172  degrees of freedom
AIC: 1557.2

Number of Fisher Scoring iterations: 6
```

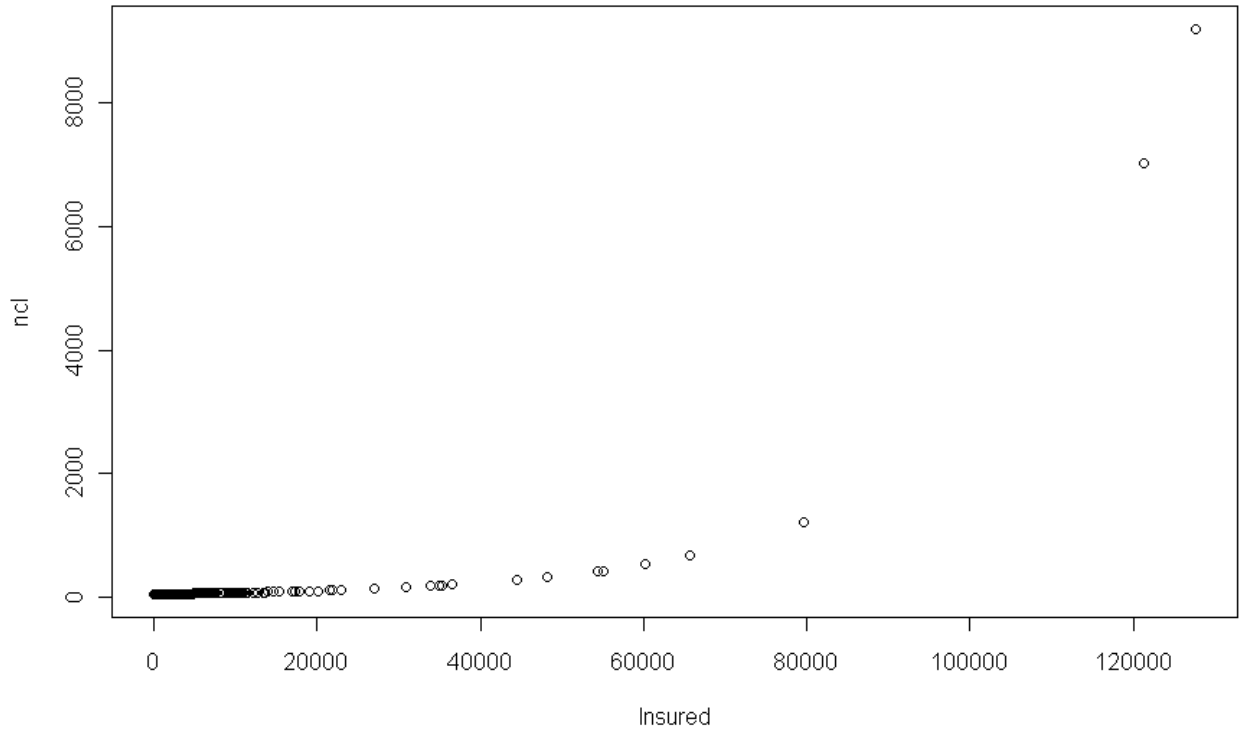
```
cl.poi <- glm(Claims~Insured,family=poisson())
summary(cl.poi)
```

```
Call:
glm(formula = Claims ~ Insured, family = poisson())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-76.964   -8.530   -7.153   -3.539   89.271

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 3.715e+00  3.295e-03  1127.6   <2e-16 ***
Insured     4.239e-05  6.729e-08   629.9   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 435505  on 2181  degrees of freedom
Residual deviance: 285638  on 2180  degrees of freedom
AIC: 293280
```

```
ncl <- predict(cl.poi,type="response")
summary(ncl)
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  41.07   41.11   41.22   51.87   41.76 9204.00
```

