Breast Cancer Data Analysis

Chris Roberts

## Frequency Distribution problem

The breast-cancer-wisconsin.csv file contains features computed from digitized images of fine needle aspirate (FNA) of breast mass. They describe characteristics of the cell nuclei present in the image.The first column is the ID number. The other columns are the ten real-valued features computed for each cell nucleus:

1. radius (mean of distances from center to points on the perimeter)
2. texture (standard deviation of gray-scale values)
3. perimeter
4. area
5. smoothness (local variation in radius lengths)
6. compactness (perimeter^2 / area - 1.0)
7. concavity (severity of concave portions of the contour)
8. concave points (number of concave portions of the contour)
9. symmetry
10. fractal dimension (“coastline approximation” - 1)

#R read the breast-cancer-wisconsin.csv file using the following command.  
bcData=read.csv("breast-cancer-wisconsin.csv", header =T)  
#Values from each column can be extracted using the $ operator. For example  
head(bcData,10)

## Idnumber radius texture perimeter area smoothness compactness concavity  
## 1 1000025 5 1 1 1 2 1 3  
## 2 1002945 5 4 4 5 7 10 3  
## 3 1015425 3 1 1 1 2 2 3  
## 4 1016277 6 8 8 1 3 4 3  
## 5 1017023 4 1 1 3 2 1 3  
## 6 1017122 8 10 10 8 7 10 9  
## 7 1018099 1 1 1 1 2 10 3  
## 8 1018561 2 1 2 1 2 1 3  
## 9 1033078 2 1 1 1 2 1 1  
## 10 1033078 4 2 1 1 2 1 2  
## concavePoints symmetry fractalDimension  
## 1 1 1 2  
## 2 2 1 2  
## 3 1 1 2  
## 4 7 1 2  
## 5 1 1 2  
## 6 7 1 4  
## 7 1 1 2  
## 8 1 1 2  
## 9 1 5 2  
## 10 1 1 2

### Write R statements to calculate the frequency (f), relative frequency (rf), cumulative frequency(cf) and cumulative relative frequency (crf) for each of the following features from the breast cancer data file:

1. Radius
2. Smoothness
3. Concavity
4. Symmetry

### Create a frequency table for each feature by combing (f,rf,cf,crf).Plot a distribution for each feature. In addition to the .Rmd and the docx files, you need to upload the following files to canvas.

1. yourLastName\_radius.csv
2. yourLastName\_smoothness.csv
3. yourLastName\_concavity.csv
4. yourLastName\_symmetry.csv

## Frequency distribution calculation for Radius.

# Your R code for calculation goes here.  
bcData=read.csv("breast-cancer-wisconsin.csv", header =T)  
  
bc.radius.f = table(bcData$radius)  
bc.radius.rf = (bc.radius.f)/sum(bc.radius.f)\*100  
bc.radius.cf = cumsum(bc.radius.f)  
bc.radius.crf = (bc.radius.cf)/sum(bc.radius.cf)\*100

## Frequency distribution calculation for Smoothness.

# Your R code for calculation goes here.  
bc.smoothness.f = table(bcData$smoothness)  
bc.smoothness.rf = (bc.smoothness.f)/sum(bc.smoothness.f)\*100  
bc.smoothness.cf = cumsum(bc.smoothness.f)  
bc.smoothness.crf = (bc.smoothness.cf)/sum(bc.smoothness.cf)\*100

## Frequency distribution calculation for concavity

# Your R code for calculation goes here.  
bc.concavity.f = table(bcData$concavity)  
bc.concavity.rf = (bc.concavity.f)/sum(bc.concavity.f)\*100  
bc.concavity.cf = cumsum(bc.concavity.f)  
bc.concavity.crf = (bc.concavity.cf)/sum(bc.concavity.cf)\*100

## Frequency distribution calculation for symmetry.

# Your R code for calculation goes here.  
bc.symmetry.f = table(bcData$symmetry)  
bc.symmetry.rf = (bc.symmetry.f)/sum(bc.symmetry.f)\*100  
bc.symmetry.cf= cumsum(bc.symmetry.f)  
bc.symmetry.crf = (bc.symmetry.cf)/sum(bc.symmetry.cf)\*100