

AS17: Kapitel 4 - Wahrscheinlichkeiten

Jürgen Hedderich

2020-04-13

Contents

Aktuelle R-Umgebung zu den Beispielen	2
4.3 Bedingte Wahrscheinlichkeiten	3
4.3.1 Beispiel: Bridge	3
4.3.2 Beispiel: Geburtstagsproblem	3
4.5 Diagnostischer Test	5
4.5.1 ROC-Analyse	7
4.5.2 Der Likelihood-Quotient (Fagan-Nomogramm)	9
4.5.3 Entscheidungsanalyse nach A.J. Vickers	10

Hinweis: Die Gliederung zu den Befehlen Und Ergebnissen in R orientiert sich an dem Inhaltsverzeichnis der 17. Auflage der ‘Angewandten Statistik’. Nähere Hinweise zu den verwendeten Formeln sowie Erklärungen zu den Beispielen sind in dem Buch (Ebook) nachzulesen!

zur Druckversion

Hinweis: Die thematische Gliederung zu den aufgeführten Befehlen und Beispielen orientiert sich an dem Inhaltsverzeichnis der 17. Auflage der ‘Angewandten Statistik’. Nähere Hinweise zu den verwendeten Formeln sowie Erklärungen zu den Beispielen sind in dem Buch (Ebook) nachzulesen!

Aktuelle R-Umgebung zu den Beispielen

The R Project for Statistical Computing

```
sessionInfo()
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18363)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.1252 LC_CTYPE=German_Germany.1252
## [3] LC_MONETARY=German_Germany.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## loaded via a namespace (and not attached):
## [1] compiler_3.6.3  magrittr_1.5    tools_3.6.3    htmltools_0.4.0
## [5] yaml_2.2.1      Rcpp_1.0.4      stringi_1.4.6  rmarkdown_2.1
## [9] knitr_1.28      stringr_1.4.0   xfun_0.12      digest_0.6.25
## [13] rlang_0.4.5     evaluate_0.14
```

4.3 Bedingte Wahrscheinlichkeiten

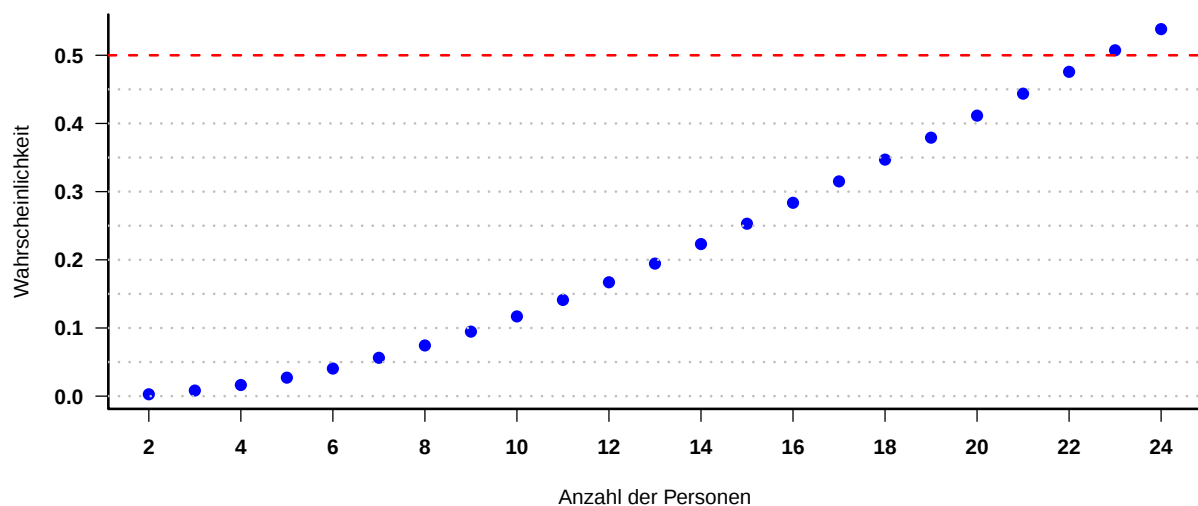
4.3.1 Beispiel: Bridge

```
P.A1 <- choose(4,1)*choose(48, 12) / choose(52, 13); P.A1
## [1] 0.4388475
P.A2 <- choose(3,1)*choose(36, 12) / choose(39, 13); P.A2
## [1] 0.4623044
P.A3 <- choose(2,1)*choose(24, 12) / choose(26, 13); P.A3
## [1] 0.52
P.A4 <- 1
P.A1 * P.A2 * P.A3 * P.A4
## [1] 0.1054982
```

