

Práctica - Reglas de Asociación en R

Laura Rodriguez Navas

April 2020

Descarga el dataset Titanic de la siguiente URL <http://www.rdatamining.com/data/titanic.raw.rdata?attredirects=0&d=1>.

Cargamos el dataset.

```
load("./titanic.raw.rdata")
```

Lo primero que haremos es comprobar que el dataset ha sido cargado correctamente, comprobando:

- Numero de registros (observaciones).
- Numero de variables.
- Tipo de variable.
- Numero de valores por cada variable.

```
str(titanic.raw)
```

```
## 'data.frame': 2201 obs. of 4 variables:
## $ Class : Factor w/ 4 levels "1st","2nd","3rd",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ Sex : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
## $ Age : Factor w/ 2 levels "Adult","Child": 2 2 2 2 2 2 2 2 2 2 ...
## $ Survived: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Describe cada una de las características antes indicadas (numero de registros, numero de variables, etc.).

- Numero de registros (observaciones): **2201**.
- Numero de variables: **4** (Class, Sex, Age y Survived).
- Tipo de variable: **nominal**.
- Numero de valores por cada variable: **4** para Class, **2** para Sex, **2** para Age y **2** para Survived.

Analizamos la distribución de los datos, comprobando cuantos registros existen para cada valor de cada variable.

```
summary(titanic.raw)
```

```
## Class Sex Age Survived
## 1st :325 Female: 470 Adult:2092 No :1490
## 2nd :285 Male :1731 Child: 109 Yes: 711
## 3rd :706
## Crew:885
```

Extraemos reglas de asociación con el algoritmo Apriori y los valores por defecto (importante asegurarse que estos son los valores por defecto):

- Soporte minimo: 0.1
- Confianza minima: 0.8
- Numero maximo de items (longitud maxima de regla): 10

```
rules <- apriori(titanic.raw, parameter = NULL, appearance = NULL, control = NULL)
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.8    0.1    1 none FALSE                TRUE     5     0.1     1
## maxlen target   ext
##          10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE     2     TRUE
##
## Absolute minimum support count: 220
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[10 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [9 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [27 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

