

plot.glm

Chris Dardis

December 20, 2021

This is a minimal example showing the diagnostic plots.

1 Effect of probability on the diagnostic statistics

This is shown in table 1.

The relationship can be shown graphically as below:

```
p <- seq(5)
h <- c(1, 3, 1.5, 3, 1)
plot(p, h, type="b",
     col="blue", axes=FALSE,
     xlab="Probability",
     ylab="Size of statistic")
axis(1, at=p, labels=c("<0.1", "0.1-0.3", "0.3-0.7", "0.7-0.9", ">0.9"))
axis(2, at=c(1, 1.5, 2, 3), labels=c("low", "low-med", "med", "high"))
dChisq <- c(2, 1.5, 2)
points(c(2, 3, 4), dChisq, type="b", col="green")
lines(x=c(1, 2), y=c(3, 2), col="green")
lines(x=c(1, 2), y=c(1, 2), col="green")
lines(x=c(4, 5), y=c(2, 1), col="green")
lines(x=c(4, 5), y=c(2, 3), col="green")
dBhat <- c(1, 3, 2, 3, 1)
points(p, dBhat, type="b", col="red")
legend(2.5, y=1.4, legend=c("h", "dChisq", "dBhat"),
      fill=c("blue", "green", "red"))
mtext("Probability vs. Size of statistic")
```

P	h	$d\chi^2$	$d\hat{\beta}$
<0.1	l	l / h	l
0.1 - 0.3	h	m	h
0.3 - 0.7	l - m	l - m	m
0.7 - 0.9	h	m	h
>0.9	l	l / h	l

Abbreviations:

l = low

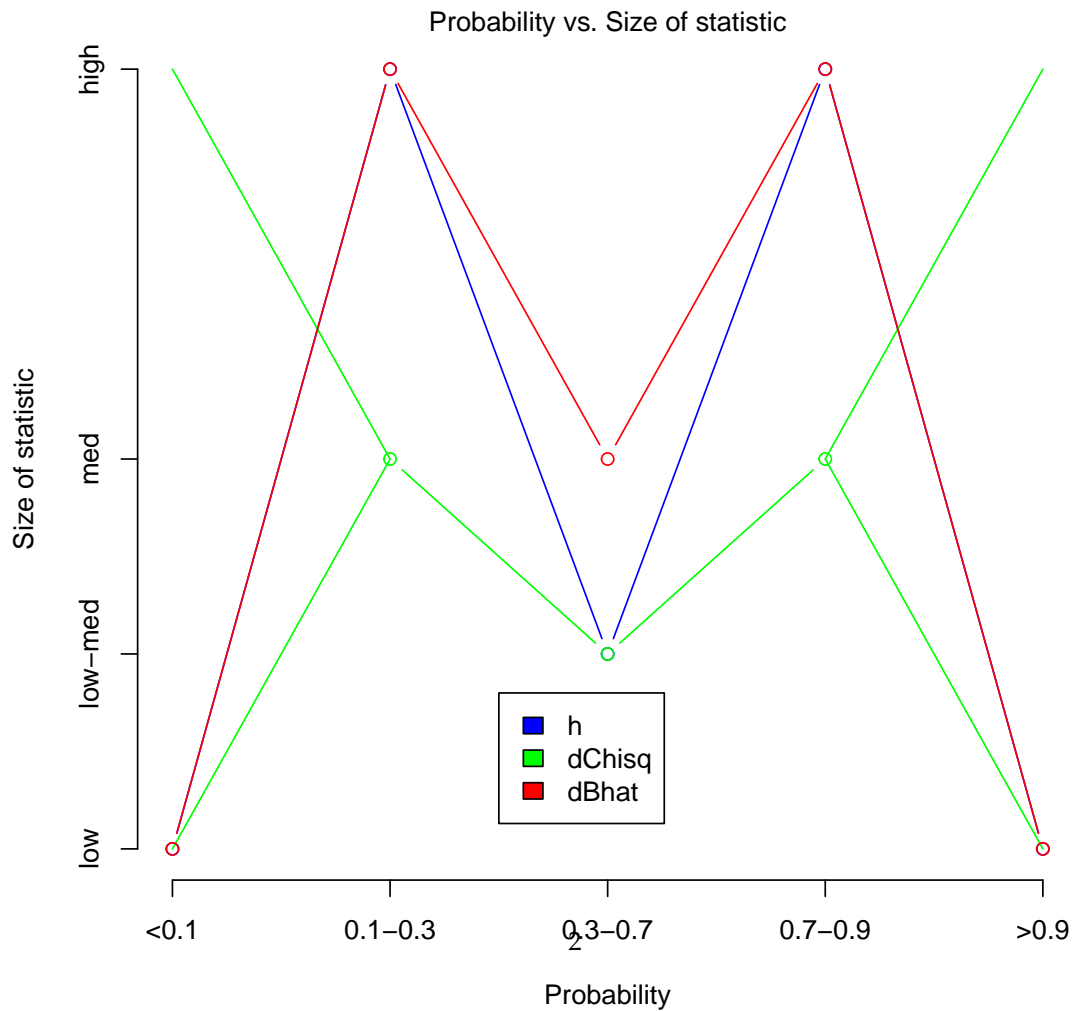
l - m = low to medium

m = medium

l / h = low or high

h = high

Table 1: Effect of probability on diagnostics



2 Sample graphical output from plot.glm

```
library("LogisticDx")
print("H&L 2nd ed. Table 4.9. Figures 5.5-5.8. Pages 177-180.")

## [1] "H&L 2nd ed. Table 4.9. Figures 5.5-5.8. Pages 177-180."

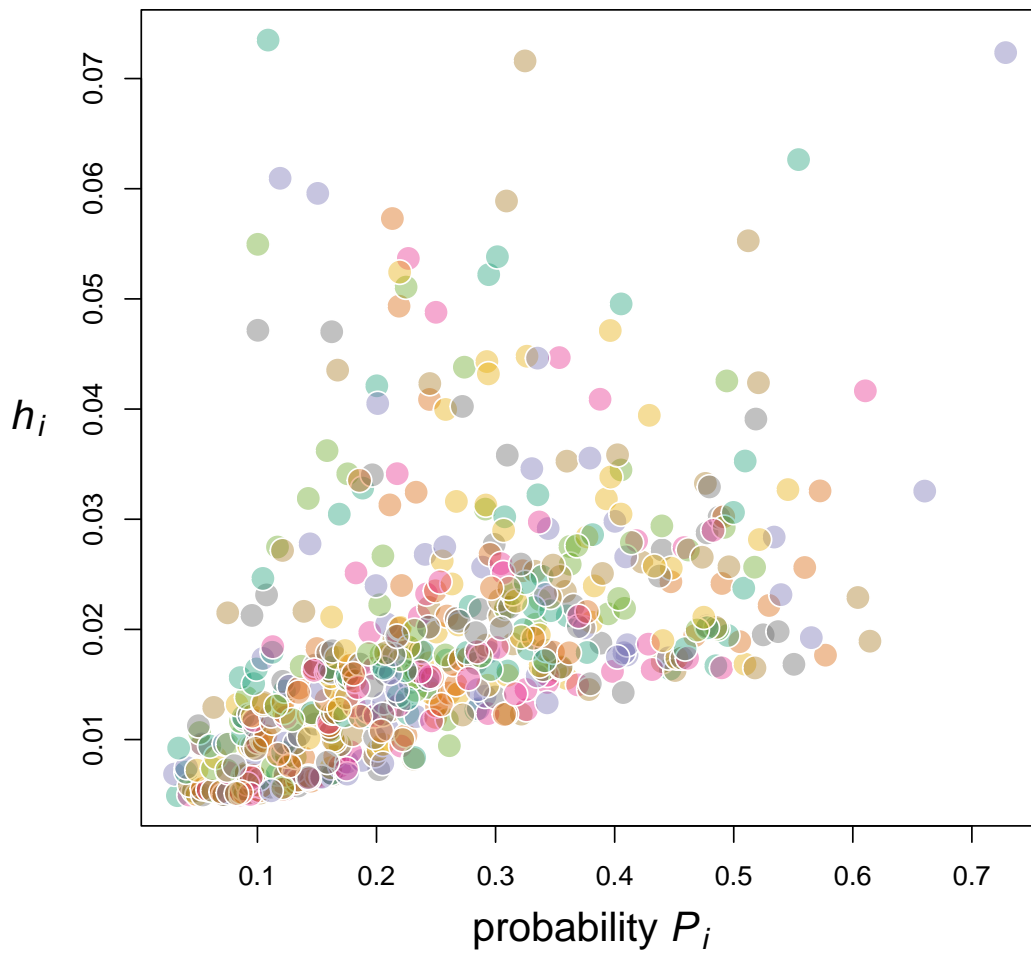
data(uis)
uis <- within(uis, {
  NDRGFP1 <- 10 / (NDRGTX + 1)
  NDRGFP2 <- NDRGFP1 * log((NDRGFP1 + 1) / 10)
})
summary(glm(DFREE ~ AGE + NDRGFP1 + NDRGFP2 + IVHX +
  RACE + TREAT + SITE +
  AGE:NDRGFP1 + RACE:SITE,
  family=binomial, data=uis))

##
## Call:
## glm(formula = DFREE ~ AGE + NDRGFP1 + NDRGFP2 + IVHX + RACE +
##   TREAT + SITE + AGE:NDRGFP1 + RACE:SITE, family = binomial,
##   data = uis)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -1.3036  -0.7914  -0.5783   0.9902   2.6024
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.607764   1.177521  -5.612 2.00e-08 ***
## AGE           0.116601   0.028859   4.040 5.34e-05 ***
## NDRGFP1      0.690191   0.209629   3.292 0.000993 ***
## NDRGFP2     -0.461492   0.123824  -3.727 0.000194 ***
## IVHXprevious -0.633915   0.298774  -2.122 0.033862 *
## IVHXrecent  -0.705257   0.261631  -2.696 0.007026 **
## RACEother    0.686261   0.264195   2.598 0.009389 **
## TREATlong   0.433709   0.203791   2.128 0.033320 *
## SITEB       0.516694   0.254937   2.027 0.042688 *
## AGE:NDRGFP1 -0.015260   0.006026  -2.532 0.011329 *
## RACEother:SITEB -1.431087   0.529872  -2.701 0.006917 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
## Null deviance: 653.73 on 574 degrees of freedom
## Residual deviance: 597.92 on 564 degrees of freedom
## AIC: 619.92
##
## Number of Fisher Scoring iterations: 4
plot(g1, devNew=FALSE)
```

