

Modele regresji i ich zastosowania

Labolatoria 9 i 10

Jan Solarz
243889

25 maja 2021

Spis treści

1	Zadanie 1	1
1.1	Podpunkt a	1
1.2	Podpunkt b	2
1.3	Podpunkt c	2
1.4	Podpunkt d	4
1.5	Podpunkt e	5
1.6	Podpunkt f	7
2	Zadanie 2	9
2.1	Podpunkt a	9
2.2	Podpunkt b	9
2.3	Podpunkt c	14

1 Zadanie 1

1.1 Podpunkt a

Zaczynamy od podzielenia zbioru na część uczącą i część testową w proporcji 70:30. Sprawdzamy na koniec podobieństwo frakcji STAN w zbiorach.

```
logistyczna <- read.delim("~/Downloads/logistyczna.txt", header=TRUE, comment.char="#")  
head(logistyczna)
```

```
## STAN RMS10 RMS20 RMS30 PWD A DD YA YDD AR  
## 1 0 0.06 0.06 0.08 1.02 112.07 0.22 44.72 0.21 0.19  
## 2 0 0.03 0.06 0.06 1.20 116.90 0.25 93.07 0.17 0.23  
## 3 0 0.05 0.05 0.08 1.23 155.88 0.47 42.79 0.29 0.85  
## 4 0 0.04 0.05 0.09 1.33 124.59 0.19 44.78 0.27 0.45  
## 5 0 0.05 0.06 0.06 1.16 133.42 0.17 40.72 0.35 0.28  
## 6 0 0.09 0.13 0.13 1.07 136.54 0.20 55.10 0.17 0.39
```

```

indices <- sort(sample(1:267, round(0.7 * 267)))

logistyczne_train<-logistyczna[indices,]
logistyczne_test<-logistyczna[-indices,]

sum(logistyczna$STAN)

## [1] 232

sum(logistyczne_train$STAN)

## [1] 164

sum(logistyczne_test$STAN)

## [1] 68

```

1.2 Podpunkt b

Konstruujemy model regresji logistycznej oparty na wszystkich zmiennych, jakie mamy do dyspozycji.

```

model1<-glm(logistyczne_train$STAN~.,data=logistyczne_train, family = binomial(link = "logit"),
summary(model1)

##
## Call:
## glm(formula = logistyczne_train$STAN ~ ., family = binomial(link = "logit"),
##      data = logistyczne_train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.10332   0.08382   0.21015   0.44111   1.40250
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -23.898454   5.740621  -4.163 3.14e-05 ***
## RMS10        28.017566  23.368747   1.199  0.2306
## RMS20       -14.150810  24.794920  -0.571  0.5682
## RMS30        15.424945  17.206585   0.896  0.3700
## PWD          12.994665   3.207363   4.052 5.09e-05 ***
## A              0.025891   0.015243   1.699  0.0894 .
## DD           22.046541   9.720716   2.268  0.0233 *
## YA          -0.001278   0.024409  -0.052  0.9582
## YDD           0.600599   5.114981   0.117  0.9065
## AR            4.670468   2.561452   1.823  0.0682 .
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 139.446  on 186  degrees of freedom
## Residual deviance:  98.544  on 177  degrees of freedom
## AIC: 118.54
##
## Number of Fisher Scoring iterations: 7
```

1.3 Podpunkt c

```
library(lmtest)

## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

model2<-glm(logistyczne_train$STAN~1,data=logistyczne_train, family = binomial(link = "logit"),
summary(model2)

##
## Call:
## glm(formula = logistyczne_train$STAN ~ 1, family = binomial(link = "logit"),
##      data = logistyczne_train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0473   0.5123   0.5123   0.5123   0.5123
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.9644     0.2227   8.822  <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: 139.45  on 186  degrees of freedom
## Residual deviance: 139.45  on 186  degrees of freedom
## AIC: 141.45
##
## Number of Fisher Scoring iterations: 4
```

```

lrtest(model1,model2)

## Likelihood ratio test
##
## Model 1: logistyczne_train$STAN ~ RMS10 + RMS20 + RMS30 + PWD + A + DD +
##      YA + YDD + AR
## Model 2: logistyczne_train$STAN ~ 1
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   10 -49.272
## 2    1 -69.723 -9 40.901  5.213e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# intercept_only <- lm(Y ~ 1, data=mydata_2)
#
# #define model with all predictors
# all <- lm(Y ~ ., data=mydata_2)
#
# #perform forward stepwise regression
# forward <- step(model2, direction='forward', scope=formula(model1), trace=0)
#
# #view results of forward stepwise regression
# forward$anova
#
# #view final model
# forward$coefficients
# summary(forward)