Mostramos todas las reglas obtenidas por el algoritmo.

```
inspect(rules)
```

| ## | lhs | rhs | support | confidence |
|---------|-------------------------------------|------------------|-----------|------------|
| ## [1] | {} | => {Age=Adult} | 0.9504771 | 0.9504771 |
| ## [2] | {Class=2nd} | => {Age=Adult} | 0.1185825 | 0.9157895 |
| ## [3] | {Class=1st} | => {Age=Adult} | 0.1449341 | 0.9815385 |
| ## [4] | {Sex=Female} | => {Age=Adult} | 0.1930940 | 0.9042553 |
| ## [5] | {Class=3rd} | => {Age=Adult} | 0.2848705 | 0.8881020 |
| ## [6] | {Survived=Yes} | => {Age=Adult} | 0.2971377 | 0.9198312 |
| ## [7] | {Class=Crew} | => {Sex=Male} | 0.3916402 | 0.9740113 |
| ## [8] | {Class=Crew} | => {Age=Adult} | 0.4020900 | 1.0000000 |
| ## [9] | {Survived=No} | => {Sex=Male} | 0.6197183 | 0.9154362 |
| ## [10] | {Survived=No} | => {Age=Adult} | 0.6533394 | 0.9651007 |
| ## [11] | {Sex=Male} | => {Age=Adult} | 0.7573830 | 0.9630272 |
| ## [12] | {Sex=Female, Survived=Yes} | => {Age=Adult} | 0.1435711 | 0.9186047 |
| ## [13] | {Class=3rd, Sex=Male} | => {Survived=No} | 0.1917310 | 0.8274510 |
| ## [14] | {Class=3rd, Survived=No} | => {Age=Adult} | 0.2162653 | 0.9015152 |
| ## [15] | {Class=3rd, Sex=Male} | => {Age=Adult} | 0.2099046 | 0.9058824 |
| ## [16] | {Sex=Male, Survived=Yes} | => {Age=Adult} | 0.1535666 | 0.9209809 |
| ## [17] | {Class=Crew, Survived=No} | => {Sex=Male} | 0.3044071 | 0.9955423 |
| ## [18] | {Class=Crew, Survived=No} | => {Age=Adult} | 0.3057701 | 1.0000000 |
| ## [19] | {Class=Crew, Sex=Male} | => {Age=Adult} | 0.3916402 | 1.0000000 |
| ## [20] | {Class=Crew, Age=Adult} | => {Sex=Male} | 0.3916402 | 0.9740113 |
| ## [21] | {Sex=Male, Survived=No} | => {Age=Adult} | 0.6038164 | 0.9743402 |
| ## [22] | {Age=Adult, Survived=No} | => {Sex=Male} | 0.6038164 | 0.9242003 |
| ## [23] | {Class=3rd, Sex=Male, Survived=No} | => {Age=Adult} | 0.1758292 | 0.9170616 |
| ## [24] | {Class=3rd, Age=Adult, Survived=No} | => {Sex=Male} | 0.1758292 | 0.8130252 |
| ## [25] | {Class=3rd, Sex=Male, Age=Adult} | => {Survived=No} | 0.1758292 | 0.8376623 |
| ## [26] | {Class=Crew, Sex=Male, Survived=No} | => {Age=Adult} | 0.3044071 | 1.0000000 |

```
## [27] {Class=Crew, Age=Adult, Survived=No} => {Sex=Male}    0.3044071 0.9955423
##      lift      count
## [1] 1.0000000 2092
## [2] 0.9635051 261
## [3] 1.0326798 319
## [4] 0.9513700 425
## [5] 0.9343750 627
## [6] 0.9677574 654
## [7] 1.2384742 862
## [8] 1.0521033 885
## [9] 1.1639949 1364
## [10] 1.0153856 1438
## [11] 1.0132040 1667
## [12] 0.9664669 316
## [13] 1.2222950 422
## [14] 0.9484870 476
## [15] 0.9530818 462
## [16] 0.9689670 338
## [17] 1.2658514 670
## [18] 1.0521033 673
## [19] 1.0521033 862
## [20] 1.2384742 862
## [21] 1.0251065 1329
## [22] 1.1751385 1329
## [23] 0.9648435 387
## [24] 1.0337773 387
## [25] 1.2373791 387
## [26] 1.0521033 670
## [27] 1.2658514 670
```

Mostramos sólo las 3 mejores reglas en base a la métrica lift.

```
inspect(head(sort(rules, by = "lift"), 3))
```

```
##      lhs                                rhs      support  confidence
## [1] {Class=Crew, Survived=No}          => {Sex=Male} 0.3044071 0.9955423
## [2] {Class=Crew, Age=Adult, Survived=No} => {Sex=Male} 0.3044071 0.9955423
## [3] {Class=Crew}                      => {Sex=Male} 0.3916402 0.9740113
##      lift      count
## [1] 1.265851 670
## [2] 1.265851 670
## [3] 1.238474 862
```

Describa cada una de las reglas obtenidas, explicando su significado así como el significado de las métricas existentes para cada regla.

- Primera regla. Los hombres de la tripulación que no sobrevivieron fueron 670, un 30% del total de las personas que iban a bordo. La proporción de hombres que no sobrevivieron y que eran de la tripulación es muy alta, un 99.5%. La probabilidad de que esta regla sea cierta es muy alta, ya que el valor de la métrica lift es superior a 1.
- Segunda regla. Los hombres adultos de la tripulación que no sobrevivieron fueron 670, un 30% del total de las personas que iban a bordo. Teniendo en cuenta la primera regla y esta, podemos afirmar que todos los hombres de la tripulación que murieron fueron adultos. Por eso el valor de las métricas de confianza y lift tienen el mismo valor que en la regla anterior.

