

Дополнительные материалы к статье

Глазков А.А., Куликов Д.А., Глазкова П.А.

Математическая биология и биоинформатика. 2020. Т. 15. № 2. С. 416–428.

doi: [10.17537/2020.15.416](https://doi.org/10.17537/2020.15.416)

ПРИЛОЖЕНИЕ

В приложении приведён пример кода на языке R, который по набору статистик (M_1, SD_1, M_2, SD_2) и заданному шагу по оси значений признаков восстанавливает выборку значений признака и параметров соответствующей ROC-кривой. Код применялся для создания таблицы со следующими столбцами:

X – в столбце содержатся значения признака x с шагом dx ;

TPR – столбец содержит значения $TPR(x)$;

FPR – столбец содержит значения $FPR(x)$;

D – столбец содержит расстояния от соответствующей текущему x точки на ROC-кривой до точки (0, 1);

S – площадь участка ROC-кривой в интервале x_i, x_{i+1} .

```
Diagnostic_characteristics_table <- function(M1, SD1, M2, SD2, dx) {
```

```
  #Определение границ оценки значений x с шагом dx
```

```
  x_start <- floor(min((M1-3*SD1)/dx, (M2-3*SD2)/dx)) * dx
```

```
  x_end <- floor(max((M1+3*SD1)/dx, (M2+3*SD2)/dx)) * dx
```

```
  #Определение количества оцениваемых интервалов
```

```
  N <- (x_end-x_start)/dx
```

```
  #Создание 5 пустых векторов для создания финальной таблицы
```

```
  X <- c()
```

```
  TPR <- c()
```

```
  FPR <- c()
```

```
  D <- c()
```

```
  S <- c()
```

```
  #Заполнение векторов значениями по формуле (1)
```

```
  if (M1 > M2){
```

```
    for (i in 1:(N+1)) {
```

```
      X[i] <- x_end-(i-1)*dx
```

```
      TPR[i] <- pnorm(X[i], M1, SD1, lower.tail = F)
```

```
      FPR[i] <- pnorm(X[i], M2, SD2, lower.tail = F)
```

```
      D[i] <- sqrt((1-TPR[i])^2 + FPR[i]^2)
```

```
    }
```

```
  } else {
```

```
    for (i in 1:(N+1)) {
```

```
      X[i] <- x_start+(i-1)*dx
```

```
      TPR[i] <- pnorm(X[i], M1, SD1)
```

```
      FPR[i] <- pnorm(X[i], M2, SD2)
```

```
      D[i] <- sqrt((1-TPR[i])^2 + FPR[i]^2)
```

```
    }
```

```
  }
```

```
  #Расчёт площадей под трапециями для N секторов ROC-кривой
```

```

S[N+1] <- 0
for (i in 1:N){
  S[i] <- abs(TPR[i]*(FPR[i+1]-FPR[i])+
             (TPR[i+1]-TPR[i])*(FPR[i+1]-FPR[i])/2)
}

Result_table <- data.frame(X, TPR, FPR, D, S)
return(Result_table)
}

#Пример работы функции:
Table1 <- Diagnostic_characteristics_table(19, 1.8, 18, 2.2, 0.1)

#Площадь под ROC-кривой вычисляется как сумма значений в столбце S:
AUC <- sum(Table1$S)

#Оптимальное пороговое значение можно определить следующим образом:
Cut_off <- Table1$X[Table1$D == min(Table1$D)]

```