4.3.2 Beispiel: Geburtstagsproblem

```
par(mfcol=c(1,1), lwd=2, font.axis=2, bty="l", ps=13)
k <- 24; p <- numeric(k)                                # Anzahl der Personen
for (i in 1:k) {                                         # 1 - W. keine Übereinstimmung
  q <- 1 - (0:(i - 1))/365
  p[i] <- 1 - prod(q) }

par(mfcol=c(1,1), lwd=2, font.axis=2, bty="l", ps=12)
plot(2:k, p[2:k], xaxp=c(2, k, (k-2)/2), las=1, pch=16, cex=1.3,
     xlab="Anzahl der Personen", ylab="Wahrscheinlichkeit", col="blue")
abline(h=0.5, col="red", lty=2)
abline(h=seq(0,0.45,0.05), col="grey", lty=3)
```



```
qbirthday(prob = 0.5, classes = 365, coincident = 2)      # Tabelle
## [1] 23
qbirthday(prob = 0.5, classes = 365, coincident = 3)
## [1] 88
qbirthday(prob = 0.5, classes = 365, coincident = 4)
## [1] 187
qbirthday(prob = 0.5, classes = 365, coincident = 5)
## [1] 313
qbirthday(prob = 0.5, classes = 365, coincident = 6)
## [1] 459
```

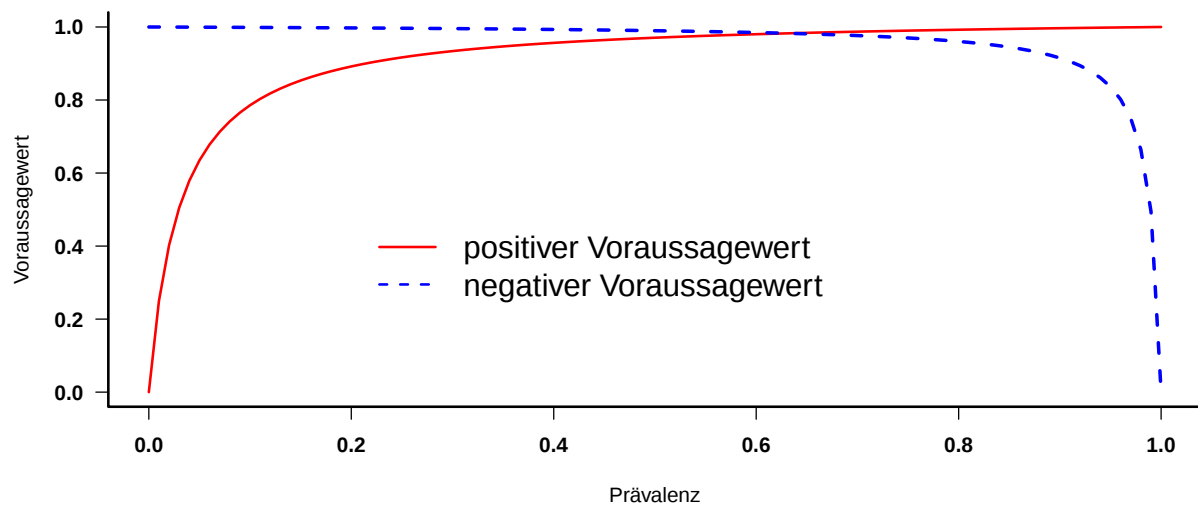
4.5 Diagnostischer Test

Sensitivität / Spezifität

```
sens <- 0.99                                # Sensitivität
spez <- 0.97                                # Spezifität
prev <- seq(0, 1, by=0.01)                  # Prävalenz

ppw <- (sens * prev) / (sens * prev + (1 - spez) * (1 - prev))
npw <- (spez * (1 - prev)) / (spez * (1 - prev) + (1 - sens) * prev)

par(mfcol=c(1,1), lwd=2, font.axis=2, bty="l", ps=12)
plot(prev, ppw, type="l", col="red", xlab="Prävalenz", las=1, cex=1.3,
      ylab="Voraussagewert", lty=1, lwd=2)
lines(prev, npw, col="blue", lty=2, lwd=2.7)
legend(0.2, 0.5, c("positiver Voraussagewert",
                   "negativer Voraussagewert"), cex=1.5,
      col=c("red", "blue"), bty="n", lty=c(1,2))
```



