plot.glm

C Dardis

July 8, 2015

This is a minimal example showing the diagnostic plots.

1 Effect of probability on the diagnostic statistics

This is shown in table 1.

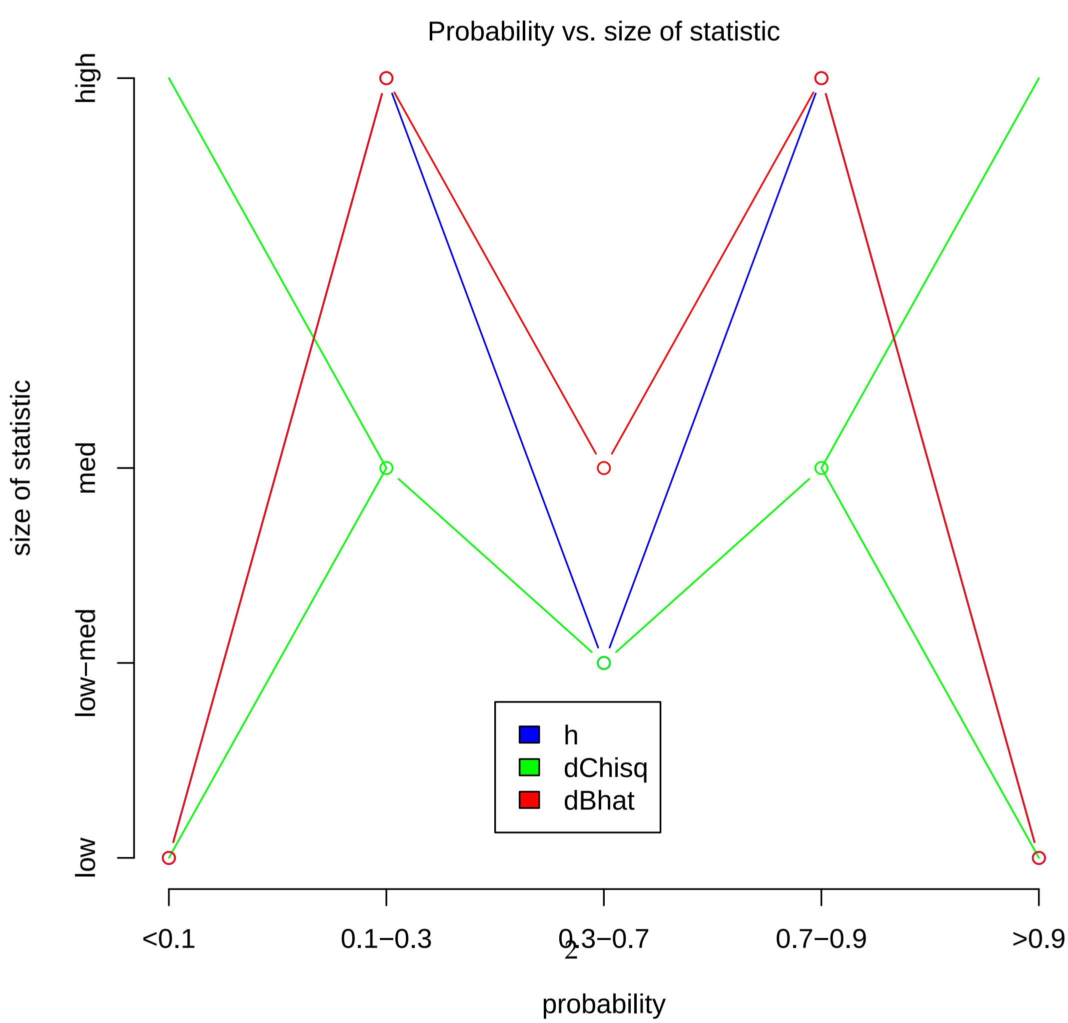
The relationship can be shown graphically as below:

```
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

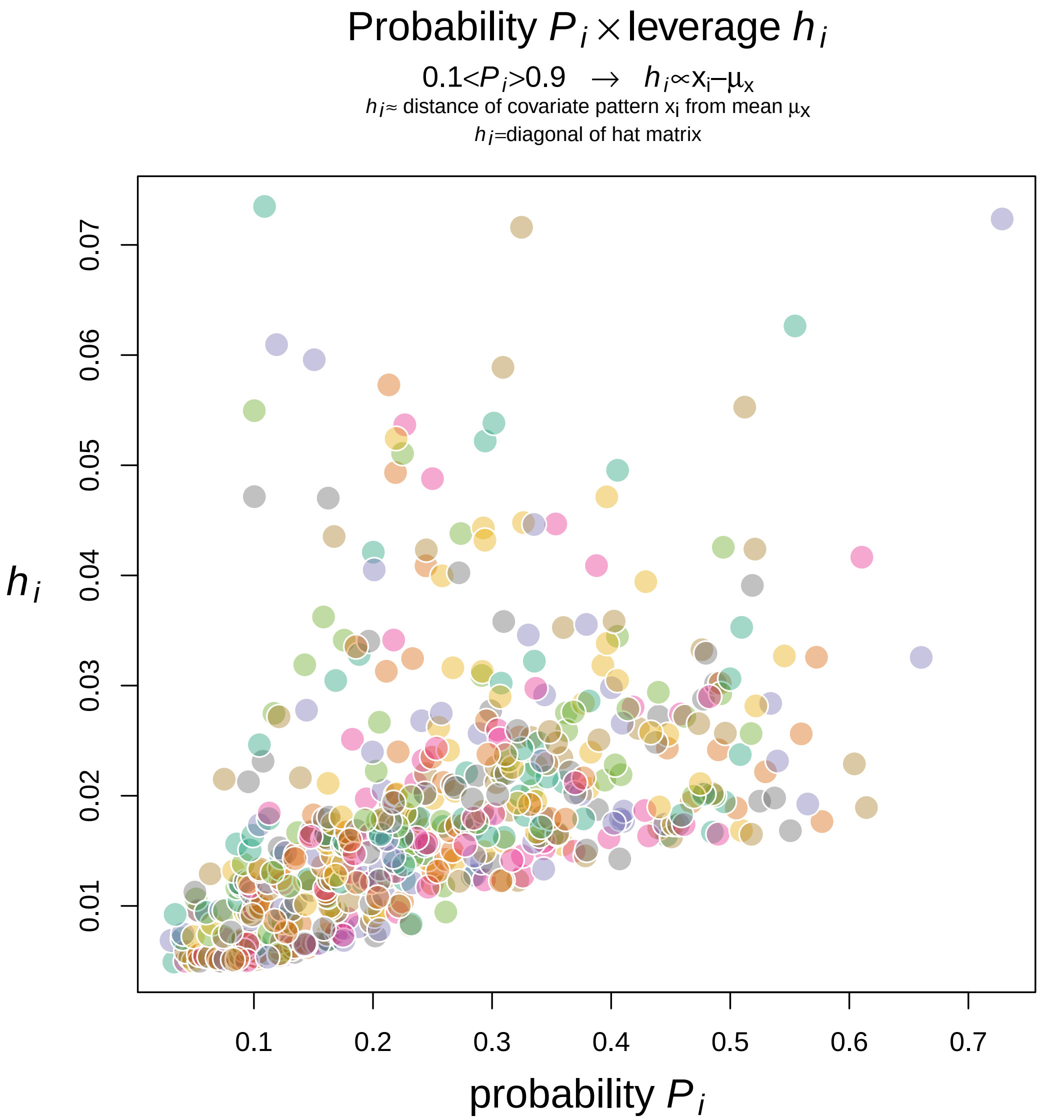


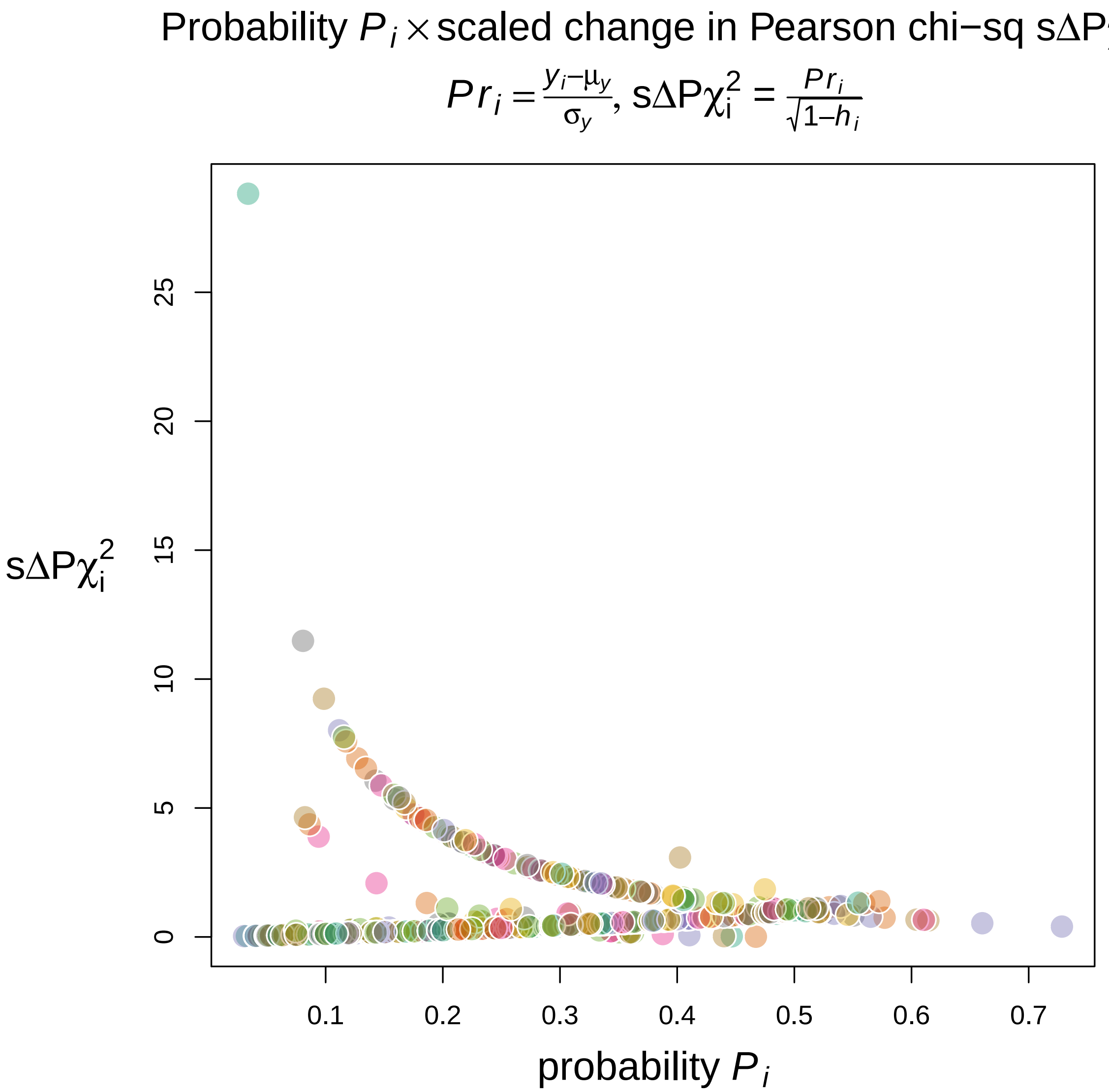
