

Annotation Data Analysis

Iván Pérez

MFF

NPFL054
March 8th, 2023

Exercise A: Preprocessing

```
##      NAME NROW NCOL MB      COLS KEY
## 1: cry.A   60    2  0 id,class id
## 2: cry.B   60    2  0 id,class id
## Total: 0MB
```

```
### join the tables using the common index
```

```
cry.AB = cry.A[cry.B]
setnames(cry.AB, c("id", "A", "B"))
```

```
### now simply make a table with the IAA
```

```
tAB = table(cry.AB[, c(2,3), with=F])
```

```
tAB
```

```
##      B
## A    1  4  7  u  x
##    1 25  3  1  4  1
##    4  4  6  2  2  0
##    7  1  1  3  1  0
##    u  0  1  0  0  0
##    x  1  0  2  1  1
```

Exercise A: Pr(a)

```
##      B
## A      1  4  7  u  x
##      1 25  3  1  4  1
##      4  4  6  2  2  0
##      7  1  1  3  1  0
##      u  0  1  0  0  0
##      x  1  0  2  1  1
```

```
Pr_a = sum(diag(tAB))/sum(tAB)
Pr_a
```

```
## [1] 0.5833333
```

Exercise A: $\Pr(e)$

```
##      B
## A    1  4  7  u  x
##    1 25  3  1  4  1
##    4  4  6  2  2  0
##    7  1  1  3  1  0
##    u  0  1  0  0  0
##    x  1  0  2  1  1
```

$$p_e = \sum_k p_{k12} = \sum_k p_{k1} \cdot p_{k2} = \sum_k \frac{n_{k1}}{N} \cdot \frac{n_{k2}}{N} = \frac{1}{N^2} \sum_k n_{k1} \cdot n_{k2}$$

```
Pr_e = table(cry.A$class) %*% table(cry.B$class) / nrow(cry.AB)^2
Pr_e
```

```
##           [,1]
## [1,] 0.3538889
```

Exercise A: Cohen's kappa

```
##      B
## A    1  4  7  u  x
##    1 25  3  1  4  1
##    4  4  6  2  2  0
##    7  1  1  3  1  0
##    u  0  1  0  0  0
##    x  1  0  2  1  1
```

$$\kappa = \frac{p_a - p_e}{1 - p_e}$$

```
ckappa = (Pr_a - Pr_e) / (1 - Pr_e)
ckappa
```

```
##           [,1]
## [1,] 0.3551161
```

Exercise B: Confusion matrix

```
### join the tables using the common index
cry.F1GS = cry.F1[cry.GS]
setnames(cry.F1GS, c("id", "F1", "GS"))

### now simply make a table with the IAA
tF1GS = table(cry.F1GS[, c(2,3), with=F])

tF1GS
```

```
##      GS
## F1    1   4   7   u   x
##    1 125   9   0   6   8
##    4   2  48   0   2   1
##    7   0   0  12   1   0
##    u   4   1   1  22   0
##    x   0   1   0   2   5
```

Exercise B: Classifier accuracy

```
##      GS
## F1    1   4   7   u   x
##    1 125   9   0   6   8
##    4   2  48   0   2   1
##    7   0   0  12   1   0
##    u   4   1   1  22   0
##    x   0   1   0   2   5
```

```
P_a = sum(diag(tF1GS))/sum(tF1GS)
P_a
```

```
## [1] 0.848
```

Exercise B: Percentages

```
##      GS
## F1    1   4   7   u   x
##  1 125   9   0   6   8
##  4   2  48   0   2   1
##  7   0   0  12   1   0
##  u   4   1   1  22   0
##  x   0   1   0   2   5
```

```
perc = round(100*prop.table(tF1GS, margin = 2),1)
perc
```

```
##      GS
## F1    1   4   7   u   x
##  1 95.4 15.3  0.0 18.2 57.1
##  4  1.5 81.4  0.0  6.1  7.1
##  7  0.0  0.0 92.3  3.0  0.0
##  u  3.1  1.7  7.7 66.7  0.0
##  x  0.0  1.7  0.0  6.1 35.7
```


Thank you!