Paik-Agresti Diagrams

January 2020

A Paik-Agresti diagram let you see the marginal (unconditional) relationship between two binary variables and their conditional relationship, conditioning on another categorical variable in the same diagram.

The height of the marginal line is a weighted average of the heights of the conditional lines. The weights are the number of observations for each point on the conditional lines.

By representing the weights by the size of a circle around each point, the Paik-Agresti diagram makes it easy to see how and when Simpson's Paradox (the slope of the conditional lines has a sign opposite to that of the marginal line) can occur.

```
library(asbio)  # install.packages('asbio')
```

```
   Loading required package: tcltk
```

```
library(spida2) # devtools::install_github('gmonette/spida2')
```

```
   spida2: development branch 0.2.0.9000.
```

```
library(magrittr)  # install.packages('magrittr') # to use 'pipes' (%>%) with Ctrl-Shift-M
```

Sentences for murder convictions in Florida

The following dataset records whether the death penalty was pronounced in 674 homicide trials in the state of Florida from 1976-1987. The data set shows the verdict, and the defendant's and victim's race.

See Radelet, M. L., and G. L. Pierce (1991) Choosing those who will die: race and the death penalty in Florida. Florida Law Review 43(1):1-34.

```
data(death.penalty) # from Agresti 2012
death.penalty
```

```
   count verdict d.race v.race
1     53       Y      W      W
2     11       Y      B      W
3      0       Y      W      B
4      4       Y      B      B
5    414       N      W      W
6     37       N      B      W
7     16       N      W      B
8    139       N      B      B
```

The relationship between defendant's race and penalty:

```
tab( count ~ verdict + d.race, death.penalty)
```

```
         d.race
verdict   B   W Total
  N     176 430   606
  Y      15  53    68
```

```
     Total 191 483   674
```

```
tab( count ~ verdict + d.race, death.penalty, pct = 2)  %>%  round(2)
```

```
, , v.race = B

       d.race
verdict      B      W    All
  N      92.15  89.03  89.91
  Y       7.85  10.97  10.09
  Total 100.00 100.00 100.00
```

Conditional counts: conditioning on race of victim

```
tab( count ~ verdict + d.race + v.race, death.penalty)
```

```
, , v.race = B

       d.race
verdict   B   W Total
  N     139  16   155
  Y       4   0     4
  Total 143  16   159

, , v.race = W

       d.race
verdict   B   W Total
  N      37 414   451
  Y      11  53    64
  Total  48 467   515

, , v.race = Total

       d.race
verdict   B   W Total
  N     176 430   606
  Y      15  53    68
  Total 191 483   674
```