2 Sample graphical output from `plot.glm`

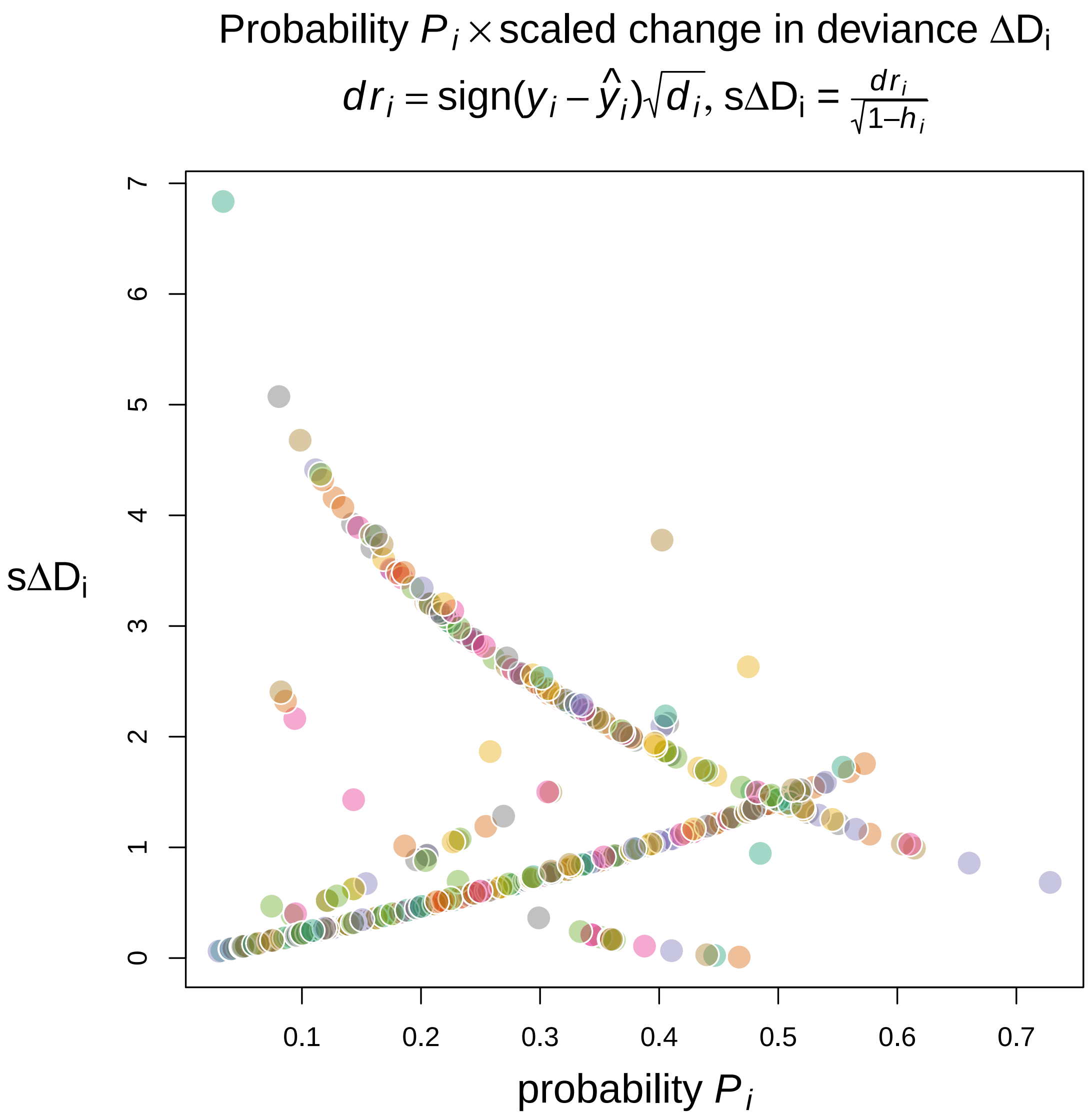
```
library("LogisticDx")