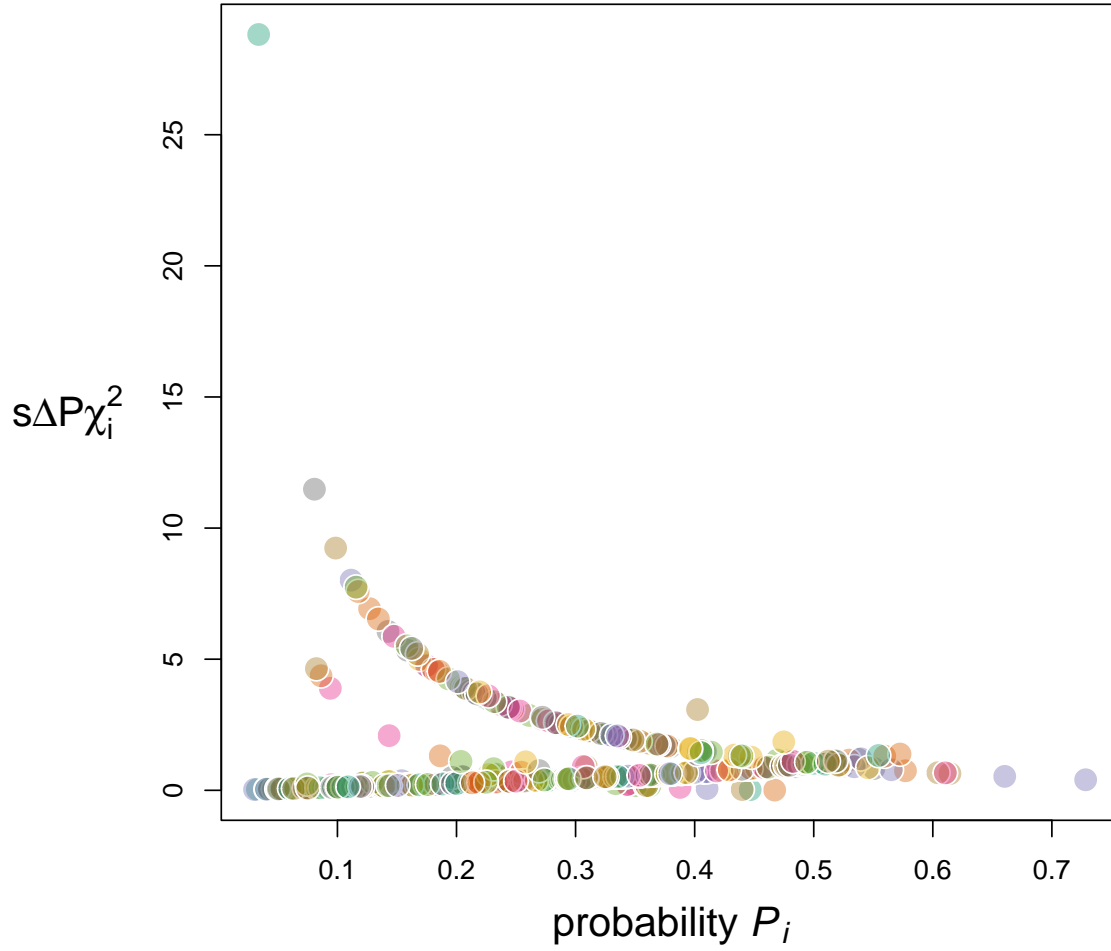
Probability P_i \times leverage h_i

$0.1 < P_i < 0.9 \rightarrow h_i \propto x_i - \mu_x$
 $h_i \approx$ distance of covariate pattern x_i from mean μ_x
 $h_i =$ diagonal of hat matrix



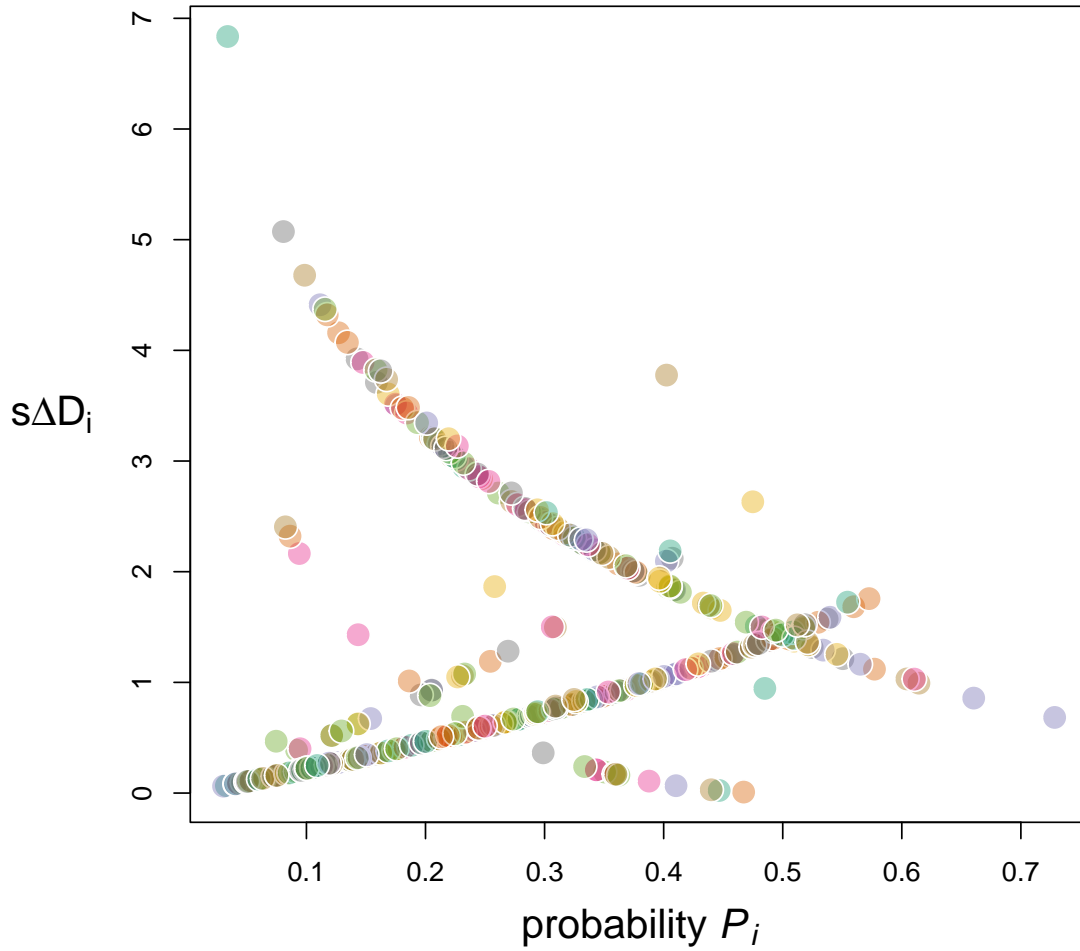
Probability P_i \times scaled change in Pearson chi-sq $s\Delta P\chi_i^2$

$$Pr_i = \frac{y_i - \mu_y}{\sigma_y}, s\Delta P\chi_i^2 = \frac{Pr_i}{\sqrt{1-h_i}}$$