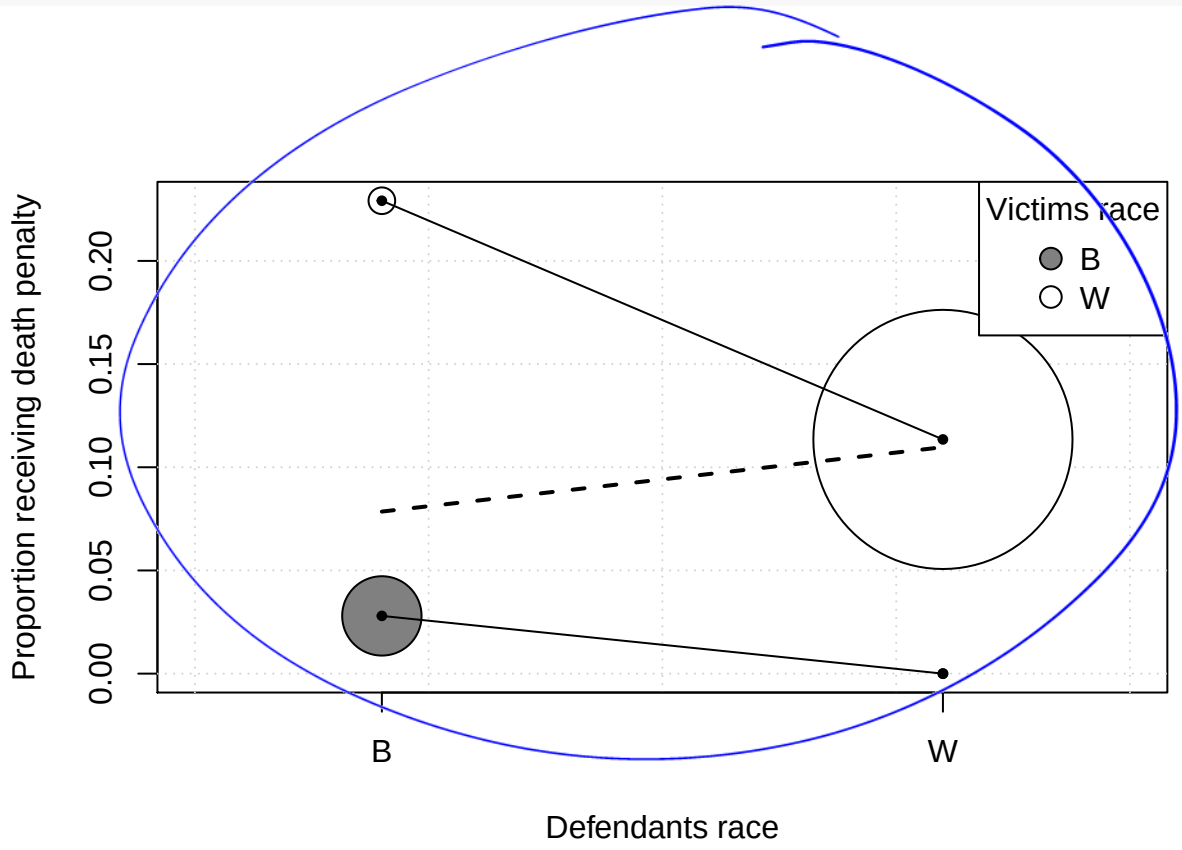
```
tab( count ~ verdict + d.race + v.race, death.penalty, pct = c(2,3))  %>%  round(2)
```

```
, , v.race = B

       d.race
verdict      B      W    All
  N      97.20 100.00  97.48
  Y       2.80   0.00   2.52
  Total 100.00 100.00 100.00

, , v.race = W

       d.race
verdict      B      W    All
  N      77.08  88.65  87.57
  Y      22.92  11.35  12.43
  Total 100.00 100.00 100.00

, , v.race = All
```

```
         d.race
verdict      B      W    All
  N      92.15  89.03  89.91
  Y       7.85  10.97  10.09
  Total 100.00 100.00 100.00
```

Paik-Agresti diagram

```
paik(verdict ~ d.race + v.race, counts = count, data = death.penalty,
     leg.title = "Victims race", xlab = "Defendants race",
     ylab = "Proportion receiving death penalty")
```



Berkeley admissions

The following data shows admissions in 1973 to graduate programmes at Berkeley by department A, B, C, D, E, gender and outcome

```
bad <- read.table(header=T, text = "
Dept  Gender  Status   count
A     Male    Admitted  512
A     Male    Denied    313
A     Female  Admitted   89
A     Female  Denied     19
B     Male    Admitted  313
B     Male    Denied    207
B     Female  Admitted   17
```

```
B     Female  Denied      8
C     Male    Admitted  120
C     Male    Denied    205
C     Female  Admitted  202
C     Female  Denied    391
D     Male    Admitted  138
D     Male    Denied    279
D     Female  Admitted  131
D     Female  Denied    244
E     Male    Admitted   53
E     Male    Denied    138
E     Female  Admitted   94
E     Female  Denied    299
F     Male    Admitted   22
F     Male    Denied    351
F     Female  Admitted   24
F     Female  Denied    317
")
bad
```

```
   Dept Gender   Status count
1     A   Male Admitted   512
2     A   Male   Denied   313
3     A Female Admitted    89
4     A Female   Denied    19
5     B   Male Admitted   313
6     B   Male   Denied   207
7     B Female Admitted    17
8     B Female   Denied     8
9     C   Male Admitted   120
10    C   Male   Denied   205
11    C Female Admitted   202
12    C Female   Denied   391
13    D   Male Admitted   138
14    D   Male   Denied   279
15    D Female Admitted   131
16    D Female   Denied   244
17    E   Male Admitted    53
18    E   Male   Denied   138
19    E Female Admitted    94
20    E Female   Denied   299
21    F   Male Admitted    22
22    F   Male   Denied   351
23    F Female Admitted    24
24    F Female   Denied   317
```