```

- P-value wynosi 0.0004606 i jest ona mniejsza od poziomu istotności na poziomie 0.05
- Poprzez oszacowanie statystyki jesteśmy w stanie odrzucić hipotezę zerową.
- Odrzucenie testu ma sens, widzieliśmy duże znaczenie na zmienną STAN przez atrybut PWD.

Poprzez oszacowanie statystyki jesteśmy w stanie odrzucić hipotezę zerową.

1.4 Podpunkt d

Liczymy wartości estymatorów i standardowych błędów predyktorów z wyjściowego modelu.

```

model1$coefficients

##      (Intercept)          RMS10          RMS20          RMS30          PWD
## -23.898453997    28.017565506   -14.150810109    15.424944737    12.994664971
##              A              DD              YA              YDD              AR
##    0.025891494    22.046541090   -0.001278148    0.600598986    4.670468404

sqrt(diag(summary(model1)$cov.unscaled)*summary(model1)$dispersion)

```

## (Intercept)	RMS10	RMS20	RMS30	PWD	A
## 5.74062068	23.36874667	24.79491970	17.20658543	3.20736292	0.01524286
## DD	YA	YDD	AR		
## 9.72071648	0.02440914	5.11498087	2.56145217		

Największe wartości błędów dotyczą zmiennych RMS10, RMS20 i RMS30.

```
confint(model1, level = 0.95)
```

```
## Waiting for profiling to be done...
```

##	2.5 %	97.5 %
## (Intercept)	-36.455414711	-13.70139645
## RMS10	-16.853960385	76.03649746
## RMS20	-66.368850297	33.56924761
## RMS30	-16.620839528	52.97215903
## PWD	7.257806160	19.95021953
## A	-0.003180568	0.05732016
## DD	4.944941093	42.86830994
## YA	-0.048136357	0.04859228
## YDD	-9.381964968	10.88636616
## AR	-0.006619474	10.09161433

```
summary(model1)
```

```
##
```

```
## Call:
```

```
## glm(formula = logistyczne_train$STAN ~ ., family = binomial(link = "logit"),  
## data = logistyczne_train)
```

```
##
```

```
## Deviance Residuals:
```

##	Min	1Q	Median	3Q	Max
##	-3.10332	0.08382	0.21015	0.44111	1.40250

```
##
```

```
## Coefficients:
```

##	Estimate	Std. Error	z value	Pr(> z)
## (Intercept)	-23.898454	5.740621	-4.163	3.14e-05 ***
## RMS10	28.017566	23.368747	1.199	0.2306
## RMS20	-14.150810	24.794920	-0.571	0.5682
## RMS30	15.424945	17.206585	0.896	0.3700
## PWD	12.994665	3.207363	4.052	5.09e-05 ***
## A	0.025891	0.015243	1.699	0.0894 .
## DD	22.046541	9.720716	2.268	0.0233 *
## YA	-0.001278	0.024409	-0.052	0.9582
## YDD	0.600599	5.114981	0.117	0.9065
## AR	4.670468	2.561452	1.823	0.0682 .

```
## ---
```

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

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 139.446  on 186  degrees of freedom
## Residual deviance:  98.544  on 177  degrees of freedom
## AIC: 118.54
##
## Number of Fisher Scoring iterations: 7
```

- Hipoteza zerowa $\beta_i = 0$ jest odrzucana tylko dla $i = 5$
- Zmienna PWD ma zdecydowanie największy wpływ na zmienną STAN, dla pozostałych zmiennych hipoteza zerowa mogłaby być przyjęta. Poziom istotności był wyższy od 0.05

1.5 Podpunkt e

Konstruujemy *classification table* i wyznaczamy parametry charakteryzujące zdolności predykcyjne modelu na dwóch poziomach odcięcia 0.5 i 0.7