```
plot(ncl~Insured)
```



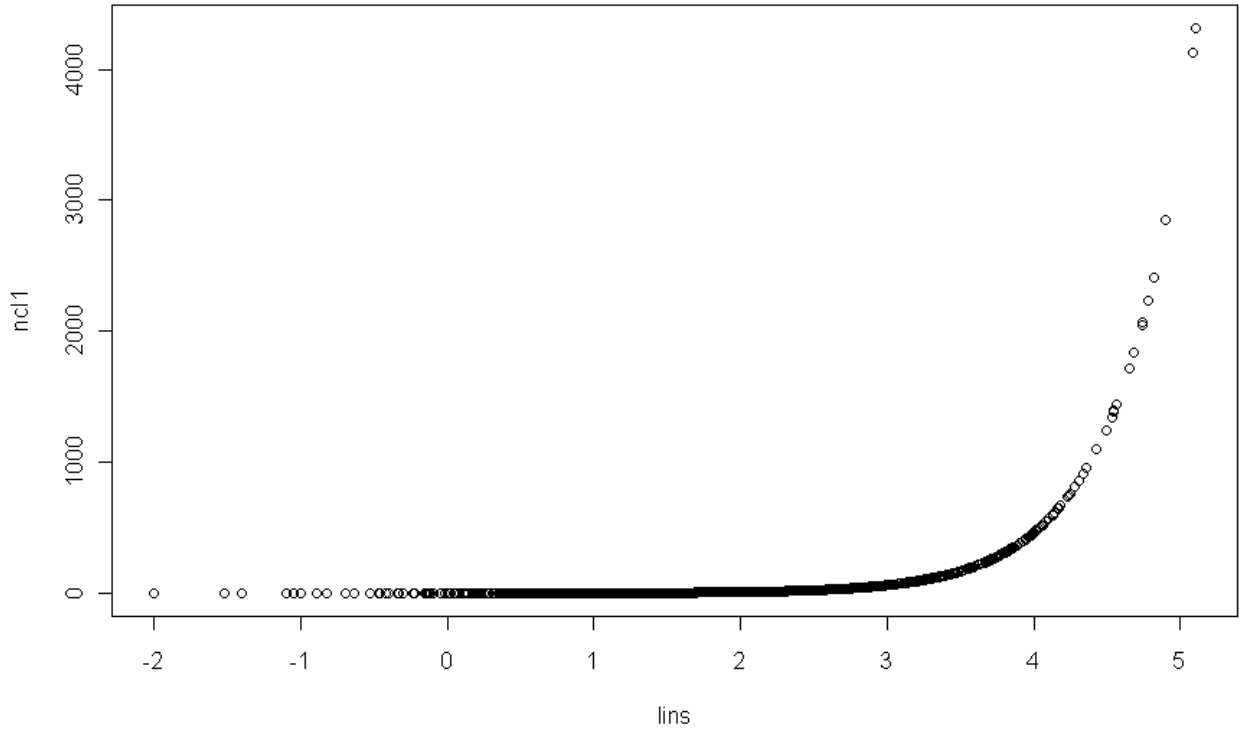

```
cl.poi2 <- glm(Claims~lins,family=quasipoisson())
summary(cl.poi2)
```

```
Call:
glm(formula = Claims ~ lins, family = quasipoisson())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-23.042   -1.716   -0.760    0.420   44.293

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.94063    0.05373  -36.12   <2e-16 ***
lins         2.01930    0.01377  146.66   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 15.50607)

    Null deviance: 435505  on 2181  degrees of freedom
Residual deviance:  27793  on 2180  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