- Tercera regla. Los hombres que formaron parte de la tripulación fueron 862, un 39% del total de las personas que iban a bordo. La proporción de hombres que eran de la tripulación es muy alta, un 97.5%. La probabilidad de que esta regla sea cierta es muy alta, ya que el valor de la métrica lift es superior a 1. Teniendo en cuenta esto, podríamos afirmar que la mayoría de personas de la tripulación fueron hombres.

Extraemos reglas de asociación con el algoritmo Apriori y los siguientes valores:

- Soporte minimo: 0.1.
- Confianza minima: 0.9.
- Numero maximo de items (longitud maxima de regla): 10.
- Los valores Age=Adult y Age=Child no pueden aparecer en ningun sitio de la regla y el resto de valores puede aparecer en ambos lugares (antecedente y consecuente).

```
rules <- apriori(titanic.raw, parameter=list(support=0.1, confidence=0.9),
               appearance = list(none = c("Age=Adult", "Age=Child"),default="both"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.9    0.1    1 none FALSE                TRUE         5     0.1    1
## maxlen target  ext
##          10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE     2    TRUE
##
## Absolute minimum support count: 220
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[10 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [8 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [3 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(rules)
```

```
##      lhs                      rhs      support  confidence lift    count
## [1] {Class=Crew}              => {Sex=Male} 0.3916402 0.9740113 1.238474 862
## [2] {Survived=No}             => {Sex=Male} 0.6197183 0.9154362 1.163995 1364
## [3] {Class=Crew,Survived=No} => {Sex=Male} 0.3044071 0.9955423 1.265851 670
```

Describe los resultados obtenidos, qué reglas son más interesantes y por qué.

- Primera regla. No es interesante. Ya la hemos descrito en el apartado anterior (tercera regla).
- Segunda regla. Los hombres que no sobrevivieron fueron 1364, un 61% del total de las personas que iban a bordo. La proporción de hombres que no sobrevivieron es muy alta, un 91.5%. La probabilidad de que esta regla sea cierta es muy alta, ya que el valor de la métrica lift es superior a 1. Esta regla es muy interesante porque nos indica que más del 50% del total de las personas que iban a bordo que murieron fueron hombres.

- Tercera regla. No es interesante. Ya la hemos descrito en el apartado anterior (primera regla).

Combinando las reglas podemos observar que solo sobrevivieron 192 hombres de la tripulación, de un total de 832. Muy pocos pudieron sobrevivir, la mayoría de los hombres que murieron pertenecían a la tripulación.

Extraemos reglas de asociación con el algoritmo Apriori y los siguientes valores:

- Soporte mínimo: 0.1.
- Confianza mínima: 0.9.
- Número máximo de items (longitud máxima de regla): 10.
- Los valores Age=Adult y Age=Child solo pueden aparecer en el antecedente. En el consecuente sólo si sobrevivieron o no.

```
rules <- apriori(titanic.raw, parameter=list(support=0.01, confidence=0.3),
               appearance = list(lhs = c("Age=Adult", "Age=Child"),
                                rhs = c("Survived=No", "Survived=Yes"),default="none"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.3    0.1    1 none FALSE                TRUE      5    0.01    1
## maxlen target   ext
##          10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 22
##
## set item appearances ...[4 item(s)] done [0.00s].
## set transactions ...[4 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [4 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [6 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(rules)
```

| ## | lhs | rhs | support | confidence | lift | count |
|--------|-------------|-------------------|------------|------------|-----------|-------|
| ## [1] | {} | => {Survived=Yes} | 0.32303498 | 0.3230350 | 1.0000000 | 711 |
| ## [2] | {} | => {Survived=No} | 0.67696502 | 0.6769650 | 1.0000000 | 1490 |
| ## [3] | {Age=Child} | => {Survived=Yes} | 0.02589732 | 0.5229358 | 1.6188209 | 57 |
| ## [4] | {Age=Child} | => {Survived=No} | 0.02362562 | 0.4770642 | 0.7047103 | 52 |
| ## [5] | {Age=Adult} | => {Survived=Yes} | 0.29713766 | 0.3126195 | 0.9677574 | 654 |
| ## [6] | {Age=Adult} | => {Survived=No} | 0.65333939 | 0.6873805 | 1.0153856 | 1438 |

Describe los resultados obtenidos, qué reglas son más interesantes y por qué.