```
library(knitr)
n=length(logistyczne_test$STAN)
real_values <- logistyczne_test$STAN
pred_values <- predict(model1, newdata = logistyczne_test)
class_tab <- table(real_values, pred_values>0.5)
true_negative <- class_tab[1,1]
false_positive <- class_tab[1,2]
false_negative <- class_tab[2,1]
true_positive <- class_tab[2,2]
paste0("Sensitivity for pi 0.5: ", round ((true_positive/(true_positive+false_negative)) * 100))

## [1] "Sensitivity for pi 0.5: 89.70588"

paste0("Specificity for pi 0.5: ", round ((true_negative/(true_negative+false_positive)) * 100))

## [1] "Specificity for pi 0.5: 8.33333"

paste0("False positive rate for pi 0.5: ", round (false_positive/(false_positive+true_negative) * 100))

## [1] "False positive rate for pi 0.5: 0.91667"

paste0("False negative rate for pi 0.5: ", round (false_negative/(false_negative+true_negative) * 100))

## [1] "False negative rate for pi 0.5: 0.875"

paste0("Overall proportion of correct classifications for pi 0.5: ", round ((true_negative+true_positive)/n * 100))

## [1] "Overall proportion of correct classifications for pi 0.5: 0.775"

kable(class_tab)
```

	FALSE	TRUE
0	1	11
1	7	61

```

n=length(logistyczne_test$STAN)
real_values <- logistyczne_test$STAN
pred_values <- predict(model1, newdata = logistyczne_test)
class_tab <- table(real_values, pred_values>0.7)
true_negative <- class_tab[1,1]
false_positive <- class_tab[1,2]
false_negative <- class_tab[2,1]
true_positive <- class_tab[2,2]
paste0("Sensitivity for pi 0.7: ", round ((true_positive/(true_positive+false_negative)) * 100))

## [1] "Sensitivity for pi 0.7: 88.23529"

paste0("Specifity for pi 0.7: ", round ((true_negative/(true_negative+false_positive)) * 100))

## [1] "Specifity for pi 0.7: 16.66667"

paste0("False positive rate for pi 0.7: ", round (false_positive/(false_positive+true_negative)) * 100)

## [1] "False positive rate for pi 0.7: 0.83333"

paste0("False negative rate for pi 0.7: ", round (false_negative/(false_negative+true_negative)) * 100)

## [1] "False negative rate for pi 0.7: 0.8"

paste0("Overall proportion of correct classifications for pi 0.7: ", round ((true_negative+true_positive)/n) * 100)

