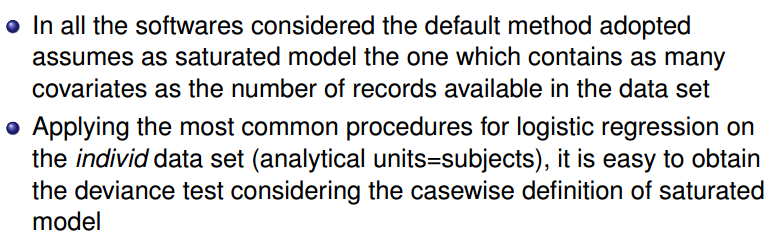
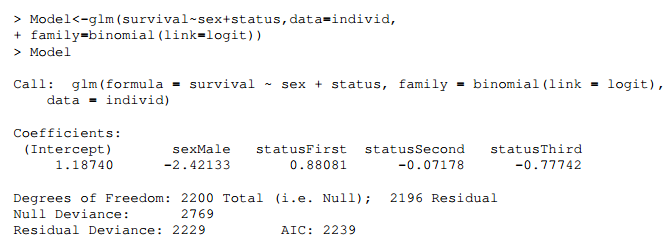
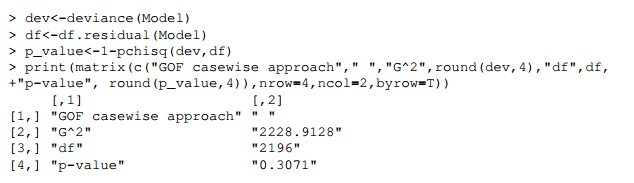
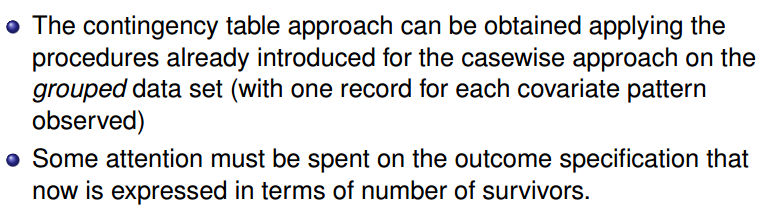


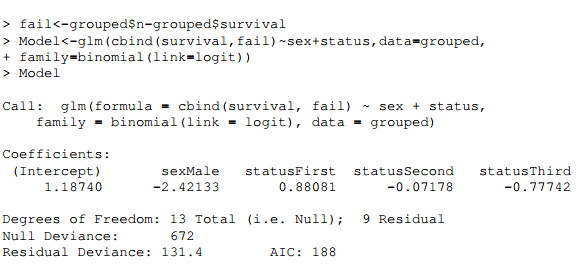
Likelihood Ratio Test implementation-casewise approach with R

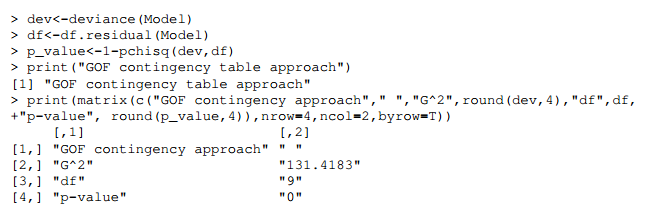


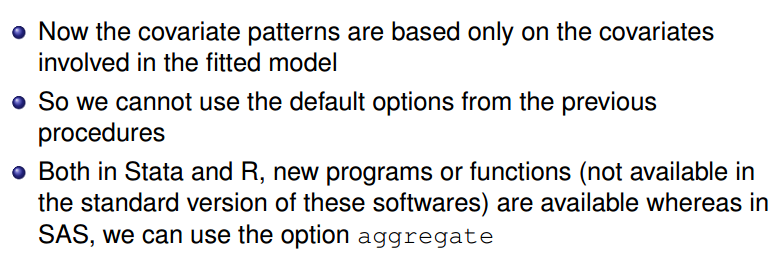


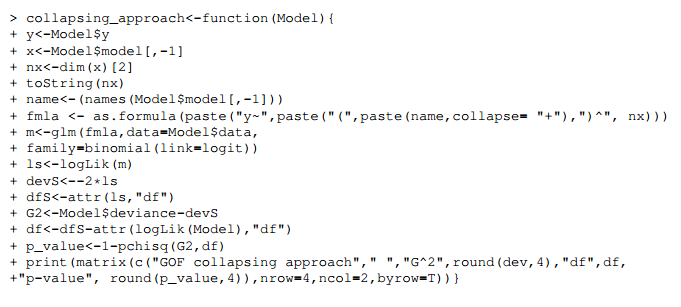


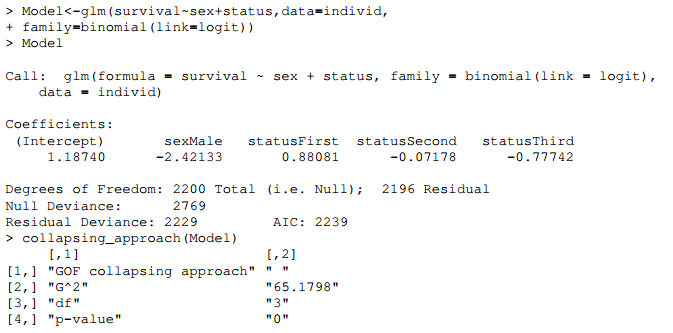
Likelihood Ratio Test implementation – Contingency approach 