## H&L 2nd ed. Table 4.9. Figures 5.5-5.8. Pages 177-180.
data(uis)
uis <- within(uis, {
    NDRGFP1 <- 10 / (NDRGTX + 1)
    NDRGFP2 <- NDRGFP1 * log((NDRGFP1 + 1) / 10)
})
summary(g1 <- glm(DFREE ~ AGE + NDRGFP1 + NDRGFP2 + IVHX +
                  RACE + TREAT + SITE +
                  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
## -1.304  -0.791  -0.578   0.990   2.602
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.60776    1.17752   -5.61  2.0e-08 ***
## AGE             0.11660    0.02886    4.04  5.3e-05 ***
## NDRGFP1         0.69019    0.20963    3.29  0.00099 ***
## NDRGFP2        -0.46149    0.12382   -3.73  0.00019 ***
## IVHXprevious   -0.63391    0.29877   -2.12  0.03386 *
## IVHXrecent     -0.70526    0.26163   -2.70  0.00703 **
## RACEother       0.68626    0.26419    2.60  0.00939 **
## TREATlong       0.43371    0.20379    2.13  0.03332 *
## SITEB           0.51669    0.25494    2.03  0.04269 *
## AGE:NDRGFP1    -0.01526    0.00603   -2.53  0.01133 *
## RACEother:SITEB -1.43109    0.52987   -2.70  0.00692 **
## ---
## 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: 619.9
```

```
##
## Number of Fisher Scoring iterations: 4
```

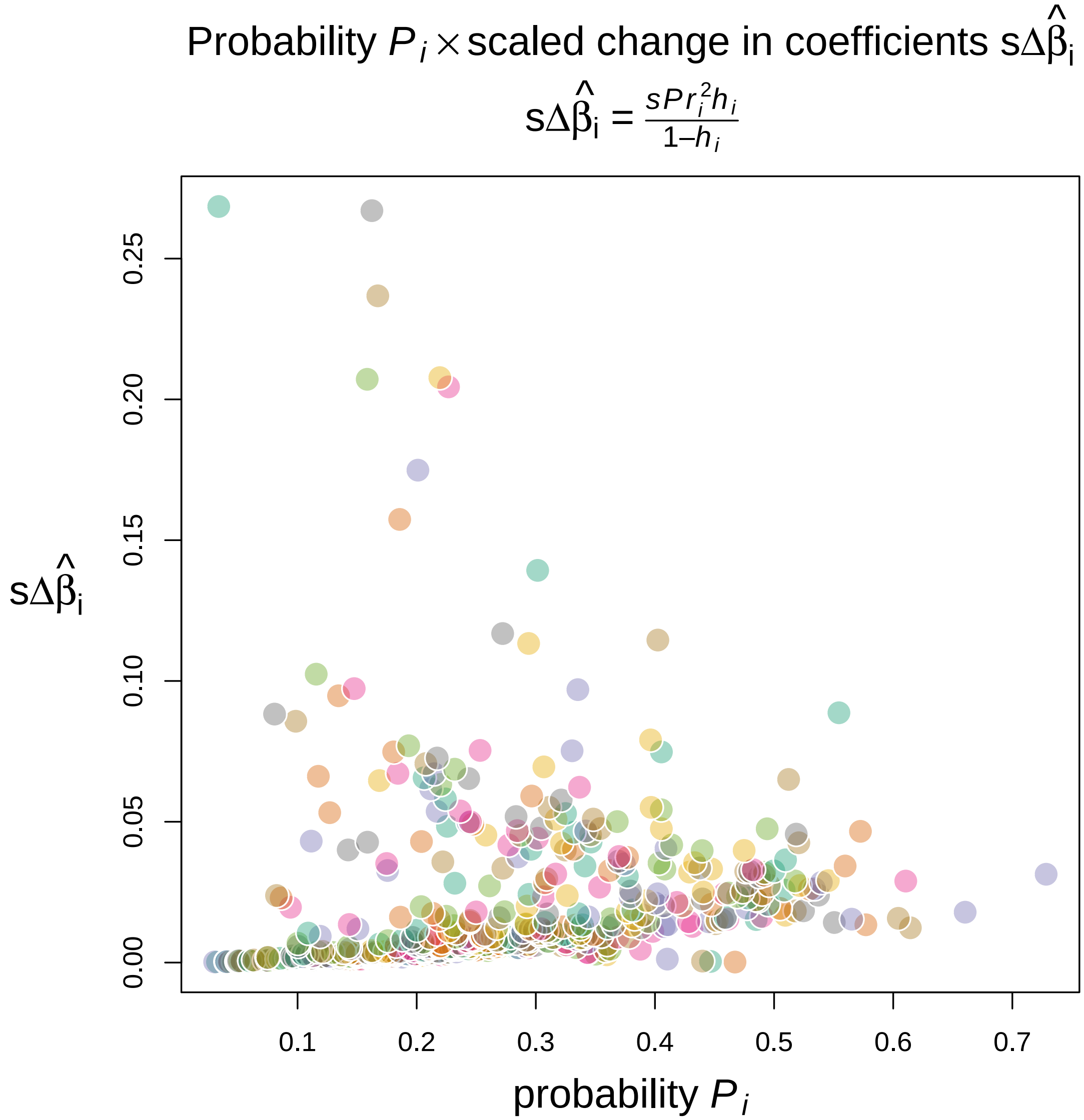
```
plot(g1, devNew=FALSE)