## [1] "Overall proportion of correct classifications for pi 0.7: 0.775"

kable(class_tab)

```

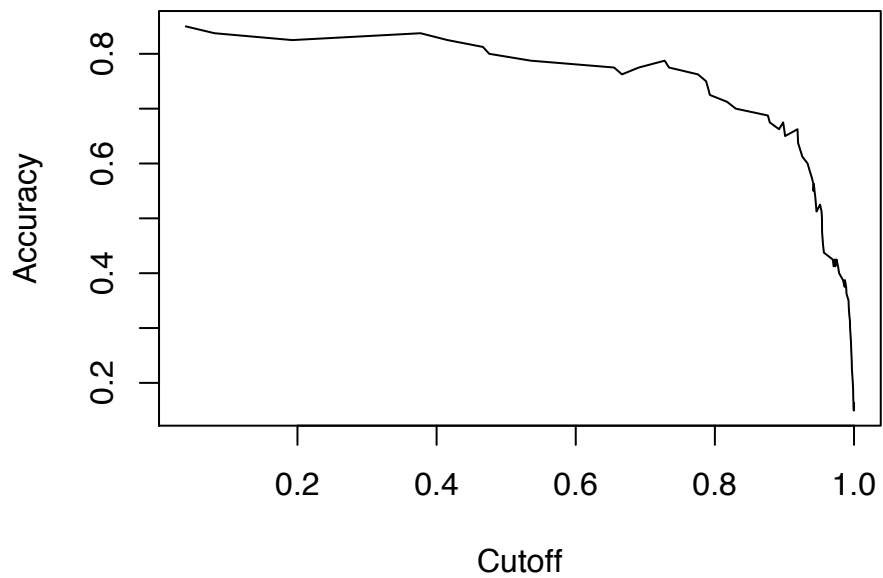
	FALSE	TRUE
0	2	10
1	8	60

1.6 Podpunkt f

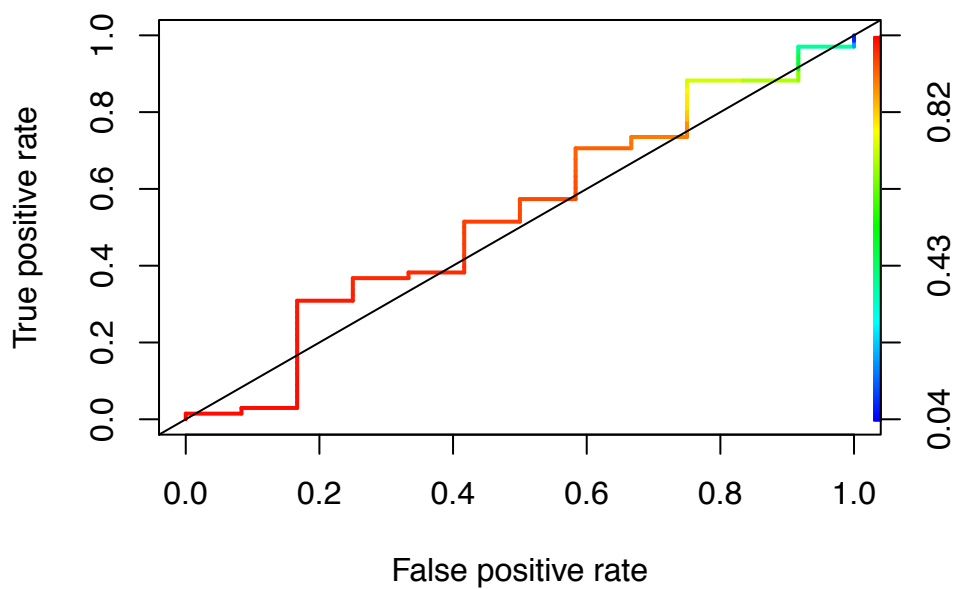
```

library(ROCR)
pred = predict(model1,logistyczne_test,type="response")
pred = prediction(pred, logistyczne_test$STAN)
perf = performance(pred, "acc")
plot(perf)

```



```
roc = performance(pred, "tpr", "fpr")
plot(roc, colorize = T, lwd = 2)
abline(a = 0, b = 1)
```



```
library(pROC)

## Type 'citation("pROC")' for a citation.
```



```
##
## Attaching package: 'pROC'
## The following objects are masked from 'package:stats':
##
##      cov, smooth, var

pred = predict(model1,logistyczne_test,type="response")
auc<-auc(logistyczne_test$STAN, pred)

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

auc

## Area under the curve: 0.5306

2*auc-1

## [1] 0.06127451
```

- Rysunek krzywej ROC wskazuje, że model na prawie każdym poziomie przewyższa skuteczność 0.5 co oznacza, że ma lepszą skuteczność dopasowania niż klasyfikator losowy.
- Pole które jest zakresłone pod krzywą ROC jest nazywane parametrem AUC. Wynosi on 0.6299 i jest powyżej 0.5 stąd lepsze skuteczność niż klasyfikatora losowego, daleko jednak do klasyfikatora idealnego -poziom AUC=1
- Współczynnik Giniego: $GC = 2*AUC-1$ (wyższość klasyfikatora nad losowym)
- Współczynnik Giniego:0.2598039

2 Zadanie 2

2.1 Podpunkt a

```
library(readxl)
cancer <- read_excel("/Users/jansolarz/Downloads/cancer remission.xlsx")

## New names:
## * '' -> ...8
## * '' -> ...9
## * '' -> ...10
## * '' -> ...11

cancer<-cancer[,1:7]
```

2.2 Podpunkt b

Zgodnie z metodą forward zaczynamy do przyłączania pojedynczych zmiennych niezależnych sprawdzając wysokość poziomu istotności zgodnie z zadanymi poziomami w zadaniu.

