Week 5: Functions

Quantitative Textanalyse mit R

Nils Reiter

May 22, 2020

R so far

- Vector, list, matrix, data.frame
  - Vectors are 1-based (i.e., start with 1)
- Types: numeric, character, logical
- Expressions: Assignments of values to variable names
- Selecting elements in containers: [ ]

R so far

- Vector, list, matrix, data.frame
  - Vectors are 1-based (i.e., start with 1)
- Types: numeric, character, logical
- Expressions: Assignments of values to variable names
- Selecting elements in containers: [ ]

```r
(1:100)[c(TRUE,FALSE)] # every second element
m <- matrix(1:100, nrow=25)
m[4,5] # row 4 col 5
m[4,]  # row 4
m[,5]  # col 5
```

Control structures

Grouped expressions: $\texttt{expr}_1, expr_2, ..., expr_n$ `expr_1, expr_2, ..., expr_n`

- Commands can be grouped with curly braces
- The value of the whole group is the value of the last expression within the group
- Grouped expressions do not form a scope on their own

Control structures

Grouped expressions: $\mathtt{expr}_1, expr_2, ..., expr_n$ `expr_1, expr_2, ..., expr_n`

- Commands can be grouped with curly braces
- The value of the whole group is the value of the last expression within the group
- Grouped expressions do not form a scope on their own

Conditional execution: `if (expr_1) expr_2 else expr_3`

- `expr_2` only executed if `expr_1` evaluates to a single logical TRUE value
- Implicit use of `as.logical(expr_1)[1]`
- Warning if `length(expr_1)>1`

Control structures

Loops: `for (name in expr) expr_2`

- `expr_2` repeatedly evaluated, while `name` is set to the elements of `expr`
- In R, no difference between ‘foreach’ and ‘for’
- `for i in 1:100 print(i)` shows their equivalence

Control structures

Loops: `for (name in expr) expr_2`

- `expr_2` repeatedly evaluated, while `name` is set to the elements of `expr`
- In R, no difference between 'foreach' and 'for'
- `for i in 1:100 print(i)` shows their equivalence
- Loops are used *much less* in R compared to other languages
  - Most operations work in a *vectorized* fashion: We call a function that gets executed to every element of the vector
  - This vectorization is much faster, because it can be processed en bloc
  - Rule of thumbs: Avoid writing for-loops
  - There is even a vectorized version of if: `ifelse(condition,a,b)`

Functions

- Functions are mini-programs
- Named collections of commands
- Similar to other programming languages
- Syntax
  - Function definition: `name <- function(arg_1, arg_2, ...) expression`
  - Function call: `name(arg_1, arg_2, ...)`

Functions

- Functions are mini-programs
- Named collections of commands
- Similar to other programming languages
- Syntax
  - Function definition: `name <- function(arg_1, arg_2, ...) expression`
  - Function call: `name(arg_1, arg_2, ...)`
  - If the name starts and ends with %, and the function takes two arguments, it can be used as a binary operator

```
"%add%" <- function(x,y) {sum(x,y)}
5 %add% 7 # returns 12
```

Example

- Accuracy: Percentage of correctly classified instances
- Input: Two lists of labels
- Output: Number

```
accuracy <- function(gold, system) {
    # let's do that together
}
```

Example

- Accuracy: Percentage of correctly classified instances
- Input: Two lists of labels
- Output: Number

```
accuracy <- function(gold, system)
    sum(gold == system)/length(gold)
```

Any remaining issues? What could go wrong?

