

# Homework 4

Gianni Spiga

2023-02-05

## Problem Set 3

### Question 8

a.)

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
767	993.4839	NA	NA	NA
761	748.0105	6	245.4734	0

```
## [1] NA 3.802907e-50
```

We first build a full model and null model, and then perform a likelihood ratio test. Performing said test gives us an extremely small p-value as shown above, thus leading us to reject  $H_0$  and conclude that there is an overall regression effect. Because of this, we can continue to develop a meaningful regression analysis.

b.)

```
library(MASS)
# We first fit with the AIC
pima.AIC <- stepAIC(pima.fit, trace = F)
pima.BIC <- stepAIC(pima.fit, trace = F, k = log(nrow(pima)))
kable(pima.AIC$coefficients)
```

	x
(Intercept)	-7.9703942
PGC	0.0330360
DBP	-0.0124186
BMI	0.0885148
Age	0.0345947

```
kable(pima.BIC$coefficients)
```

	x
(Intercept)	-8.3937430
PGC	0.0325117
BMI	0.0815896
Age	0.0301570

The first set of coefficients is one that was picked by AIC and the second set of coefficients is those picked by the BIC criterion. We can see that the AIC coefficients contains all the coefficients from BIC with the addition of Diastolic Blood Pressure.

c.)

We will continue with the model chosen by the BIC criterion, in effort to keep our model as simple and understandable as possible. Let's create a fit for this new model.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.3937430	0.6660667	-12.601955	0.00e+00
PGC	0.0325117	0.0033288	9.766912	0.00e+00
BMI	0.0815896	0.0135258	6.032159	0.00e+00
Age	0.0301570	0.0076323	3.951247	7.77e-05

From the table above, we can see that all of the coefficients in this model are significant at any reasonable choice of  $\alpha$ . Thus we can conclude categorizing whether a Pima Indian is diabetic or not is predicted by their PGC, Body Mass Index, and Age.

## Appendix

```
library(knitr)
pima <- read.table("pima-indians-diabetes.dat", sep = ",")

names(pima) <- c("Unknown", "PGC", "DBP", "Tricep", "Serum.Insulin", "BMI", "Unknown", "Age", "Diabetic")
# Description of data on page 146
pima <- pima[,-c(1, 7)]

# Carry out fit with binomial regression
pima.fit <- glm(Diabetic ~ ., data = pima, family = "binomial")
#summary(pima.fit)

# To test the overall regression effect, we use LRT (pg 44)
pima.fit.null <- glm(Diabetic ~ 1, data = pima, family = "binomial")

#LRT
pima.lrt <- anova(pima.fit.null, pima.fit, test = "Chi")
kable(pima.lrt)
pima.lrt$`Pr(>Chi)`
library(MASS)
# We first fit with the AIC
pima.AIC <- stepAIC(pima.fit, trace = F)
```

```
pima.BIC <- stepAIC(pima.fit, trace = F, k = log(nrow(pima)))
kable(pima.AIC$coefficients)
kable(pima.BIC$coefficients)
pima.fit.red <- glm(Diabetic ~ PGC + BMI + Age, data = pima, family = "binomial")
kable(summary(pima.fit.red)$coefficients)
```