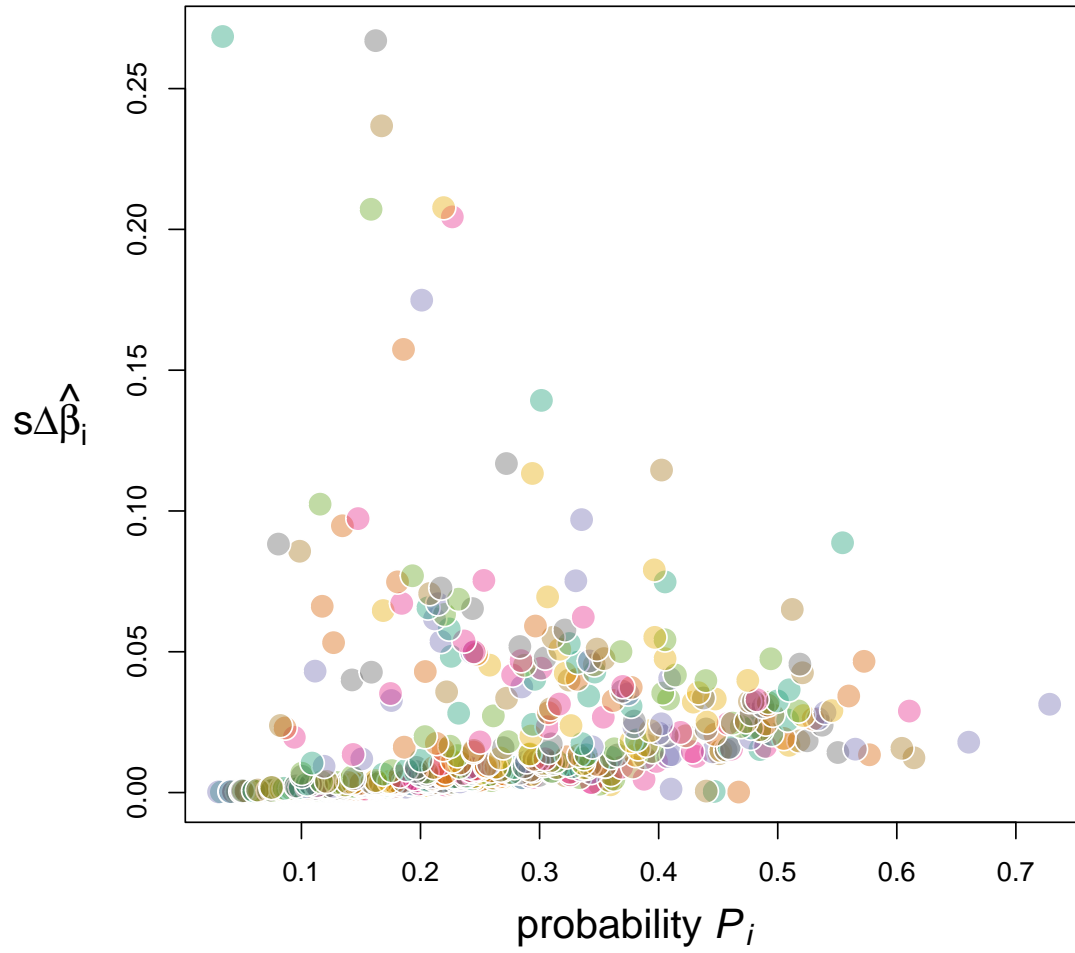
Probability P_i \times scaled change in deviance ΔD_i

$$dr_i = \text{sign}(y_i - \hat{y}_i) \sqrt{d_i}, \quad s\Delta D_i = \frac{dr_i}{\sqrt{1-h_i}}$$



