

Binned Residual Plot

Description

A function that plots averages of y versus averages of x and can be useful to plot residuals for logistic regression.

Usage

```
binnedplot(x ,y, nclass=NULL,
           xlab="Expected Values", ylab="Average residual",
           main="Binned residual plot",
           cex.pts=0.8, col.pts=1, col.int="gray", ...)
```

Arguments

| | |
|---------|--|
| x | The expected values from the logistic regression. |
| y | The residuals values from logistic regression (observed values minus expected values). |
| nclass | Number of categories (bins) based on their fitted values in which the data are divided. Default=NULL and will take the value of nclass according to the \$n\$ such that if \$n \geq 100\$, $nclass = \text{floor}(\sqrt{\text{length}(x)})$; if $10 < n < 100$, $nclass = 10$; if $n < 10$, $nclass = \text{floor}(n/2)$. |
| xlab | a label for the x axis, default is "Expected Values". |
| ylab | a label for the y axis, default is "Average residual". |
| main | a main title for the plot, default is "Binned residual plot". See also title. |
| cex.pts | The size of points, default=0.8. |
| col.pts | color of points, default is black |
| col.int | color of intervals, default is gray |
| ... | Graphical parameters to be passed to methods |

Details

In logistic regression, as with linear regression, the residuals can be defined as observed minus expected values. The data are discrete and so are the residuals. As a result, plots of raw residuals from logistic regression are generally not useful. The binned residuals plot instead, after dividing the data into categories (bins) based on their fitted values, plots the average residual versus the average fitted value for each bin.

Value

A plot in which the gray lines indicate plus and minus 2 standard-error bounds, within which one would expect about 95% of the binned residuals to fall, if the model were actually true.

Note

There is typically some arbitrariness in choosing the number of bins: each bin should contain enough points so that the averaged residuals are not too noisy, but it helps to have also many bins so as to see more local patterns in the residuals (see Gelman and Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, pag 97).

Author(s)

M. Grazia Pittau grazia@stat.columbia.edu; Yu-Sung Su suyusung@tsinghua.edu.cn

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

[par, plot](#)

Examples

```
old.par <- par(no.readonly = TRUE)
data(lalonde)
attach(lalonde)
fit <- glm(treat ~ re74 + re75 + educ + black + hisp + married
          + nodegr + u74 + u75, family=binomial(link="logit"))
x <- predict(fit)
y <- resid(fit)
binnedplot(x,y)
par(old.par)
```