Example

Edge cases

Not the same length

- We should verify that the vectors have the same length

Different types

- Types are coerced to character if necessary
  - Depending on use case: Maybe issue a warning

Different factors

- Not a problem, factor levels can be compared as character vectors

Named Arguments

- Function arguments can be named
- Named arguments can be given in any order
  - If both named and positional arguments are mixed, positional arguments go first
- Using named arguments increases readability of your code

  ```
  accuracy(system=listOfLabels, gold=realLabels)
  ```

Default Values

- Specified in the function definition

```
accuracy <- function(gold, system, times=100)
    times * (sum(gold == system)/length)
```

- Can be defined based on other arguments (!)

```
accuracy <- function(gold, system, limit=1:length(gold))
    sum(gold[limit] == system[limit])/length(limit)
```

Passing on arguments

- Sometimes we want to pass on arguments to other functions, in particular if we write ‘wrapper functions’ (i.e., functions that wrap another function)
- In these cases, we can use the ... argument (yes, it’s three regular dots)
- All arguments not defined by our function are passed on to another function

```
partOfRandomVector <- function(n, x, ...) {
    v <- runif(n, ...)
    v[v<x]
}
```

We can now use all (named) arguments of `runif` as arguments for `partOfRandomVector` without further declaration

Anonymous functions (and apply)

- Similar to passing literal arguments to functions (as we have done before), we can pass anonymous functions
- They look the same, but are not assigned to a name:
  `function(arg_1, arg_2) expression`
- The function `apply()` expects a function as third argument
  - `apply()` applies this function to each row or column of a given matrix

Example

```
m <- matrix(1:100, ncol=5)
apply(m, 1, function(x) { sum(x**2) } )
```

Exercise 5

- At the usual place: `https://github.com/idh-cologne-sprachverarbeitung-mit-r/exercise-05`
  - Includes these slides
- Less exercises, but a bit more complicated

ATM

- When we withdraw money from an ATM, it has to decide which bills it gives
- We will implement this function in R
- The function `atm(v)` takes a numeric value and returns a named list
  - The names are the bill types (i.e., 50)
  - The values are the number of bills to give us
- How to decide?
  - We want as few bills as possible (this is not what real ATMs to ...)

ATM

- When we withdraw money from an ATM, it has to decide which bills it gives
- We will implement this function in R
- The function `atm(v)` takes a numeric value and returns a named list
  - The names are the bill types (i.e., 50)
  - The values are the number of bills to give us
- How to decide?
  - We want as few bills as possible (this is not what real ATMs to ...)

Example

```
> unlist(atm(50))
500 200 100  50  20  10   5
  0   0   0   1   0   0   0
> unlist(atm(195))
500 200 100  50  20  10   5
  0   0   1   1   2   0   1
```

Fleiss' Kappa

- Fleiss' Kappa is a metric to calculate inter-annotator agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- Observed agreement: How many agreements have been achieved?
- Expected agreement: How many agreements can be expected if everyone guesses?

Inter-Annotator Agreement

Observed Agreement

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j} - 1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j}-1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\hat{P}_i = \underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j} - 1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\hat{P}_i = \underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j}-1)}_{\text{abs. pairwise agr. for item } i}$$

Normalized observed agreement for all items

$$P_o = \frac{1}{N}\sum_{i=1}^{N} \hat{P}_i$$

(this is just the arithmetic mean, a.k.a. average)

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$\underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Inter-Annotator Agreement

Expected Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

- Probability that category j gets selected (by one annotator)

$$\underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Inter-Annotator Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Inter-Annotator Agreement

Expected Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

- Probability that two annotators select category j

$$p_j \times p_j = p_j^2$$

Inter-Annotator Agreement

Expected Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

- Probability that two annotators select category j

$$p_j \times p_j = p_j^2$$

- Probability that two annotators are in agreement (over all categories):

$$P_e = \sum_{j=1}^{k} p_j^2$$

Fleiss' Kappa (**Fleiss:1971aa**)

p_j	Probability for j
$\hat{P}_i$	Observed agreement for i

Fleiss' Kappa (**Fleiss:1971aa**)

p_j	Probability for j
$\hat{P}_i$	Observed agreement for i

$$P_o = \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{array}{lll}
p_j & & \text{Probability for } j \\
\hat{P}_i & & \text{Observed agreement for } i \\
P_o & = & \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \\
P_e & = & \sum_{j=1}^{k} p_j^2
\end{array}
$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{aligned}
p_j & \quad \text{Probability for } j \\
\hat{P}_i & \quad \text{Observed agreement for } i \\
P_o &= \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \\
P_e &= \sum_{j=1}^{k} p_j^2 \\
\kappa &= \frac{P_o - P_e}{1 - P_e}
\end{aligned}
$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{array}{rcl}
p_j & & \text{Probability for } j \\
\hat{P}_i & & \text{Observed agreement for } i \\
P_o & = & \dfrac{1}{N} \sum_{i=1}^{N} \hat{P}_i \\
P_e & = & \sum_{j=1}^{k} p_j^2 \\
\kappa & = & \dfrac{P_o - P_e}{1 - P_e}
\end{array}
$$