Ausbreitung von Krankheiten

```
pfn <- function(n, prev, sens, spec=1) {
  npv <- (1-prev)*spec / ((1-prev)*spec + prev*(1-sens))
  p    <- 1- npv^n; return(round(100*p, 1)) }

pfn(n=c(1000, 5000), prev=0.01, sens=0.95, spec=1.00)
## [1] 39.6 92.0
pfn(n=c(1000, 5000), prev=0.01, sens=0.95, spec=0.90)
## [1] 42.9 93.9

ni    <- c(500, 1000, 2000, 5000)           # Tabelle
previ <- c(0.02, 0.01)
sensi <- c(0.90, 0.95, 0.99)
tab <- matrix(rep(NA, 36), nrow=6, byrow=T)
k    <- 0
for (i in 1:3) {
  for (j in 1:2) {
    k <- k + 1
    tab[k,] <- c(sensi[i], previ[j], pfn(ni, previ[j], sensi[i], spec=1))
  } }

colnames(tab) <- c("Sensitivität", "Prävalenz", "500", "1000", "2000", "5000")
tab
##      Sensitivität Prävalenz  500 1000 2000 5000
## [1,]          0.90      0.02 63.9 87.0 98.3 100.0
## [2,]          0.90      0.01 39.6 63.6 86.7 99.4
## [3,]          0.95      0.02 39.9 63.9 87.0 99.4
## [4,]          0.95      0.01 22.3 39.6 63.6 92.0
## [5,]          0.99      0.02  9.7 18.5 33.5 64.0
## [6,]          0.99      0.01  4.9  9.6 18.3 39.7
```

4.5.1 ROC-Analyse

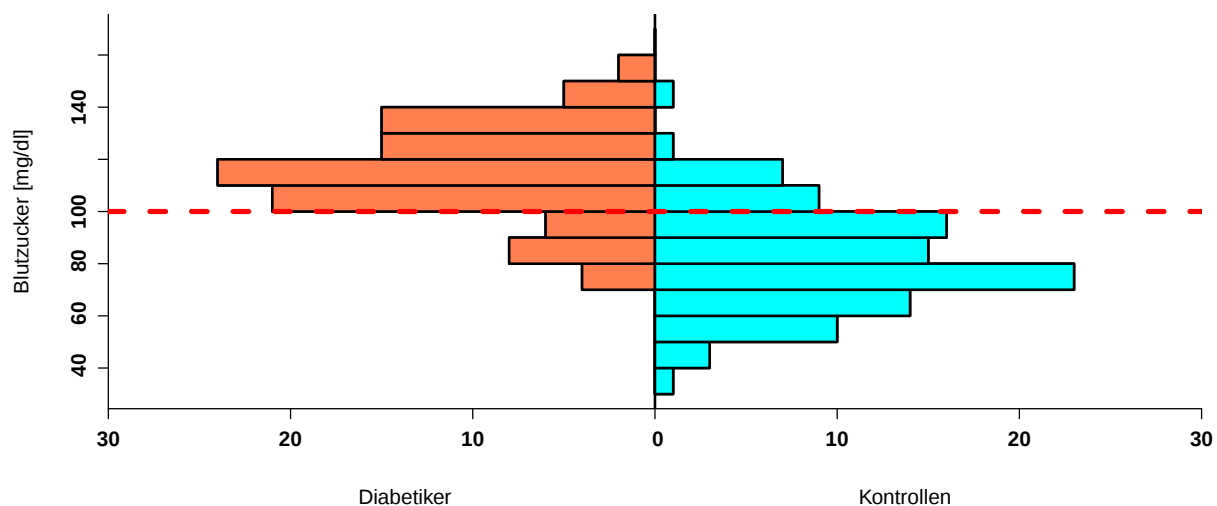
Funktion `histbackback()` in `library(Hmisc)`

