```r
set.seed(42)
```

Arbeitsblatt 9

Aufgabe 1

Aufgabe 2

```r
load("diab.RData")

log_reg <- glm(diabetes~pgc+bmi, family=binomial(link='logit'), data=diab)

pred = predict.glm(log_reg, data.frame(pgc=120, bmi=27), type="response")[1]
pred = as.vector(pred)

logit = function(L) {
  return(1/2*(1+tanh(L/2)))
}

pred2 = as.vector(logit(log_reg$coefficients[1]+log_reg$coefficients[2]
                        *120+log_reg$coefficients[3]*27))
delta = round((pred - pred2)^2, 4)
delta
#> [1] 0

pred
#> [1] 0.2254225
pred2
#> [1] 0.2254225
```

b)

```r
summary(log_reg)
#>
#> Call:
#> glm(formula = diabetes ~ pgc + bmi, family = binomial(link = "logit"),
#>     data = diab)
#>
#> Deviance Residuals:
#>     Min       1Q   Median       3Q      Max
#> -2.1803  -0.7752  -0.4693   0.7758   3.0164
#>
#> Coefficients:
#>              Estimate Std. Error z value Pr(>|z|)
#> (Intercept) -7.515639   0.605236 -12.418  < 2e-16 ***
#> pgc          0.035169   0.003289  10.694  < 2e-16 ***
#> bmi          0.076334   0.013338   5.723 1.05e-08 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
```

```
#>     Null deviance: 993.48  on 767  degrees of freedom
#> Residual deviance: 771.40  on 765  degrees of freedom
#> AIC: 777.4
#>
#> Number of Fisher Scoring iterations: 4
```

Die Koeffizientenschätzung der Variablen pgc ist $\beta = 0.035169$, ein positiver Wert. Dies bedeutet, dass ein Anstieg der Glukosekonzentration mit einer Erhöhung der Wahrscheinlichkeit verbunden ist, an Diabetes erkrankt zu sein.

c)

Ein wichtiges zu verstehendes Konzept für die Interpretation der logistischen Beta-Koeffizienten ist die Odds Ratio. Die Odds Ratio misst die Zuordnung zwischen einer Prädiktorvariablen (x) und der Ergebnisvariablen (y). Es stellt das Verhältnis der Chancen dar, dass ein Ereignis eintritt (Ereignis = 1) bei Vorhandensein des Prädiktors x (x = 1), verglichen mit den Chancen des Ereignisses, das in Abwesenheit dieses Prädiktors eintritt (x = 0).

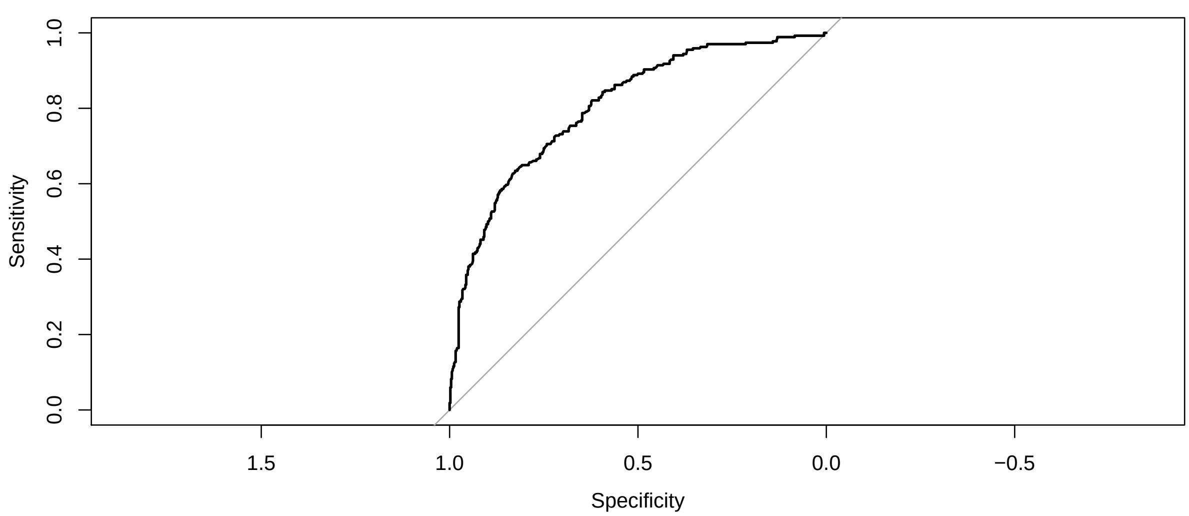
Für einen gegebenen Prädiktor (z.B. x1) entspricht der zugehörige Beta-Koeffizient (b1) in der logistischen Regressionsfunktion dem Logarithmus der Odds Ratios für diesen Prädiktor.