```
model_1<-glm(cancer$remission~1,data=cancer, family = binomial(link = "logit"))
model_2<-glm(cancer$remission~cancer$cell,data=cancer, family = binomial(link = "logit"))

lrtest(model_1,model_2)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$cell
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -15.896  1  2.58    0.1082

model_3<-glm(cancer$remission~cancer$smear,data=cancer, family = binomial(link = "logit"))
lrtest(model_1,model_3)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$smear
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -16.640  1 1.0921    0.296

model_4<-glm(cancer$remission~cancer$infil,data=cancer, family = binomial(link = "logit"))
lrtest(model_1,model_4)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$infil
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -16.201  1 1.9698    0.1605

model_5<-glm(cancer$remission~cancer$li,data=cancer, family = binomial(link = "logit"))
lrtest(model_1,model_5)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$li
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -13.037  1 8.2988  0.003967 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

model_6<-glm(cancer$remission~cancer$blast,data=cancer, family = binomial(link = "logit"))
lrtest(model_1,model_6)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$blast
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -15.410  1 3.5513    0.0595 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model_7<-glm(cancer$remission~cancer$temp,data=cancer, family = binomial(link = "logit"))
lrtest(model_1,model_7)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ 1
## Model 2: cancer$remission ~ cancer$temp
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -17.186
## 2    2 -16.837  1 0.698    0.4035

# #define intercept-only model
# intercept_only <- lm(Y ~ 1, data=mydata_2) #define model with all predictors
# all <- lm(Y ~ ., data=mydata_2) #perform forward stepwise regression
# forward <- step(intercept_only, direction='forward', scope=formula(all), trace=0)

```

Przeprowadzone testy wskazują na najniższą wartość $\text{Pr}(\chi^2_{\text{Chisq}})$ przy zmiennej li 0.003967 dlatego przyłączamy ją do naszego modelu.

```

model_1<-glm(cancer$remission~cancer$li,data=cancer, family = binomial(link = "logit"))
model_2<-glm(cancer$remission~cancer$li+cancer$cell,data=cancer, family = binomial(link

lrtest(model_1,model_2)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li
## Model 2: cancer$remission ~ cancer$li + cancer$cell
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.037
## 2    3 -12.170  1 1.7322    0.1881

model_3<-glm(cancer$remission~cancer$li+cancer$smear,data=cancer, family = binomial(link
lrtest(model_1,model_3)

## Likelihood ratio test

```

```
##
## Model 1: cancer$remission ~ cancer$li
## Model 2: cancer$remission ~ cancer$li + cancer$smear
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.037
## 2    3 -12.969  1 0.1356    0.7127

model_4<-glm(cancer$remission~cancer$li+cancer$infil,data=cancer, family = binomial(link=logit))
lrtest(model_1,model_4)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li
## Model 2: cancer$remission ~ cancer$li + cancer$infil
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.037
## 2    3 -12.745  1 0.5825    0.4453

model_6<-glm(cancer$remission~cancer$li+cancer$blast,data=cancer, family = binomial(link=logit))
lrtest(model_1,model_6)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li
## Model 2: cancer$remission ~ cancer$li + cancer$blast
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.037
## 2    3 -12.991  1 0.0918    0.7619

model_7<-glm(cancer$remission~cancer$li+cancer$temp,data=cancer, family = binomial(link=logit))
lrtest(model_1,model_7)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li
## Model 2: cancer$remission ~ cancer$li + cancer$temp
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -13.037
## 2    3 -12.324  1 1.4251    0.2326
```

Przeprowadzone testy wskazują na najniższą wartość $\Pr(\chi^2_{\text{Chisq}})$ przy zmiennej *cell* 0.1881 dlatego przyłączamy ją do naszego modelu.