- Primera regla. Todas las personas que iban a bordo y que sobrevivieron fueron 711, un 32% del total. La proporción de personas que sobrevivieron es muy baja. Muy pocas personas sobrevivieron, solo un 32% del total. La probabilidad de que esta regla sea cierta es total, ya que el valor de la métrica lift es 1.

- Segunda regla. Todas las personas que iban a bordo y que no sobrevivieron fueron 1490, un 68% del total. La proporción de personas que no sobrevivieron es muy alta, un 68%. La mayoría de las personas a bordo murió. La probabilidad de que esta regla sea cierta es total, ya que el valor de la métrica lift es 1.

Si sumamos el total de personas que sobrevivieron (711) de la primera regla, con el total de personas que murieron de la segunda regla (1490), el resultado es el total de personas que iban a bordo en el Titanic (2201).

- Tercera regla y Cuarta regla. Los niños que sobrevivieron fueron 57, un 2.6% del total de las personas que iban a bordo. Y los niños que no sobrevivieron fueron 52, un 2.4% del total de las personas que iban a bordo. La proporción de los niños que sobrevivieron y la proporción de los que no sobrevivieron es muy parecida. Y es más probable que sobrevivieran ya que el valor de la métrica lift en este caso es superior a 1.
- Quinta regla y Sexta regla. Los adultos que sobrevivieron fueron 654, un 29.7% del total de las personas que iban a bordo. Y los adultos que no sobrevivieron fueron 1438, un 65.3% del total de las personas que iban a bordo. La proporción de adultos que no sobrevivieron es mucho mayor (68.7%) que la proporción de los adultos que sobrevivieron (31.2%).

Es muy interesante observar que la proporción de los niños que sobrevivieron y los que no, está muy equilibrada. En cambio, en el caso de los adultos no fue así. La mayoría murió.

Buscamos la regla “Sex=Female” THEN “Survived=Yes” para ver cuantas mujeres sobrevivieron.

```
rules <- apriori(titanic.raw, parameter=list(support=0, confidence=0, minlen=2),
               appearance = list(lhs = c("Sex=Female"), rhs = c("Survived=Yes"),default="none"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0    0.1   1 none FALSE                TRUE     5     0     2
## maxlen target   ext
##      10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[2 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [2 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [1 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
inspect(rules)
```

```
##      lhs                rhs          support  confidence lift      count
## [1] {Sex=Female} => {Survived=Yes} 0.1562926 0.7319149 2.265745 344
```

Describe los resultados obtenidos, qué reglas son más interesantes y por qué.

Solo se ha encontrado una regla. La regla muestra que las mujeres que sobrevivieron fueron 344, un 15.6% del total de las personas que iban a bordo. Parecen pocas, pero eso se debe a que había pocas mujeres a bordo. Pero muchas de ellas sobrevivieron, ya que la regla presenta valores de confianza (73%) y lift muy altos.