Wenn die Odds Ratio 2 ist, dann sind die Chancen, dass das Ereignis eintritt (Ereignis = 1), doppelt so hoch, wenn der Prädiktor x vorhanden ist (x = 1) als für ein Fehlen des Prädiktors (x = 0).

So beträgt beispielsweise der Regressionskoeffizient für BMI 0.0763342. Dies deutet darauf hin, dass eine Erhöhung des BMI um 5 Einheiten die Wahrscheinlichkeit, an Diabetes erkrankt zu sein, um das 4.6325327-fache erhöht $\left(\frac{5}{exp(\beta_{bmi})}\right)$.

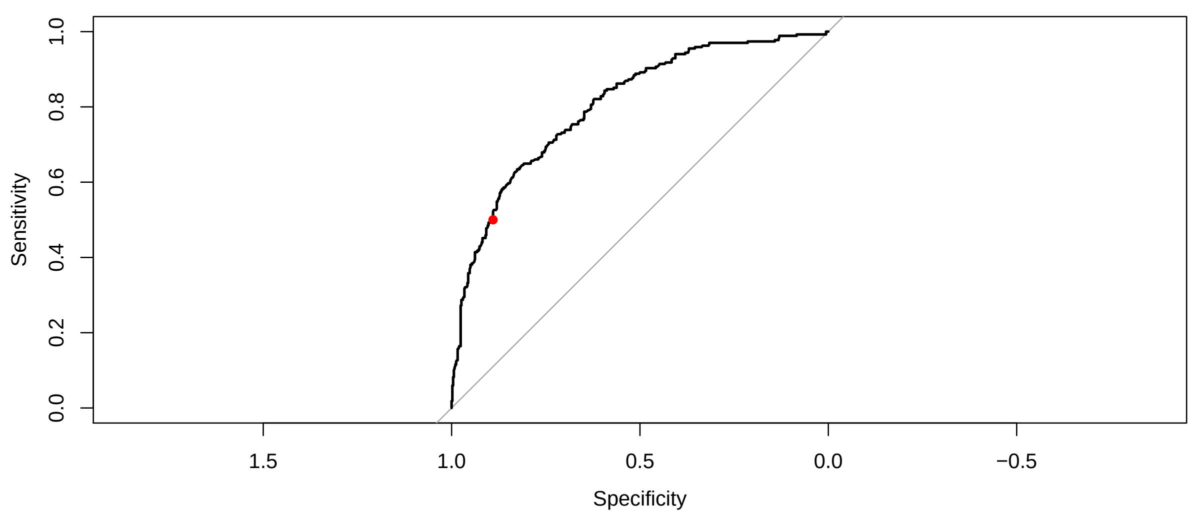
d)

```r
packageTest("pROC")

plot.roc(diab$diabetes, diab$pgc+diab$bmi)
```



e)

```
test <- NULL
test$y.true <- diab$diabetes
test$y.pred <- predict.glm(log_reg, type="response")

test$y.pred[test$y.pred > 0.5] = 1
test$y.pred[test$y.pred <= 0.5] = 0

confusion_matrix <- table(test)
confusion_matrix
#>         y.pred
#> y.true    0   1
#>       0 445  55
#>       1 126 142

misclass_rate <- 1-sum(diag(confusion_matrix))/sum(confusion_matrix)

sensitivity <- confusion_matrix[2,2]/(confusion_matrix[2,2] + confusion_matrix[2,2])

specificity <- confusion_matrix[1,1]/(confusion_matrix[1,1] + confusion_matrix[1,2])

misclass_rate
#> [1] 0.2356771
sensitivity
#> [1] 0.5
specificity
#> [1] 0.89

plot.roc(diab$diabetes, diab$pgc+diab$bmi)
points(specificity, sensitivity, col="red", pch = 16)
```



Aufgabe 3