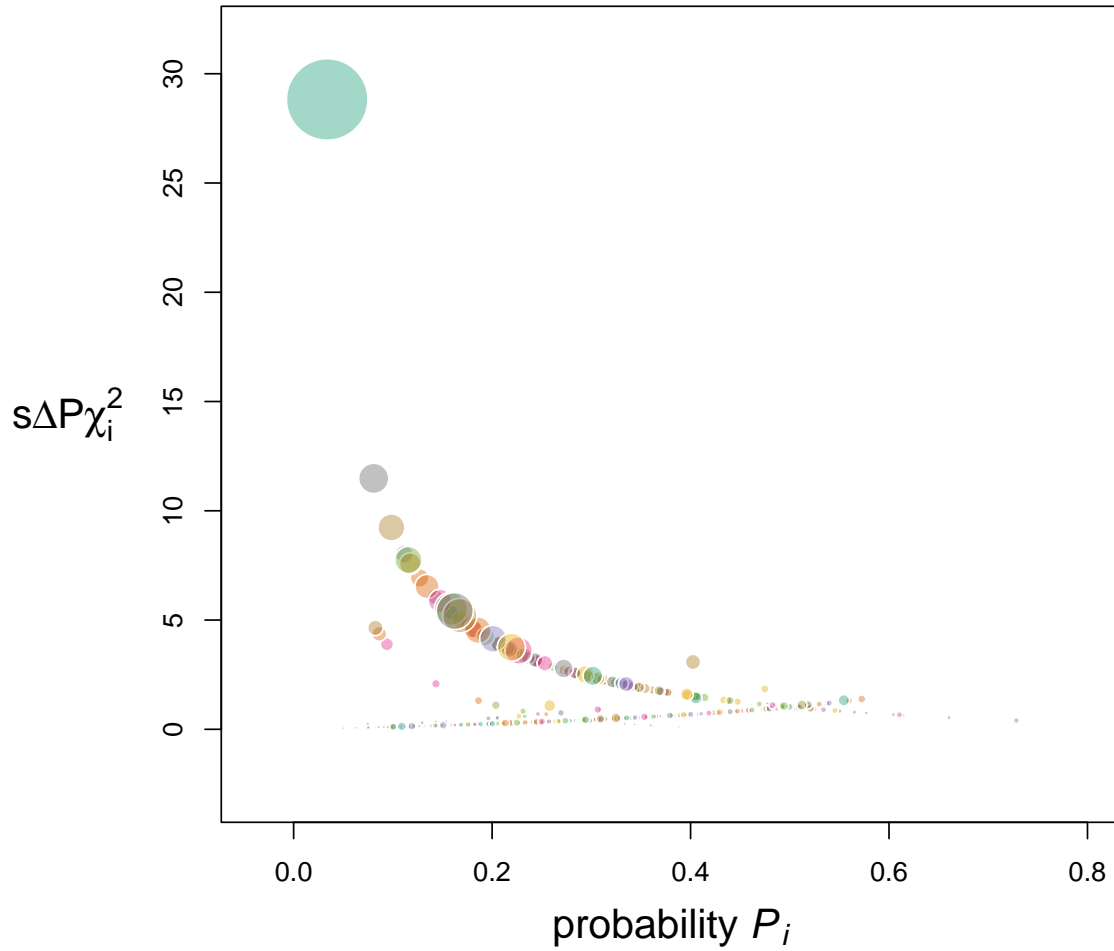
Probability P_i \times scaled change in coefficients $s\Delta\hat{\beta}_i$

$$s\Delta\hat{\beta}_i = \frac{sPr_i^2 h_i}{1-h_i}$$



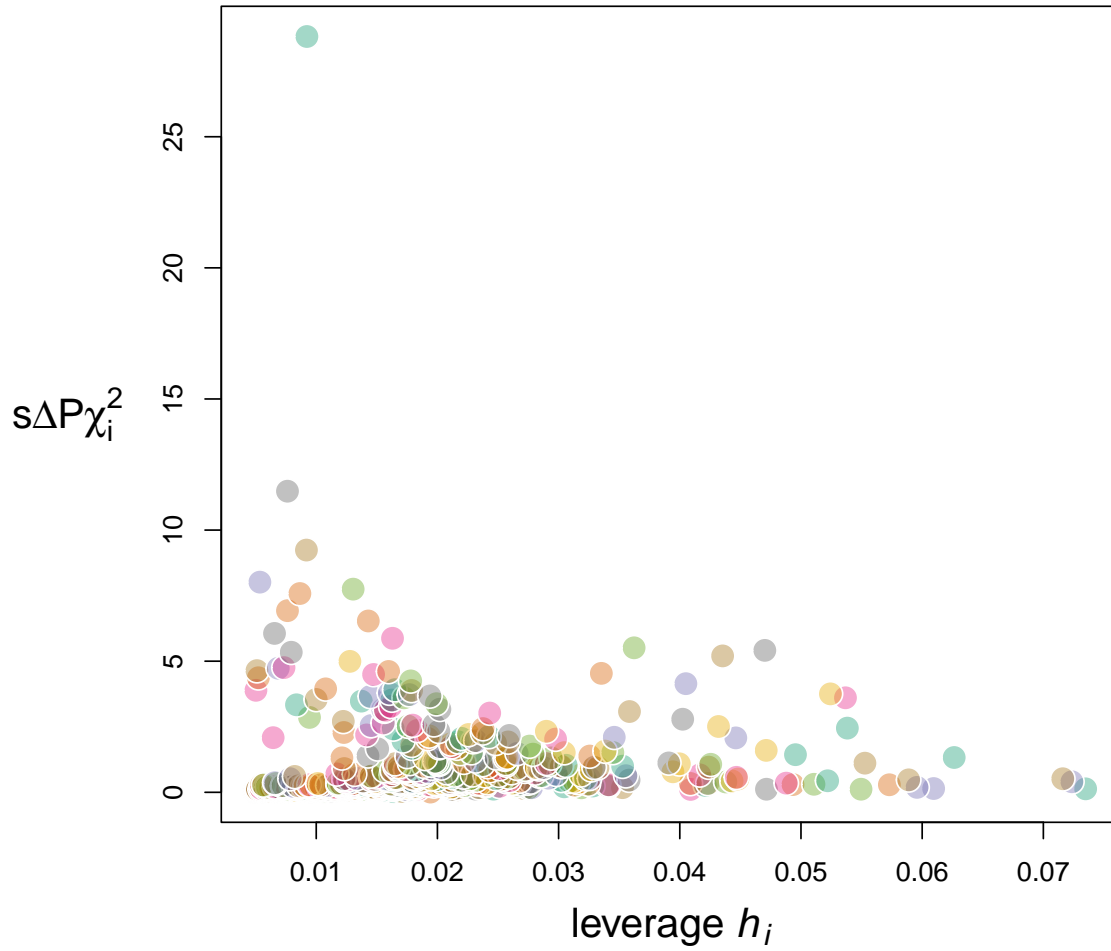
Probability P_i \times scaled change in Pearson chi-sq $s\Delta P_i$