Funktion `prediction()` und `performance()` in `library(ROCR)`

```
library(Hmisc)
library(ROCR)

# Blutzucker [mg/dl]
blz.diab <- c(
  119, 89, 136, 89, 107, 137, 126, 101, 105, 134, 131, 133, 117, 102, 108, 112, 94, 90,
  102, 74, 119, 130, 105, 111, 130, 114, 110, 124, 71, 138, 96, 112, 121, 124, 102, 109,
  108, 130, 89, 114, 153, 109, 103, 77, 133, 122, 87, 102, 130, 120, 118, 100, 114, 138,
  136, 111, 100, 100, 118, 127, 105, 126, 127, 81, 117, 104, 105, 84, 149, 132, 104, 114,
  116, 142, 110, 139, 131, 120, 114, 156, 106, 112, 125, 112, 89, 143, 108, 133, 139, 116,
  115, 149, 94, 128, 80, 119, 122, 148, 117, 135)
blz.cntr <- c(
  93, 78, 42, 70, 109, 75, 80, 73, 115, 100, 60, 58, 70, 53, 94, 80, 71, 95,
  71, 127, 66, 75, 68, 92, 66, 108, 117, 41, 80, 71, 86, 68, 75, 87, 110, 104,
  53, 82, 120, 89, 77, 77, 78, 52, 65, 101, 69, 53, 61, 99, 73, 89, 102, 78,
  110, 113, 143, 73, 87, 88, 76, 90, 75, 102, 60, 62, 115, 69, 113, 97, 78, 98,
  62, 86, 81, 96, 61, 60, 91, 98, 94, 87, 59, 94, 47, 95, 80, 114, 74, 73,
  55, 85, 91, 85, 31, 84, 106, 91, 65, 84)

par(mfcol=c(1,1), lwd=2, font.axis=2, bty="n", ps=12, las=0, col="black")
options(digits=0);
out <- histbackback(blz.diab, blz.cntr, brks=seq(30, 170, by=10),
  axes=TRUE, las=1, cex.axis=1.5,
  xlab=c("Diabetiker", "Kontrollen"),
  ylab="Blutzucker [mg/dl]")
barplot(-out$left, col="coral", horiz=TRUE, space=0, add=TRUE, axes=FALSE)
barplot(out$right, col="cyan", horiz=TRUE, space=0, add=TRUE, axes=FALSE)
abline(h=7, col="red", lwd=4, lty=2)
```



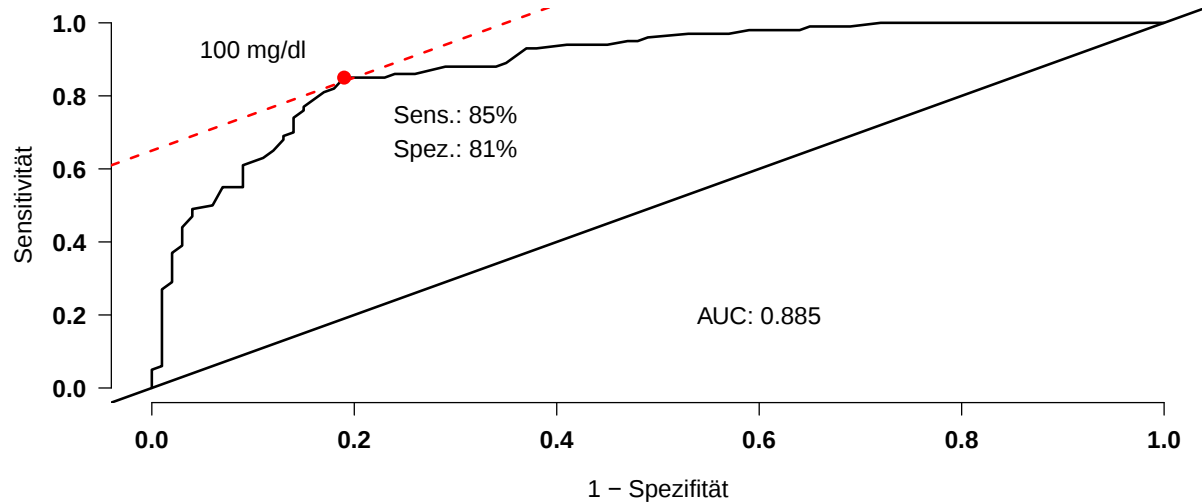
```

blz <- c(blz.diab, blz.cntr)
grp <- c(rep(1, length(blz.diab)), rep(0, length(blz.cntr)))
pred <- prediction(blz, grp)

roc <- performance(pred, "tpr", "fpr")

par(mfcol=c(1,1),lwd=2, font.axis=2, bty="n", ps=14, las=1)
plot(roc, xlab="1 - Spezifität", ylab="Sensitivität")
abline(0,1); abline(0.65, 1, lty=2.5, col="red")
points(0.19, 0.85, pch=16, cex=1.5, col="red")
text(0.3, 0.75, "Sens.: 85%"); text(0.3, 0.65, "Spez.: 81%")
text(0.6, 0.20, "AUC: 0.885"); text(0.1, 0.920, "100 mg/dl")