Likelihood Ratio Test implementation – Collapsing approach 



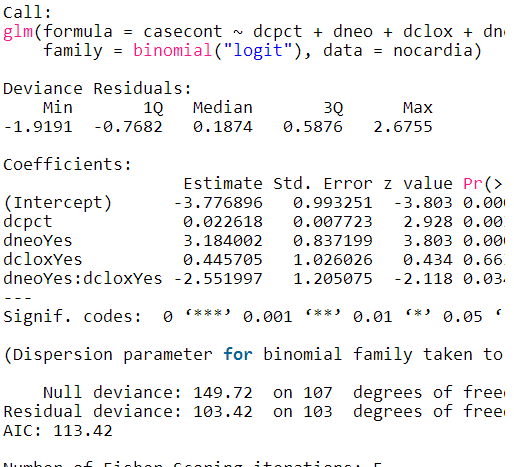


Example 2

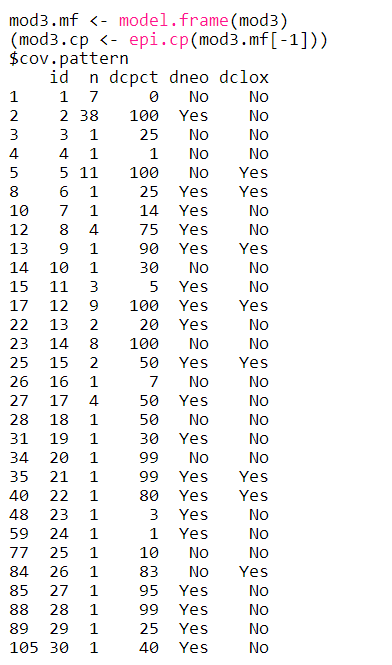
Two steps in assessing the fit of the model: first is to determine if the model fits using summary measures of goodness of fit or by assessing the predictive ability of the model; second is to determine if there’s any observations that do not fit the model or that have an influence on the model.

mod3 <- glm(casecont ~ dcpct + dneo + dclox + dneo\*dclox,family = binomial("logit"), data = nocardia)

summary(mod3)



library(epiR)



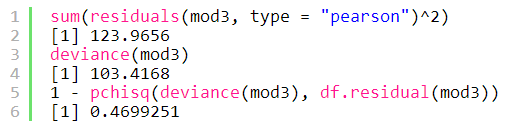
There are 30 covariate patterns in the dataset. The pattern dcpct=100, dneo=Yes, dclox=No appears 38 times.

**Pearson and deviance residuals**  
Residuals represent the difference between the data and the model. The Pearson residuals are comparable to standardized residuals used for linear regression models. Deviance residuals represent the contribution of each observation to the overall deviance.



**Goodness-of-fit test**  
All goodness-of-fit tests are based on the premise that the data will be divided into subsets and within each subset the predicted number of outcomes will be computed and compared to the observed number of outcomes. The Pearson \chi^2  and the deviance \chi^2  are based on dividing the data up into the natural covariate patterns. The Hosmer-Lemeshow test is based on a more arbitrary division of the data.

The Pearson \chi^2  is similar to the residual sum of squares used in linear models. It will be close in size to the deviance, but the model is fit to minimize the deviance and not the Pearson \chi^2 . It is thus possible even if unlikely that the \chi^2 could increase as a predictor is added to the model.



The p-value is large indicating no evidence of lack of fit. However, when using the deviance statistic to assess the goodness-of-fit for a nonsaturated logistic model, the \chi^2  approximation for the likelihood ratio test is questionable. When the covariate pattern is almost as large as N, the deviance cannot be assumed to have a \chi^2  distribution.

Now the Hosmer-Lemeshow test, usually dividing by 10 the data:

hosmerlem <- function (y, yhat, g = 10) {

cutyhat <- cut(yhat, breaks = quantile(yhat, probs = seq(0, 1, 1/g)),

                  include.lowest = TRUE)

obs <- xtabs(cbind(1 - y, y) ~ cutyhat)

expect <- xtabs(cbind(1 - yhat, yhat) ~ cutyhat)

chisq <- sum((obs - expect)^2 / expect)

P <- 1 - pchisq(chisq, g - 2)

c("X^2" = chisq, Df = g - 2, "P(>Chi)" = P)

}

hosmerlem(y = nocardia$casecont, yhat = fitted(mod3))

Erreur dans cut.default(yhat, breaks = quantile(yhat, probs = seq(0, 1, 1/g)),  (depuis #2) :

  'breaks' are not unique

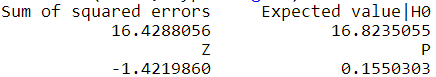
The model used has many ties in its predicted probabilities (too few covariate values?) resulting in an error when running the Hosmer-Lemeshow test. Using fewer cut-points (g = 5 or 7) does not solve the problem. This is a typical example when not to use this test. A better goodness-of-fit test than Hosmer-Lemeshow and Pearson / deviance \chi^2  tests is the [le Cessie – van Houwelingen – Copas – Hosmer unweighted sum of squares test for global goodness of fit](http://www.jstor.org/stable/2532385)(also [here](http://onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-0258%2819970515%2916:9%3C965::AID-SIM509%3E3.0.CO;2-O/abstract;jsessionid=693E96DAE9C9D28A0FA58C5F3D288520.d04t04)) implemented in the rms package (but you have to implement your model with the lrm function of this package):

mod3b <- lrm(casecont ~ dcpct + dneo + dclox + dneo\*dclox, nocardia,

method = "lrm.fit", model = TRUE, x = TRUE, y = TRUE,

linear.predictors = TRUE, se.fit = FALSE)

residuals(mod3b, type = "gof")





The p-value is 0.16 so there’s no evidence the model is incorrect. Even better than these tests would be to check for linearity of the predictors.