Me quedo sólo con los valores de soporte, confianza y lift.

```
soporte <- quality(rules)$support
confianza <- quality(rules)$confidence
lift <- quality(rules)$lift
soporte
```

```
## [1] 0.1562926
```

```
confianza
```

```
## [1] 0.7319149
```

```
lift
```

```
## [1] 2.265745
```

Guardar en la variable “numSobreviven” el numero de mujeres de tercera clase que sobrevivieron al hundimiento del Titanic. Para ello, buscar la regla de asociación específica suponiendo que el consecuente incluye el item Survived=Yes.

```
rules <- apriori(titanic.raw, parameter=list(support=0, confidence=0, minlen=2),
               appearance = list(lhs = c("Class=3rd", "Sex=Female"),
                                rhs = c("Survived=Yes"), default="none"))
```

```
## Apriori
```

```
##
```

```
## Parameter specification:
```

```
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0    0.1   1 none FALSE             TRUE     5       0       2
```

```
## maxlen target  ext
```

```
##      10  rules FALSE
```

```
##
```

```
## Algorithmic control:
```

```
## filter tree heap memopt load sort verbose
```

```
##    0.1 TRUE TRUE  FALSE TRUE     2     TRUE
```

```
##
```

```
## Absolute minimum support count: 0
```

```
##
```

```
## set item appearances ...[3 item(s)] done [0.00s].
```

```
## set transactions ...[3 item(s), 2201 transaction(s)] done [0.00s].
```

```
## sorting and recoding items ... [3 item(s)] done [0.00s].
```

```
## creating transaction tree ... done [0.00s].
```

```
## checking subsets of size 1 2 3 done [0.00s].
```

```
## writing ... [3 rule(s)] done [0.00s].
```

```
## creating S4 object ... done [0.00s].
```

```
inspect(rules)
```

```
##      lhs                                rhs          support  confidence lift
## [1] {Sex=Female}                      => {Survived=Yes} 0.15629259 0.7319149 2.2657450
## [2] {Class=3rd}                      => {Survived=Yes} 0.08087233 0.2521246 0.7804871
## [3] {Class=3rd,Sex=Female} => {Survived=Yes} 0.04089050 0.4591837 1.4214673
##      count
```

```
## [1] 344
## [2] 178
## [3] 90

numSobreviven <- quality(rules)$count[3]
numSobreviven
```

```
## [1] 90
```

Describe los resultados obtenidos, qué regla o reglas son más interesantes y por qué.

Primera regla. No es interesante. Ya la hemos descrito en el apartado anterior.

Segunda regla. No es muy interesante. Lo sería más si lo comparáramos con otro conjunto de ítems. Que es el caso de la tercera regla.

Tercera regla. Las mujeres de la tercera clase que sobrevivieron fueron 90, un 0.41% del total de personas a bordo. Parecen pocas, pero eso se debe a que había pocas mujeres de la tercera clase a bordo. Pero la mitad de ellas sobrevivieron, ya que la regla presenta valores de confianza (50%) y el valor de la métrica lift es superior a 1.

Sumar el número de mujeres de cada clase que no sobrevivieron al hundimiento del Titanic. Comprobar que la suma es igual al número de mujeres que no sobrevivieron al hundimiento del Titanic. Para ello, buscar las reglas de asociación específicas suponiendo que el consecuente incluye el ítem Survived=No. La variable sumaMujeresMuertas (obtenida de la suma de los resultados de las reglas específicas) tiene que ser igual a la variable numMujeresMuertas obtenida a partir de la regla de asociación específica.

Primero buscamos la regla del numMujeresMuertas, fueron 126.

```
rules <- apriori(titanic.raw, parameter=list(support=0, confidence=0, minlen=2),
               appearance = list(lhs = c("Sex=Female"), rhs = c("Survived=No"),default="none"))

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0   0.1   1 none FALSE                TRUE     5       0       2
## maxlen target  ext
##      10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[2 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [2 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [1 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].

numMujeresMuertas <- quality(rules)$count[1]
numMujeresMuertas

## [1] 126
```


Después buscamos las reglas para saber el número de las mujeres que murieron para cada variable Class {"1st", "2nd", "3rd", "Crew"}.

Guardamos esos valores en las siguientes variables: MujeresMuertas1st (4), MujeresMuertas2nd (13), MujeresMuertas3rd (106) y MujeresMuertasCrew (3).