```

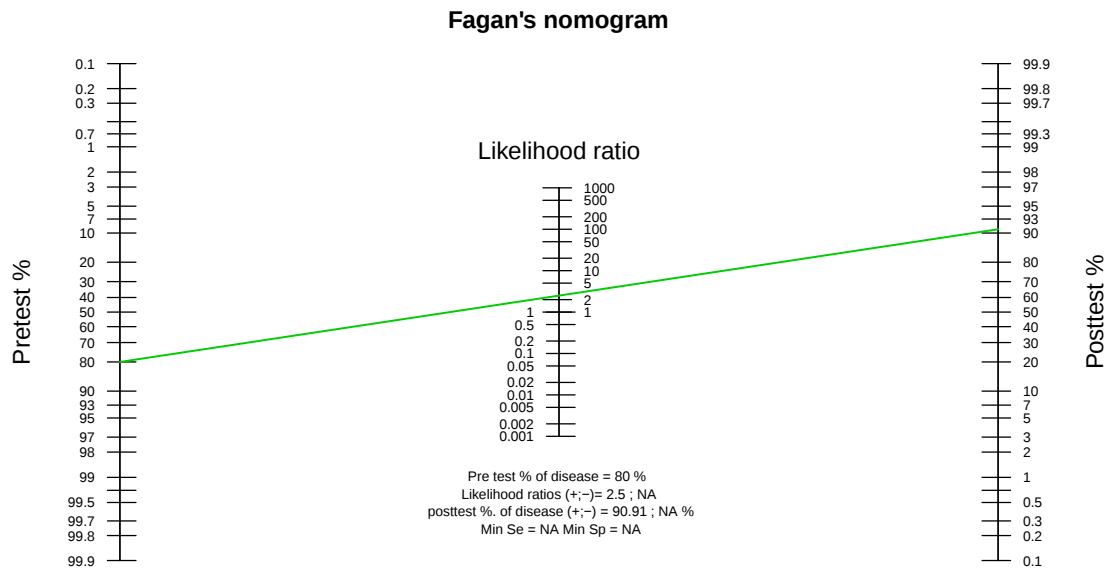


4.5.2 Der Likelihood-Quotient (Fagan-Nomogramm)

Funktion `nomogram()` in `library(UncertainInterval)`

```
library(UncertainInterval)

par(opar)
nomogram(prob.pre.test = .80, LR=c(pos=2.5, neg=NULL))
```



4.5.3 Entscheidungsanalyse nach A.J. Vickers

```
sens <- 0.85
spec <- 0.81
prev <- 0.50

p_t <- seq(0, 1, by=0.01)
n_b <- prev * sens - (1-prev) * (1-spec) * (p_t/(1-p_t))
tempnb <- prev - (1-prev) * p_t / (1 - p_t)

par(mfcol=c(1,1),lwd=2, font.axis=2, bty="n", ps=12, las=0, col="black")
plot(p_t, n_b, type="l", ylim=c(-0.1, prev+0.01), xlim=c(0,1), lty=1, las=1,
     col="black", ylab="(möglicher) Nutzen", xlab="Schwellenwahrscheinlichkeit")
abline(h=0, lty=2, col="darkgrey")
lines(p_t, tempnb, lty = 3, col="darkgrey")
legendlabel <- c("Modell", "null", "alle")
legendcolor <- c("black", "darkgrey", "darkgrey")
legendpattern <- c(1,2,3)
legend("topright", legendlabel, cex=0.9, bty="n", col=legendcolor, lty=legendpattern)
```