```
model_1<-glm(cancer$remission~cancer$li+cancer$cell,data=cancer, family = binomial(link=logit))
model_2<-glm(cancer$remission~cancer$li+cancer$cell+cancer$smear,data=cancer, family = binomial(link=logit))
lrtest(model_1,model_2)

## Likelihood ratio test
##
```

```
## Model 1: cancer$remission ~ cancer$li + cancer$cell
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$smear
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -12.170
## 2    4 -12.146  1 0.0483    0.8261

model_4<-glm(cancer$remission~cancer$li+cancer$cell+cancer$infil,data=cancer, family = b
lrtest(model_1,model_4)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$infil
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -12.170
## 2    4 -12.151  1 0.0391    0.8433

model_6<-glm(cancer$remission~cancer$li+cancer$cell+cancer$blast,data=cancer, family = b
lrtest(model_1,model_6)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$blast
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -12.170
## 2    4 -12.093  1 0.1545    0.6943

model_7<-glm(cancer$remission~cancer$li+cancer$cell+cancer$temp,data=cancer, family = bi
lrtest(model_1,model_7)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$temp
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -12.170
## 2    4 -10.977  1 2.3874    0.1223
```

Przeprowadzone testy wskazują na najniższą wartość $Pr(\chi^2_{Chisq})$ przy zmiennej *temp* 0.1223 dlatego przyłączamy ją do naszego modelu.

```
model_1<-glm(cancer$remission~cancer$li+cancer$cell+cancer$temp,data=cancer, family = bi
model_2<-glm(cancer$remission~cancer$li+cancer$cell+cancer$temp+cancer$smear,data=cancer
lrtest(model_1,model_2)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell + cancer$temp
```

```
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$temp + cancer$smear
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    4 -10.977
## 2    5 -10.929  1 0.0954    0.7574

model_4<-glm(cancer$remission~cancer$li+cancer$cell+cancer$temp+cancer$infil,data=cancer)
lrtest(model_1,model_4)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell + cancer$temp
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$temp + cancer$infil
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    4 -10.977
## 2    5 -10.935  1 0.0842    0.7716

model_6<-glm(cancer$remission~cancer$li+cancer$cell+cancer$temp+cancer$blast,data=cancer)
lrtest(model_1,model_6)

## Likelihood ratio test
##
## Model 1: cancer$remission ~ cancer$li + cancer$cell + cancer$temp
## Model 2: cancer$remission ~ cancer$li + cancer$cell + cancer$temp + cancer$blast
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    4 -10.977
## 2    5 -10.966  1 0.0209    0.8852
```

Przeprowadzone testy wskazują na najniższą wartość $Pr(>Chisq)$ przy zmiennej *smear* 0.7574 nie przyłączamy jej jednak do modelu ze względu na to że jest wyższa od przyjętego poziomu istotności.

Nasz ostateczny model

```
summary(model_1)

##
## Call:
## glm(formula = cancer$remission ~ cancer$li + cancer$cell + cancer$temp,
##      family = binomial(link = "logit"), data = cancer)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.02043  -0.66313  -0.08323   0.81282   1.65887
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   67.634     56.888   1.189   0.2345
## cancer$li      3.867      1.778   2.175   0.0297 *
## cancer$cell    9.652      7.751   1.245   0.2130
## cancer$temp  -82.074     61.712  -1.330   0.1835
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 34.372  on 26  degrees of freedom
## Residual deviance: 21.953  on 23  degrees of freedom
## AIC: 29.953
##
## Number of Fisher Scoring iterations: 7

sqrt(diag(summary(model_1)$cov.unscaled)*summary(model_1)$dispersion)

## (Intercept)    cancer$li cancer$cell cancer$temp
##    56.887547     1.778278     7.751076    61.712382
```

- Predyktorami zostają zmienne $li, cell, temp$
- Wartości estymatorów wynoszą: $\hat{\beta}_1 = 67.634$, $\hat{\beta}_2 = 3.867$, $\hat{\beta}_3 = 9.652$, $\hat{\beta}_4 = -82.074$
- Wartości SE odpowiednio 56.887547, 1.778278, 7.751076, 61.712382
- AIC: 29.953

2.3 Podpunkt c

- Dodajemy do tabeli *cancer remission* $\pi(x) := P(remiss = 1 \mid X = x)$ oraz lewy i prawy koniec przedziału ufności na poziomie 95 procent.