$$\text{area} \propto s\Delta\hat{\beta}_i, \text{ radius} = \sqrt{\frac{s\Delta\hat{\beta}_i}{P_i}}$$



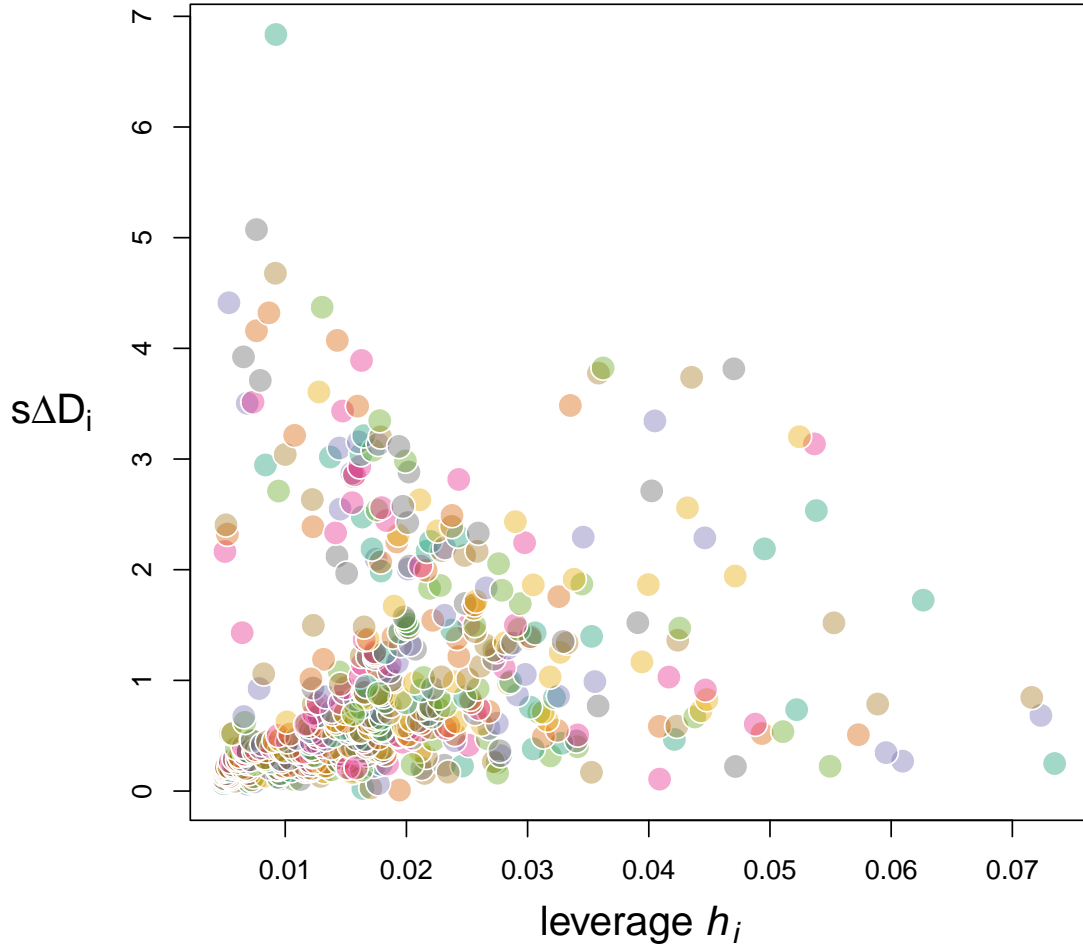
Leverage h_i \times scaled change in Pearson chi-sq $s\Delta P\chi$