- $P_o - P_e$: Tatsächlich erreichtes, nicht-zufälliges Agreement
- $1 - P_e$: Maximal erreichbares, nicht-zufälliges Agreement

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{aligned}
p_j & \quad \text{Probability for } j \\
\hat{P}_i & \quad \text{Observed agreement for } i \\
P_o &= \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \\
P_e &= \sum_{j=1}^{k} p_j^2 \\
\kappa &= \frac{P_o - P_e}{1 - P_e}
\end{aligned}
$$

- $P_o - P_e$: Tatsächlich erreichtes, nicht-zufälliges Agreement
- $1 - P_e$: Maximal erreichbares, nicht-zufälliges Agreement
- $-\infty < \kappa < 1$: Je höher desto besser
  - Extremfälle?

Section 3

Fleiss' Kappa

Inter-Annotator Agreement

- Warum quantitativ messen?
  - Weil wir vergleichen wollen, über verschiedene Konfigurationen hinweg
    - (5 oder 6 Kategorien, 2 oder 3 AnnotatorInnen, 10 oder 20 Instanzen)

Inter-Annotator Agreement

- Warum quantitativ messen?
  - Weil wir vergleichen wollen, über verschiedene Konfigurationen hinweg
    - (5 oder 6 Kategorien, 2 oder 3 AnnotatorInnen, 10 oder 20 Instanzen)
- Wie messen?
  - Problem: Wir wissen nicht was richtig ist
    - IAA ist ausschließlich eine Aussage über die Übereinstimmung, nicht über die Korrektheit

Inter-Annotator Agreement

- Warum quantitativ messen?
  - Weil wir vergleichen wollen, über verschiedene Konfigurationen hinweg
    - (5 oder 6 Kategorien, 2 oder 3 AnnotatorInnen, 10 oder 20 Instanzen)
- Wie messen?
  - Problem: Wir wissen nicht was richtig ist
    - IAA ist ausschließlich eine Aussage über die Übereinstimmung, nicht über die Korrektheit
  - Verfahren das für beliebige Zahlen von Kategorien, AnnotatorInnen oder Instanzen funktioniert
  - Wie oft haben zwei AnnotatorInnen gleich gewählt? (paarweise Übereinstimmung)

Inter-Annotator Agreement

Datenstrukturen

Word	A1	A2	A3
Zwei	ART	ART	CARD
Hunde	NN	NNS	NNS
bellen	VVFIN	VVINF	VVFINX
.	$.	$.	$.

Word	A1	A2	A3
Zwei	ART	ART	CARD
Hunde	NN	NNS	NNS
bellen	VVFIN	VVINF	VVFINX
.	$.	$.	$.

$\Downarrow$ Konversion $\Downarrow$

Word	ART	CARD	NN	NNS	VVFIN	VVINF	VVFINX	$.
Zwei	2	1						
Hunde			1	2				
bellen					1	1	1	
.								3

Inter-Annotator Agreement

Paarungen

Word	ART	CARD	NN	NNS	VVFIN	VVINF	VVFINX	$.
Zwei	2	1						
Hunde			1	2				
bellen					1	1	1	
.								3

- Wieviele paarweise Übereinstimmungen gibt es (absolut gemessen)?