Finalmente sumamos estos valores y comprobamos que su suma nos da el mismo valor que sumaMujeresMuertas.

```
inspect(MujeresMuertas1st)
```

```
##      lhs                rhs      support    confidence lift
## [1] {Class=1st}          => {Survived=No} 0.055429350 0.37538462 0.55451110
## [2] {Sex=Female}         => {Survived=No} 0.057246706 0.26808511 0.39601028
## [3] {Class=1st,Sex=Female} => {Survived=No} 0.001817356 0.02758621 0.04074983
##      count
## [1] 122
## [2] 126
## [3] 4
```

```
MujeresMuertas1st <- quality(MujeresMuertas1st)$count[3]
MujeresMuertas1st
```

```
## [1] 4
```

```
inspect(MujeresMuertas2nd)
```

```
##      lhs                rhs      support    confidence lift
## [1] {Class=2nd}          => {Survived=No} 0.075874602 0.5859649 0.8655764
## [2] {Sex=Female}         => {Survived=No} 0.057246706 0.2680851 0.3960103
## [3] {Class=2nd,Sex=Female} => {Survived=No} 0.005906406 0.1226415 0.1811637
##      count
## [1] 167
## [2] 126
## [3] 13
```

```
MujeresMuertas2nd <- quality(MujeresMuertas2nd)$count[3]
MujeresMuertas2nd
```

```
## [1] 13
```

```
inspect(MujeresMuertas3rd)
```

```
##      lhs                rhs      support    confidence lift
## [1] {Sex=Female}         => {Survived=No} 0.05724671 0.2680851 0.3960103
## [2] {Class=3rd}          => {Survived=No} 0.23989096 0.7478754 1.1047474
## [3] {Class=3rd,Sex=Female} => {Survived=No} 0.04815993 0.5408163 0.7988837
##      count
## [1] 126
## [2] 528
## [3] 106
```

```
MujeresMuertas3rd <- quality(MujeresMuertas3rd)$count[3]
MujeresMuertas3rd
```

```
## [1] 106
```

```
inspect(MujeresMuertasCrew)
```

```
##      lhs                rhs      support    confidence lift
```

```
## [1] {Sex=Female}          => {Survived=No} 0.057246706 0.2680851 0.3960103
## [2] {Class=Crew}         => {Survived=No} 0.305770104 0.7604520 1.1233254
## [3] {Class=Crew,Sex=Female} => {Survived=No} 0.001363017 0.1304348 0.1926758
##      count
## [1] 126
## [2] 673
## [3]   3
```

```
MujeresMuertasCrew <- quality(MujeresMuertasCrew)$count[3]
MujeresMuertasCrew
```

```
## [1] 3
```

```
sumaMujeresMuertas <- MujeresMuertas1st + MujeresMuertas2nd +
  MujeresMuertas3rd + MujeresMuertasCrew
sumaMujeresMuertas
```

```
## [1] 126
```

Para dar respuesta a las siguientes cuestiones que se plantean, es necesario buscar reglas específicas para el conocimiento que estamos buscando. Lea detenidamente las cuestiones que se plantean y responde a ellas de manera clara y concisa. Explica de manera clara y concisa cómo has obtenido dichos resultados o cómo has llegado a dichas conclusiones.

Las cuestiones planteadas son:

a) ¿Se cumplió? la norma de los niños y las mujeres primero?

```
rules <- apriori(titanic.raw, parameter=list(support=0.01, confidence=0, minlen=3),
  appearance = list(lhs = c("Age=Adult", "Age=Child", "Sex=Male", "Sex=Female"),
    rhs = c("Survived=No"),default="none"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0    0.1   1 none FALSE                TRUE     5    0.01    3
## maxlen target  ext
##      10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 22
##
## set item appearances ...[5 item(s)] done [0.00s].
## set transactions ...[5 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [5 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [3 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(rules)
```

```
##      lhs                      rhs          support  confidence lift
## [1] {Sex=Male,Age=Child}    => {Survived=No} 0.01590186 0.5468750 0.8078335
```

```
## [2] {Sex=Female, Age=Adult} => {Survived=No} 0.04952294 0.2564706 0.3788535
## [3] {Sex=Male, Age=Adult}   => {Survived=No} 0.60381645 0.7972406 1.1776688
##      count
## [1]    35
## [2]   109
## [3]  1329
```