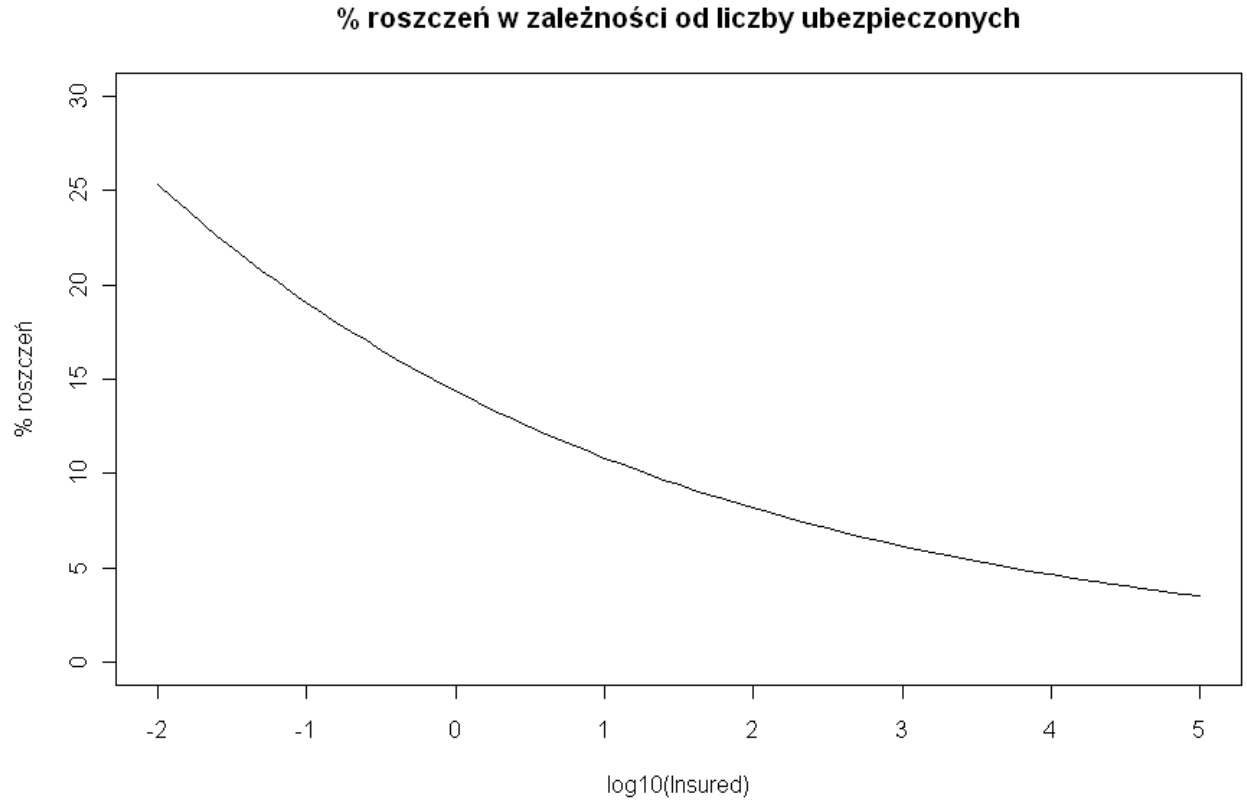
```
ncl2 <- predict(cl.poi2,type="response")
summary(ncl2)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   0.003    2.126    6.813   51.870   26.870 4316.000
```

```
plot(ncl1~lins)
```



```
plot(c(-2,5), c(0,30), type = "n", xlab = "log10(Insured)",
     ylab = "% roszczeń",main="% roszczeń w zależności od liczby
ubezpieczonych")
ll <- seq(-2,5,0.1)
lines(ll, predict(cl.poi2,
                  data.frame(lins=ll),type = "response")/10^(ll-2))
```




```
cl.poi3 <- glm(Claims~lins+fzone,family=quasipoisson())
summary(cl.poi3)
Call:
glm(formula = Claims ~ lins + fzone, family = quasipoisson())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-16.259   -1.542   -0.600    0.514   34.611

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.60456    0.05324 -30.136  < 2e-16 ***
lins               2.02713    0.01300 155.978  < 2e-16 ***
fzonelarge cities -0.23594    0.03329  -7.087 1.84e-12 ***
fzonesouth cities -0.39061    0.03390 -11.522  < 2e-16 ***
fzonesouth rural  -0.51638    0.03075 -16.792  < 2e-16 ***
fzonenorth cities -0.48060    0.05122  -9.383  < 2e-16 ***
fzonenorth rural  -0.59337    0.04160 -14.264  < 2e-16 ***
fzoneGotland      -1.09993    0.14345  -7.667 2.63e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 12.28943)

    Null deviance: 435505  on 2181  degrees of freedom
Residual deviance:  22836  on 2174  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

```
plot(c(-2,5), c(0,40), type = "n", xlab = "log10(Insured)",
     ylab = "% roszczeń",main="% roszczeń w zależności od liczby
ubezpieczonych")
lines(ll, predict(cl.poi3,
                  data.frame(lins=ll,
                             fzone=factor(rep("StoGoMa", length(ll)),
                                          levels=levels(fzone))),
                  type = "response")/10^(ll-2))
lines(ll, predict(cl.poi3,
                  data.frame(lins=ll,
                             fzone=factor(rep("Gotland", length(ll)),
                                          levels=levels(fzone))),
                  type = "response")/10^(ll-2),lty=2)
legend("topright",legend=c("StoGoMa","Gotland"),lty=1:2,bty="n")
```