$$h_i \approx x_i - \mu_x, s\Delta P\chi_i^2 = sPr_i^2$$



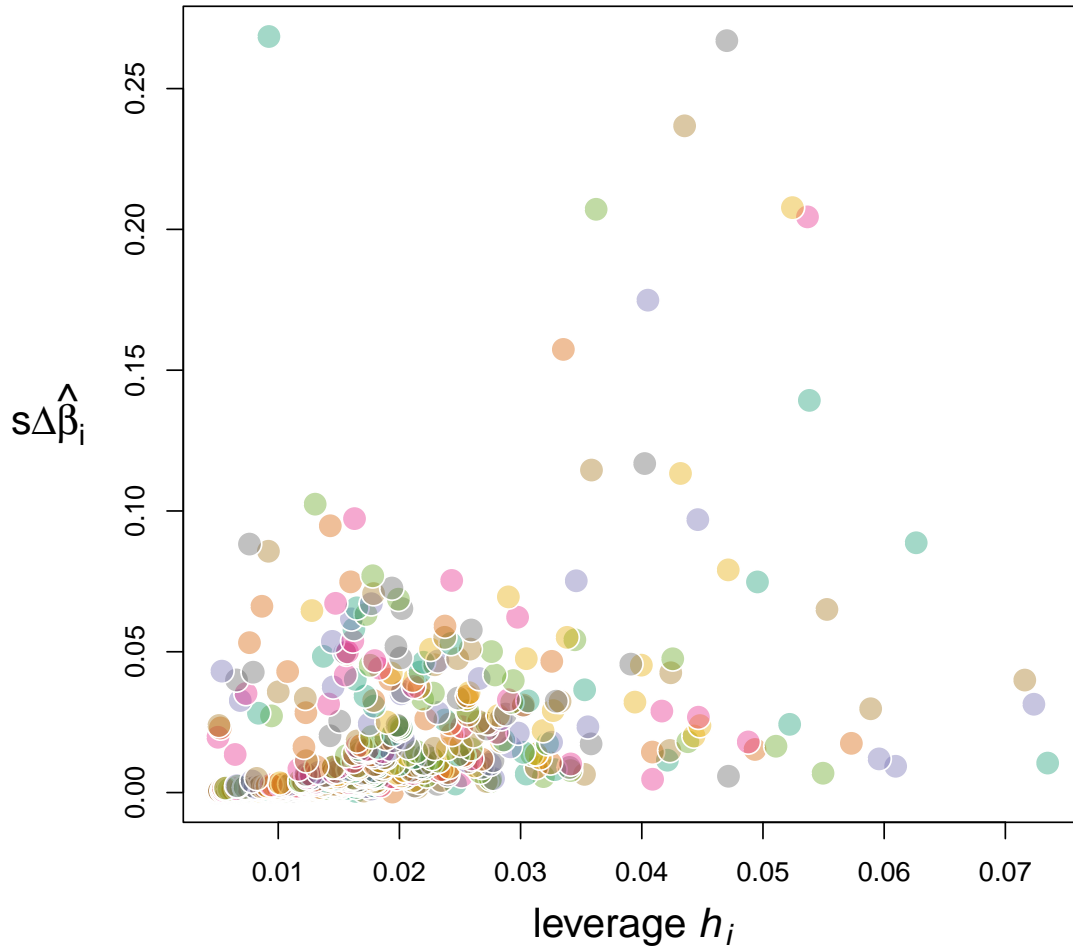
Leverage h_i \times scaled change in deviance $s\Delta D_i$

$$dr_i = \text{sign}(y_i - \hat{y}_i)\sqrt{d_i}, s\Delta D_i = \frac{dr_i}{\sqrt{1-h_i}}$$



Leverage h_i \times scaled change in coefficients $s\Delta\hat{\beta}_i$

$$h_i \approx x_i - \bar{x}, s\Delta\hat{\beta}_i = \frac{sPr_i^2 h_i}{1-h_i}$$



Correlation between $s\Delta P\chi_i^2$, $s\Delta D_i$ and $s\Delta\hat{\beta}_i$