Marginal counts: not conditioning on department

```
tab(bad, count ~ Gender + Status)
```

```
          Status
Gender   Admitted Denied Total
  Female      557   1278  1835
  Male       1158   1493  2651
  Total      1715   2771  4486
```

```
tab(bad, count ~ Gender + Status, pct = 1) %>% round(2)
```

```
          Status
Gender    Admitted Denied  Total
  Female    30.35  69.65 100.00
  Male      43.68  56.32 100.00
  All       38.23  61.77 100.00
```

Conditional counts: conditioning on department

```
tab(bad, count ~ Gender + Status + Dept)
```

```
, , Dept = A

          Status
Gender    Admitted Denied Total
  Female        89     19   108
  Male         512    313   825
  Total        601    332   933

, , Dept = B

          Status
Gender    Admitted Denied Total
  Female        17      8    25
  Male         313    207   520
  Total        330    215   545

, , Dept = C

          Status
Gender    Admitted Denied Total
  Female       202    391   593
  Male         120    205   325
  Total        322    596   918

, , Dept = D

          Status
Gender    Admitted Denied Total
  Female       131    244   375
  Male         138    279   417
  Total        269    523   792

, , Dept = E

          Status
Gender    Admitted Denied Total
  Female        94    299   393
  Male          53    138   191
  Total        147    437   584

, , Dept = F

          Status
```

```
Gender   Admitted Denied Total
  Female       24    317   341
  Male         22    351   373
  Total        46    668   714

, , Dept = Total

        Status
Gender   Admitted Denied Total
  Female      557   1278  1835
  Male       1158   1493  2651
  Total      1715   2771  4486
```

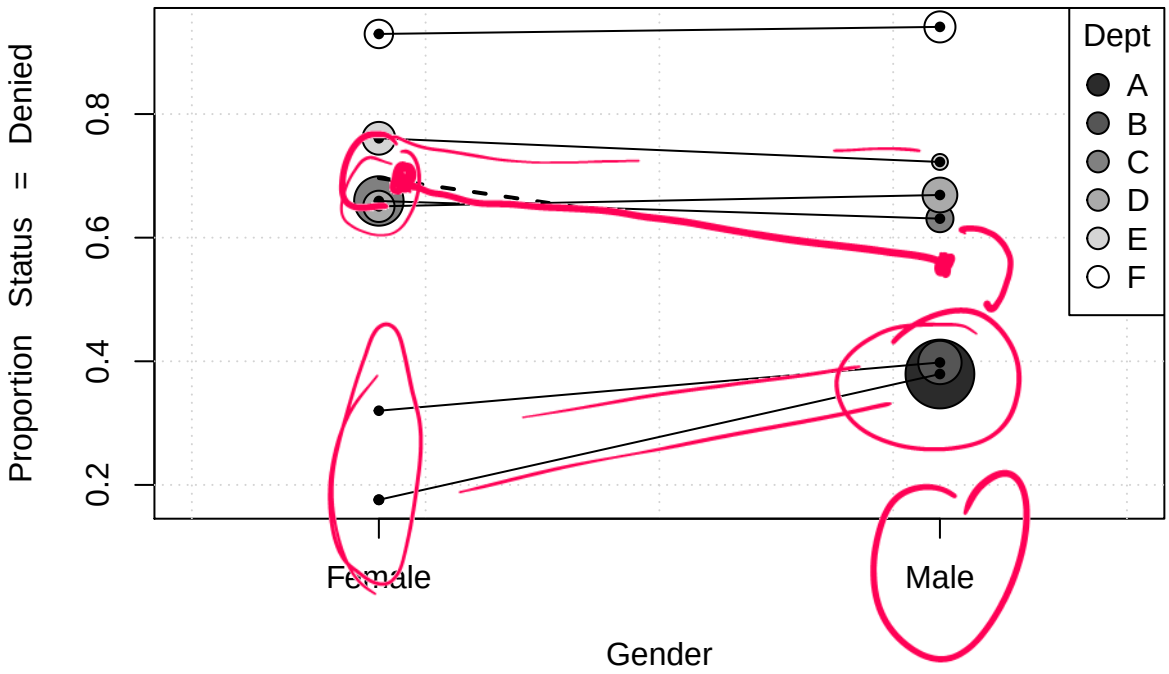
```
tab(bad, count ~ Gender + Status + Dept, pct = c(1,3))  %>% round(2)
```

```
, , Dept = A

        Status
Gender   Admitted Denied  Total
  Female    82.41  17.59 100.00
  Male      62.06  37.94 100.00
  All       64.42  35.58 100.00

, , Dept = B

        Status
Gender   Admitted Denied  Total
  Female    68.00  32.00 100.00
  Male      60.19  39.81 100.00
  All       60.55  39.45 100.00

, , Dept = C

        Status
Gender   Admitted Denied  Total
  Female    34.06  65.94 100.00
  Male      36.92  63.08 100.00
  All       35.08  64.92 100.00

, , Dept = D

        Status
Gender   Admitted Denied  Total
  Female    34.93  65.07 100.00
  Male      33.09  66.91 100.00
  All       33.96  66.04 100.00

, , Dept = E

        Status
Gender   Admitted Denied  Total
  Female    23.92  76.08 100.00
  Male      27.75  72.25 100.00
  All       25.17  74.83 100.00
```

