SPH R Appreciation Society: Small cell suppression

Malcolm Gillies <malcolm.gillies@unsw.edu.au>

25 February 2025

Small cell suppression

- What even is it?
- What packages are available
- Quick `easySdcTable` walkthrough
- Work in progress

Privacy, confidentiality, disclosure, reidentification

- Privacy act requirements for ‘de-identification’
- Harm due to disclosure of personal information
- Data custodian requirements
- AIHW/ABS requirements

Statistical disclosure control

- Deidentification, confidentialisation, ...
- Methods to achieve this aim include:
- *Small cell suppression*

Rule of 6 (or 5, or 10, ...)

Age cat	Count
15-19	16
20-24	8
25-29	3
30-34	4
Total	31

Secondary suppression

Age cat	Male Count	Female Count	Total Count
15-19	16	0	16
20-24	8	10	18
25-29	3	8	11
30-34	4	5	9
Total	31	23	54

There's a package (or two) for that!

- GaussSuppression
- sdcTable
- easySdcTable
- pTable
- ACRO-R
- cellKey
- modulartabler

easySdcTable (1)

```
> agesex
# A tibble: 8 × 3
  agecat sex    count
  <chr>  <chr>  <dbl>
1 15-19  Male      16
2 20-24  Male       8
3 25-29  Male       3
4 30-34  Male       4
5 15-19  Female     0
6 20-24  Female    10
7 25-29  Female     8
8 30-34  Female     5
```

easySdcTable (2)

```
> pivot_wider(agesex, id_cols=agecat, names_from=sex,
              values_from=count)
# A tibble: 4 × 3
  agecat  Male Female
  <chr>  <dbl>  <dbl>
1 15-19     16      0
2 20-24      8     10
3 25-29      3      8
4 30-34      4      5
```

easySdcTable (3)

```
> agesex.p <- ProtectTable(agesex, dimVar=c("agecat", "sex"),
              freqVar=c("count"), maxN=5, protectZeros=FALSE)
> pivot_wider(agesex.p$data, id_cols=agecat, names_from=sex,
              values_from=suppressed)
# A tibble: 5 × 4
  agecat Total Female  Male
  <chr>  <dbl>  <dbl> <dbl>
1 Total     54     23    31
2 15-19     16      0    16
3 20-24     18     10     8
4 25-29     11     NA    NA
5 30-34      9     NA    NA
```

Next steps

- easySdcTable + gtSummary = Table 1 !?!?