Inter-Annotator Agreement

Paarungen

Word	ART	CARD	NN	NNS	VVFIN	VVINF	VVFINX	$.
Zwei	2	1						
Hunde			1	2				
bellen					1	1	1	
.								3

- Wieviele paarweise Übereinstimmungen gibt es (absolut gemessen)?
  - $1 + 1 + 0 + 3 = 5$

Inter-Annotator Agreement

Paarungen

Word	ART	CARD	NN	NNS	VVFIN	VVINF	VVFINX	$.
Zwei	2	1						
Hunde			1	2				
bellen					1	1	1	
.								3

- Wieviele paarweise Übereinstimmungen gibt es (absolut gemessen)?
  - $1 + 1 + 0 + 3 = 5$
- Warum 3? Binomialkoeffizient!

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Fakultät:

$$n! = n(n-1)(n-2)(n-3)\cdots 1$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Fakultät:

$$n! = n(n-1)(n-2)(n-3)\cdots 1$$

Für $k = 2$:

$$\binom{n}{2} = \frac{n!}{(n-2)!2!}$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Fakultät:

$$n! = n(n-1)(n-2)(n-3)\cdots 1$$

Für $k = 2$:

$$\binom{n}{2} \quad = \quad \frac{n!}{(n-2)!2!} = \frac{1}{2}\frac{n!}{(n-2)!}$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Fakultät:

$$n! = n(n-1)(n-2)(n-3)\cdots 1$$

Für $k = 2$:

$$\begin{aligned}\binom{n}{2} &= \frac{n!}{(n-2)!2!} = \frac{1}{2}\frac{n!}{(n-2)!} \\ &= \frac{1}{2}\frac{n(n-1)(n-2)(n-3)\cdots 1}{(n-2)(n-3)\cdots 1}\end{aligned}$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Fakultät:

$$n! = n(n-1)(n-2)(n-3)\cdots 1$$

Für $k = 2$:

$$\begin{aligned}\binom{n}{2} &= \frac{n!}{(n-2)!2!} = \frac{1}{2}\frac{n!}{(n-2)!} \\ &= \frac{1}{2}\frac{n(n-1)(n-2)(n-3)\cdots 1}{(n-2)(n-3)\cdots 1} \\ &= \frac{1}{2}n(n-1)\end{aligned}$$

Binomialkoeffizient / ‘n choose k’

Anzahl möglicher k-elementiger Teilmengen aus n Dingen

Reihenfolge: Sind (a, b) und (b, a) das gleiche Paar?

- Ja: $\frac{1}{2}n(n-1)$
- Nein: $n(n-1)$

Inter-Annotator Agreement

Paarungen

n	$\binom{n}{2}$
1	0
2	1
3	3
4	6
5	10
10	45
100	4.950

Table: $\binom{n}{2}$ Werte für steigende n

Inter-Annotator Agreement

Problems

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- 3 annotators (=n)
- 5 pairwise agreements

Inter-Annotator Agreement

Problems

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- 3 annotators (=n)
- 5 pairwise agreements

Situation 2:

Word	A	B	C	D
1	2	1	1	
2		1	1	2
3	1	1	1	1
4	3	1		

- 4 annotators (=n)
- 5 pairwise agreements

Inter-Annotator Agreement

Problems

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- 3 annotators (=n)
- 5 pairwise agreements

Situation 2:

Word	A	B	C	D
1	2	1	1	
2		1	1	2
3	1	1	1	1
4	3	1		

- 4 annotators (=n)
- 5 pairwise agreements

How much worse is Situation 2 compared to 1?
$\rightarrow$ Scaling!

Inter-Annotator Agreement

Scaling

- Sometimes, values have different scales
  - i.e., different min and max values
- Scaling: Apply a function to values so that they are comparable
  - Simplest way: Divide by the (theoretical) maximum

Inter-Annotator Agreement

Scaling

- Sometimes, values have different scales
  - i.e., different min and max values
- Scaling: Apply a function to values so that they are comparable
  - Simplest way: Divide by the (theoretical) maximum
- What's the theoretical maximum here?
  - If all annotators agree: $\binom{n}{2} = \frac{1}{2}n(n-1)$