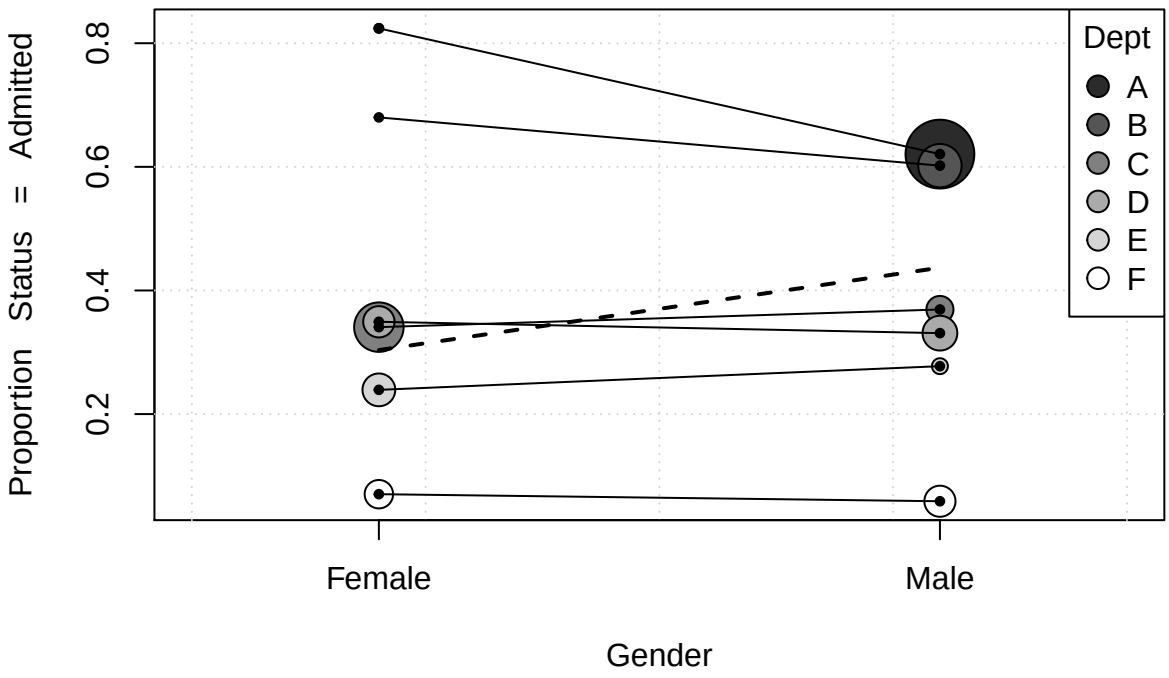
```
, , Dept = F

        Status
Gender   Admitted Denied  Total
  Female     7.04  92.96 100.00
  Male       5.90  94.10 100.00
  All        6.44  93.56 100.00

, , Dept = All

        Status
Gender   Admitted Denied  Total
  Female    30.35  69.65 100.00
  Male      43.68  56.32 100.00
  All       38.23  61.77 100.00
```

Paik diagram

```
paik( Status ~ Gender + Dept,  data = bad) # 'paik' expects 'counts' to be called 'count' in data.frame
```




```
bad$Status <- factor(bad$Status, levels = c('Denied','Admitted')) # changes ordering of levels in facto

paik( Status ~ Gender + Dept,  data = bad)
```



But does this mean that Berkeley is not 'discriminating' against female applicants?