Sí. Como podemos observar, el número de niños y mujeres que no sobrevivieron fueron un total de 144 (35 + 109). Muy pocos niños y muy pocas mujeres murieron en general. Acorde a eso, los valores del soporte, confianza y lift són muy bajos. En cambio, sí que murieron muchos hombres, un total de 1329. Un 60% de las personas que murieron y que iban a bordo fueron los hombres.

b) ¿Tuvo mayor peso la clase a la que pertenecía el pasajero?

```
rules <- apriori(titanic.raw, parameter=list(support=0, confidence=0, minlen=2),
              appearance = list(lhs = c("Class=Crew", "Class=1st", "Class=2nd", "Class=3rd"),
                                rhs = c("Survived=No"), default="none"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0    0.1   1 none FALSE                TRUE     5      0      2
## maxlen target  ext
##      10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[5 item(s)] done [0.00s].
## set transactions ...[5 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [5 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 done [0.00s].
## writing ... [4 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(rules)
```

```
##      lhs          rhs      support  confidence lift      count
## [1] {Class=1st} => {Survived=No} 0.05542935 0.3753846 0.5545111 122
## [2] {Class=2nd} => {Survived=No} 0.07587460 0.5859649 0.8655764 167
## [3] {Class=3rd} => {Survived=No} 0.23989096 0.7478754 1.1047474 528
## [4] {Class=Crew} => {Survived=No} 0.30577010 0.7604520 1.1233254 673
```

Sí. Se puede observar muy bien en estas cuatro reglas. Que representan la división por Class de las personas que murieron a bordo. Vemos como las personas de la tripulación y de la tercera clase son los que más número de muertos representan.

c) ¿Podemos saber si la tripulación se comportó heroicamente?

Sí, por la regla número 4 anterior. Podemos observar que las personas que más murieron fueron los de la tripulación.

Obtener las reglas de mayor longitud (las que incluyan un mayor numero de variables). Utilizar diferentes umbrales de soporte y confianza y mostrar qué reglas son las de mayor longitud para los diferentes umbrales.

```
rules <- apriori(titanic.raw, parameter = list(support=0.5, confidence=0.9, target="rules"))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.9    0.1    1 none FALSE                TRUE     5     0.5     1
## maxlen target   ext
##          10  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE     2     TRUE
##
## Absolute minimum support count: 1100
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[10 item(s), 2201 transaction(s)] done [0.00s].
## sorting and recoding items ... [3 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [6 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
inspect(sort(rules, by ="count"))
```

| | lhs | rhs | support | confidence | lift | count |
|--------|--------------------------|----------------|-----------|------------|----------|-------|
| ## [1] | {} | => {Age=Adult} | 0.9504771 | 0.9504771 | 1.000000 | 2092 |
| ## [2] | {Sex=Male} | => {Age=Adult} | 0.7573830 | 0.9630272 | 1.013204 | 1667 |
| ## [3] | {Survived=No} | => {Age=Adult} | 0.6533394 | 0.9651007 | 1.015386 | 1438 |
| ## [4] | {Survived=No} | => {Sex=Male} | 0.6197183 | 0.9154362 | 1.163995 | 1364 |
| ## [5] | {Sex=Male, Survived=No} | => {Age=Adult} | 0.6038164 | 0.9743402 | 1.025106 | 1329 |
| ## [6] | {Age=Adult, Survived=No} | => {Sex=Male} | 0.6038164 | 0.9242003 | 1.175139 | 1329 |