Inter-Annotator Agreement

Scaling

- Sometimes, values have different scales
  - i.e., different min and max values
- Scaling: Apply a function to values so that they are comparable
  - Simplest way: Divide by the (theoretical) maximum
- What's the theoretical maximum here?
  - If all annotators agree: $\binom{n}{2} = \frac{1}{2}n(n-1)$
- Scaling $\simeq$ Normalization

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- 3 annotators (=n)
- 5 pairwise agreements
- Scaled: $\frac{5}{4\binom{3}{2}} = \frac{5}{4\times3} = \frac{5}{12} = 0.416$

Situation 2:

Word	A	B	C	D
1	2	1	1	
2		1	1	2
3	1	1	1	1
4	3	1		

- 4 annotators (=n)
- 5 pairwise agreements
- Scaled: $\frac{5}{4\binom{4}{2}} = \frac{5}{4\times6} = \frac{5}{24} = 0.208$

Inter-Annotator Agreement

Observed Agreement

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j}-1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

$$\underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j} - 1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\hat{P}_i = \underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j}-1)}_{\text{abs. pairwise agr. for item } i}$$

Inter-Annotator Agreement

Observed Agreement

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

$N = 4, k = 4, n = 3$

Normalized observed agreement for item i

Problem: k categories, n annotators, N items

$$\hat{P}_i = \underbrace{\frac{1}{n(n-1)}}_{\text{Scaling for annotators}} \times \underbrace{\sum_{j=1}^{k} n_{i,j}(n_{i,j}-1)}_{\text{abs. pairwise agr. for item } i}$$

Normalized observed agreement for all items

$$P_o = \frac{1}{N} \sum_{i=1}^{N} \hat{P}_i$$

(this is just the arithmetic mean, a.k.a. average)

Large Operators

$$\sum_{i=0}^{n} f(i) = f(0) + f(1) + f(2) + \cdots f(n)$$

Large Operators

$$\sum_{i=0}^{n} f(i) = f(0) + f(1) + f(2) + \cdots f(n)$$

Example

$$\sum_{i=0}^{4} i^5 \quad = \quad 0^5 + 1^5 + 2^5 + 3^5 + 4^5 = 1300$$

Large Operators

$$\sum_{i=0}^{n} f(i) = f(0) + f(1) + f(2) + \cdots f(n)$$

Example

$$\begin{aligned}
\sum_{i=0}^{4} i^5 &= 0^5 + 1^5 + 2^5 + 3^5 + 4^5 = 1300 \\
\prod_{i=1}^{3} (5i + 1) &= (5(1) + 1) \times (5(2) + 1) \times (5(3) + 1) \\
&= 6 + 11 + 16 = 33
\end{aligned}$$

Inter-Annotator Agreement

Expected Agreement

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

- $n = 3$ annotators
- $k = 4$ categories
- 5 pairwise agreements

Situation 2:

Word	A	B	C
1	2	1	
2	2		1
3	1	1	1
4	3		

- $n = 3$ annotators
- $k = 3$ categories
- 5 pairwise agreements

What situation had the better agreement? How much better?

Inter-Annotator Agreement

Expected Agreement

- How likely is it (in general) that category A is selected?

Inter-Annotator Agreement

Expected Agreement

- How likely is it (in general) that category A is selected?

Excurs: Relative Frequencies

- How often do you win a game?

Inter-Annotator Agreement

Expected Agreement

- How likely is it (in general) that category A is selected?

Excurs: Relative Frequencies

- How often do you win a game?
- $\frac{\text{Number of positive events } (= \text{wins})}{\text{Number of total events} (= \text{games played})}$

Inter-Annotator Agreement

Expected Agreement

- How likely is it (in general) that category A is selected?

Excurs: Relative Frequencies

- How often do you win a game?
- $\frac{\text{Number of positive events } (= \text{wins})}{\text{Number of total events} (= \text{games played})}$
- $p(\text{"selecting category A"}) = ?$

Inter-Annotator Agreement

Expected Agreement

- How likely is it (in general) that category A is selected?