```
func = function(x){ exp(x)/(1+exp(x))}
tabela <- data.frame(cancer)
predicted_values <- predict(model_1, data = cancer, type='link', se.fit=TRUE)
pi <- func(predicted_values$fit)
high <- func(predicted_values$fit + (predicted_values$se.fit*qnorm(0.95)))
low <- func(predicted_values$fit - (predicted_values$se.fit*qnorm(0.95)))
tabela<-cbind(tabela,pi,low,high)
kable(tabela)
```

remission	cell	smear	infil	li	blast	temp	pi	low	high
1	0.80	0.83	0.66	1.9	1.100	0.996	0.7226489	0.2344817	0.9568289
1	0.90	0.36	0.32	1.4	0.740	0.992	0.5787391	0.3115861	0.8065760
0	0.80	0.88	0.70	0.8	0.176	0.982	0.1045990	0.0119993	0.5291077
0	1.00	0.87	0.87	0.7	1.053	0.986	0.2825773	0.0946241	0.5974888
1	0.90	0.75	0.68	1.3	0.519	0.980	0.7141804	0.3175562	0.9306408
0	1.00	0.65	0.65	0.6	0.519	0.982	0.2708868	0.0765172	0.6248959
1	0.95	0.97	0.92	1.0	1.230	0.992	0.3215554	0.1549074	0.5506637
0	0.95	0.87	0.83	1.9	1.354	1.020	0.6072319	0.1516246	0.9304290
0	1.00	0.45	0.45	0.8	0.322	0.999	0.1663164	0.0402579	0.4868618
0	0.95	0.36	0.34	0.5	0.000	1.038	0.0015693	0.0000036	0.4090487
0	0.85	0.39	0.33	0.7	0.279	0.988	0.0728520	0.0092136	0.3990210
0	0.70	0.76	0.53	1.2	0.146	0.982	0.1728570	0.0110953	0.7956046
0	0.80	0.46	0.37	0.4	0.380	1.006	0.0034575	0.0000336	0.2636162
0	0.20	0.39	0.08	0.8	0.114	0.990	0.0001850	0.0000000	0.8017311
0	1.00	0.90	0.90	1.1	1.037	0.990	0.5712204	0.2968420	0.8078416
1	1.00	0.84	0.84	1.9	2.064	1.020	0.7146954	0.2167842	0.9577549
0	0.65	0.42	0.27	0.5	0.114	1.014	0.0006223	0.0000008	0.3203593
0	1.00	0.75	0.75	1.0	1.322	1.004	0.2228888	0.0590836	0.5671126
0	0.50	0.44	0.22	0.6	0.114	0.990	0.0015425	0.0000022	0.5260312
1	1.00	0.63	0.63	1.1	1.072	0.986	0.6491095	0.3174187	0.8803675
0	1.00	0.33	0.33	0.4	0.176	1.010	0.0169297	0.0005605	0.3459119
0	0.90	0.93	0.84	0.6	1.591	1.020	0.0062175	0.0000721	0.3517630
1	1.00	0.58	0.58	1.0	0.531	1.002	0.2526057	0.0784679	0.5729315
0	0.95	0.32	0.30	1.6	0.886	0.988	0.8701089	0.4993059	0.9782600
1	1.00	0.60	0.60	1.7	0.964	0.990	0.9313166	0.5549406	0.9932640
1	1.00	0.69	0.69	0.9	0.398	0.986	0.4605092	0.2011016	0.7432325
0	1.00	0.73	0.73	0.7	0.398	0.986	0.2825773	0.0946241	0.5974888