Excurs: Relative Frequencies

- How often do you win a game?
- $\frac{\text{Number of positive events } (= \text{wins})}{\text{Number of total events} (= \text{games played})}$
- $p(\text{"selecting category A"}) = ?$
- Relative frequency is an *estimate* of the probability
  - Probability: Theoretical concept
  - Relative frequency: Result of actual experiments
  - Assumption: The more experiments I do, the more similar is the relative frequency to the probability

Inter-Annotator Agreement

Expected Agreement

Situation 1:

Word	A	B	C	D
1	2	1		
2			1	2
3	1	1		1
4	3			

Situation 2:

Word	A	B	C
1	2	1	
2	2		1
3	1	1	1
4	3		

- $p(\text{"selecting category A"}) = \frac{\text{Number of positive events}}{\text{Number of possible events}}$
  - Positive events: How often category A has been selected
  - Possible events: How many decisions have been made
- $p(\text{"selecting category A"}) =$
  - Situation 1: $\frac{6}{12}$
  - Situation 2: $\frac{8}{12}$

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$\underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$\underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

- Probability that two annotators select category j

$$p_j \times p_j = p_j^2$$

Inter-Annotator Agreement

Expected Agreement

- Probability that category j gets selected (by one annotator)

$$p_j = \underbrace{\frac{1}{nN}}_{\text{Possible events (all annotations)}} \times \underbrace{\sum_{i=1}^{N} n_{i,j}}_{\text{positive events (=annotations with } j\text{)}}$$

- Probability that two annotators select category j

$$p_j \times p_j = p_j^2$$

- Probability that two annotators are in agreement (over all categories):

$$P_e = \sum_{j=1}^{k} p_j^2$$

Fleiss' Kappa (**Fleiss:1971aa**)

p_j	Probability for j
$\hat{P}_i$	Observed agreement for i

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{array}{lcl}
p_j & & \text{Probability for } j \\
\hat{P}_i & & \text{Observed agreement for } i \\
P_o & = & \dfrac{1}{N} \sum_{i=1}^{N} \hat{P}_i
\end{array}
$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{aligned}
p_j & \quad \text{Probability for } j \\
\hat{P}_i & \quad \text{Observed agreement for } i \\
P_o &= \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \\
P_e &= \sum_{j=1}^{k} p_j^2
\end{aligned}
$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{aligned}
p_j & \quad \text{Probability for } j \\
\hat{P}_i & \quad \text{Observed agreement for } i \\
P_o &= \frac{1}{N} \sum_{i=1}^{N} \hat{P}_i \\
P_e &= \sum_{j=1}^{k} p_j^2 \\
\kappa &= \frac{P_o - P_e}{1 - P_e}
\end{aligned}
$$

Fleiss' Kappa (**Fleiss:1971aa**)

$$\begin{array}{lcl} p_j & & \text{Probability for } j \\ \hat{P}_i & & \text{Observed agreement for } i \\ P_o & = & \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \hat{P}_i \\ P_e & = & \displaystyle\sum_{j=1}^{k} p_j^2 \\ \kappa & = & \dfrac{P_o - P_e}{1 - P_e} \end{array}$$

- $P_o - P_e$: Tatsächlich erreichtes, nicht-zufälliges Agreement
- $1 - P_e$: Maximal erreichbares, nicht-zufälliges Agreement

Fleiss' Kappa (**Fleiss:1971aa**)

$$
\begin{array}{lcl}
p_j & & \text{Probability for } j \\
\hat{P}_i & & \text{Observed agreement for } i \\
P_o & = & \frac{1}{N}\sum_{i=1}^{N}\hat{P}_i \\
P_e & = & \sum_{j=1}^{k} p_j^2 \\
\kappa & = & \frac{P_o - P_e}{1 - P_e}
\end{array}
$$

- $P_o - P_e$: Tatsächlich erreichtes, nicht-zufälliges Agreement
- $1 - P_e$: Maximal erreichbares, nicht-zufälliges Agreement
- $-\infty < \kappa < 1$: Je höher desto besser
  - Extremfälle?