```

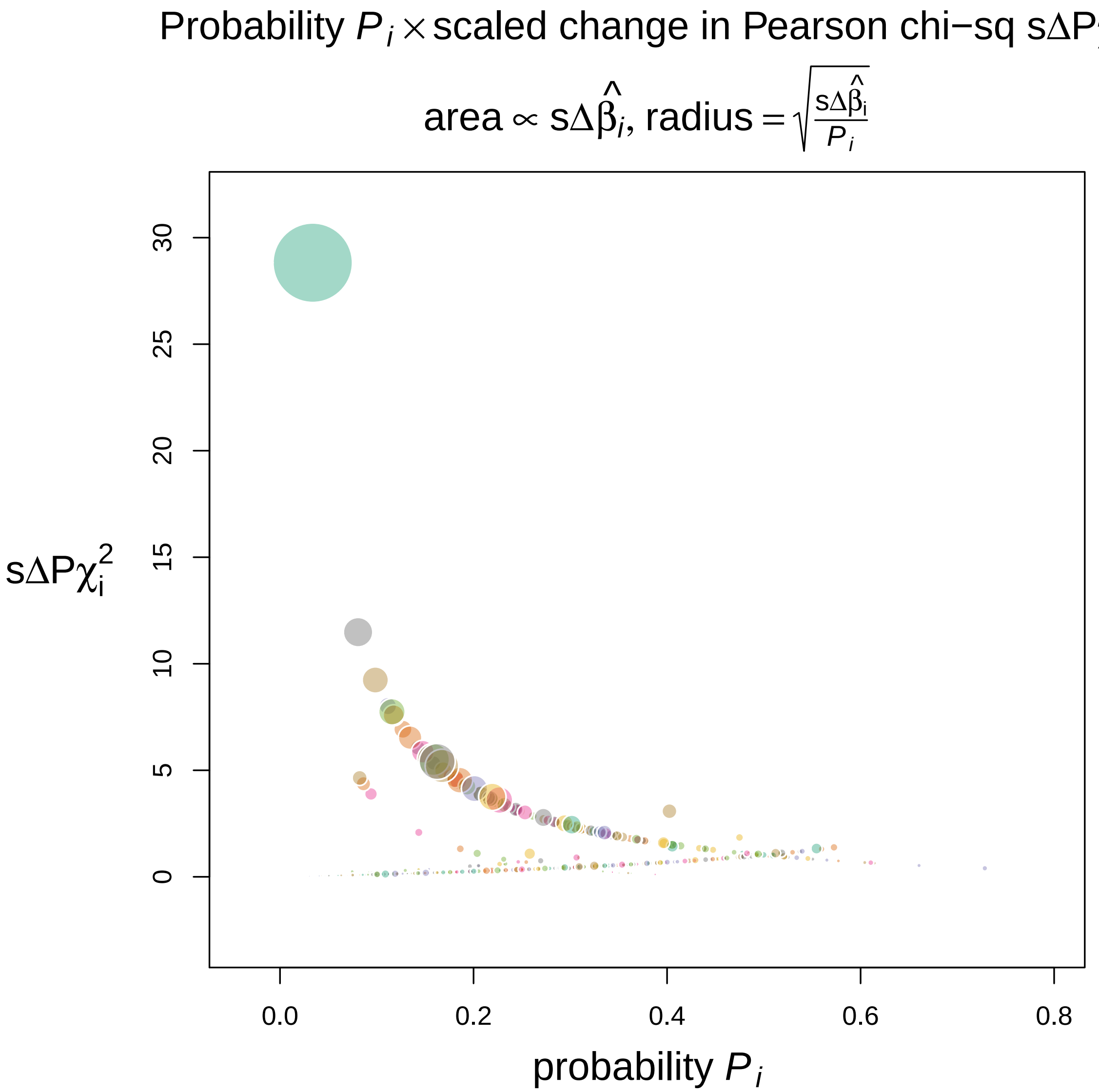



Probability $P_i \times$ scaled change in Pearson chi−sq $s\Delta P\chi$
$Pr_i = \frac{y_i - \mu_y}{\sigma_y}$, $s\Delta P\chi_i^2 = \frac{Pr_i}{\sqrt{1-h_i}}$
$s\Delta P\chi_i^2$
0
5
10
15
20
25
0.1
0.2
0.3
0.4
0.5
0.6
0.7
probability P_i

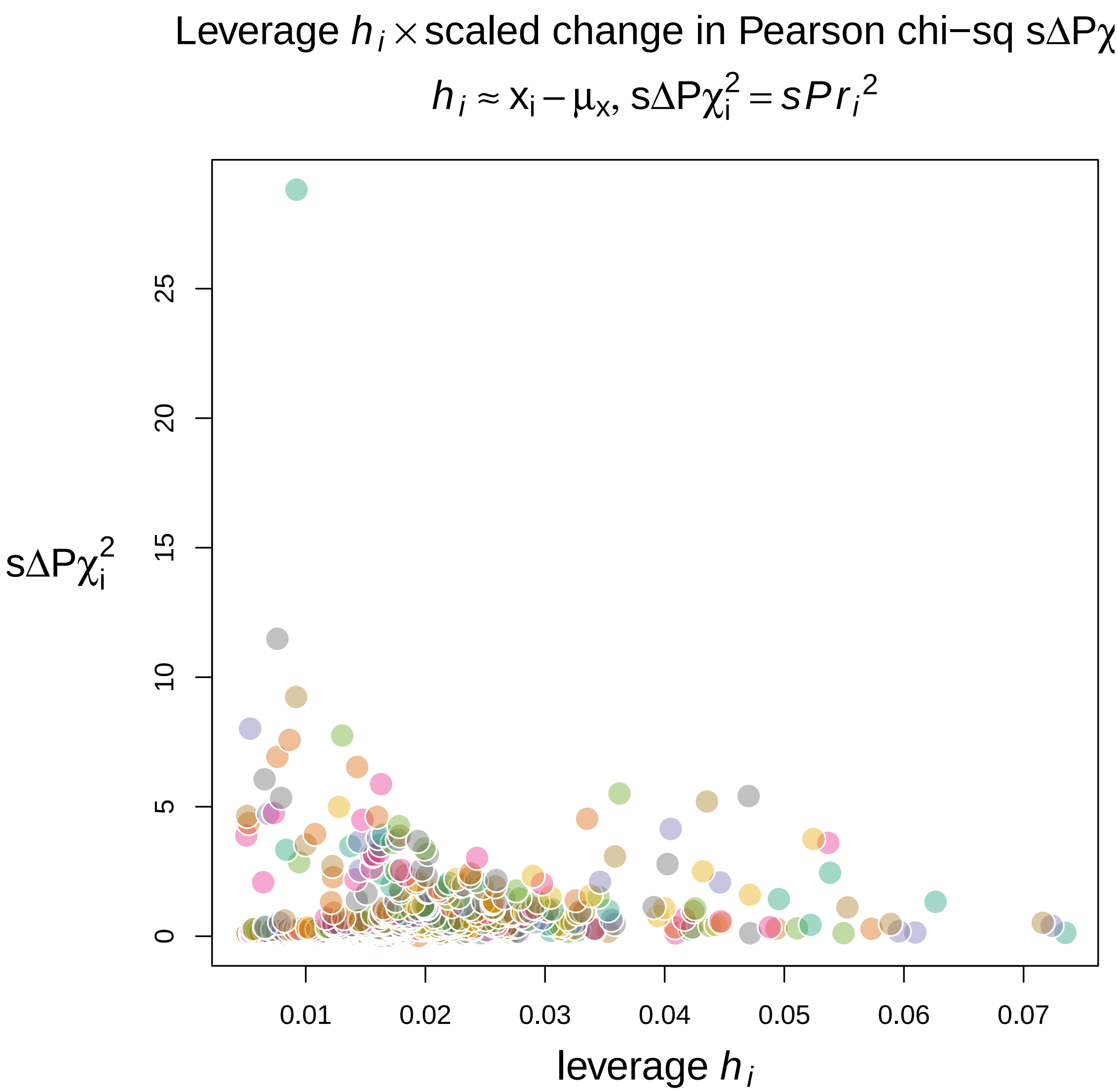


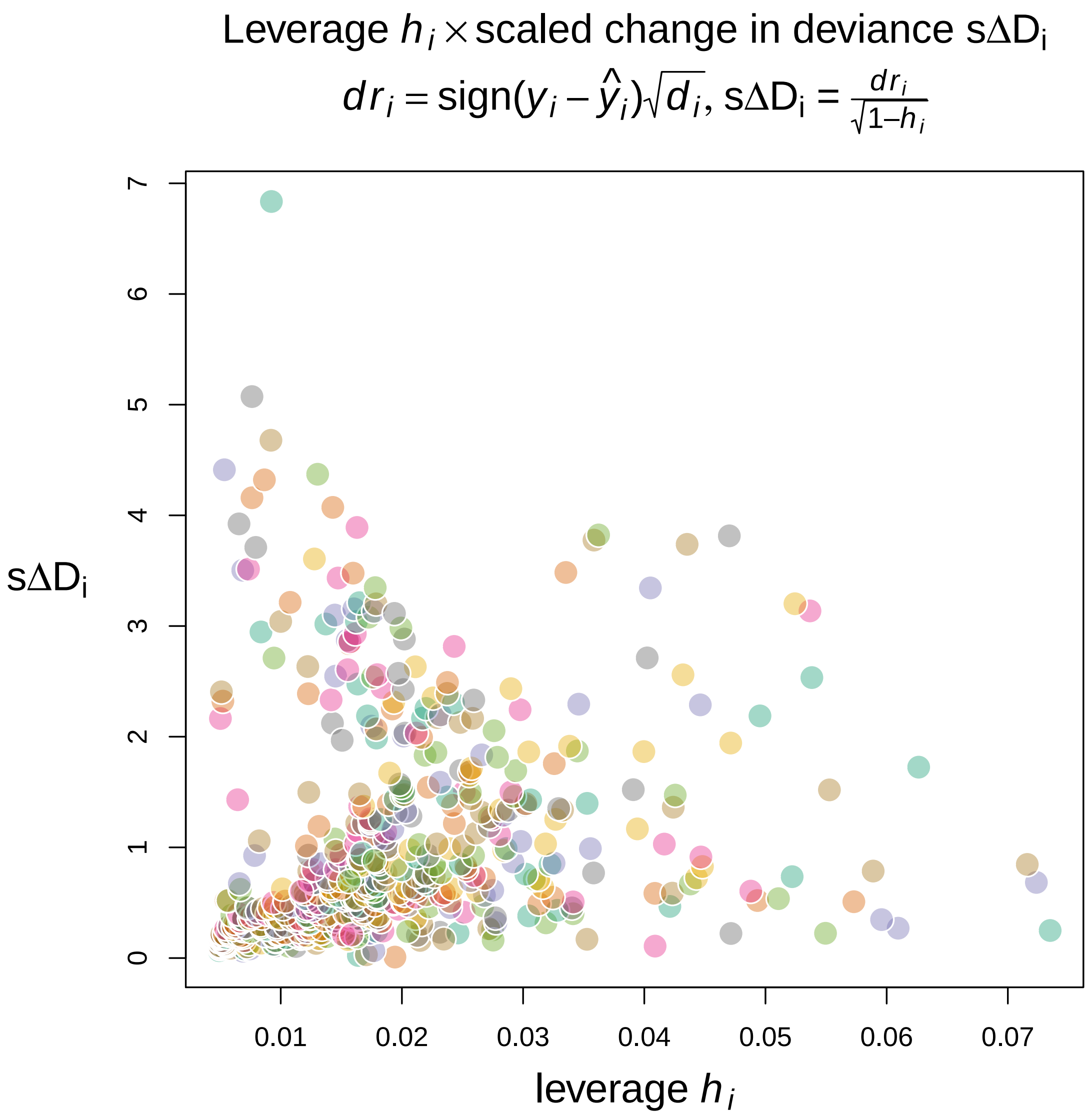
Probability $P_i \times$ scaled change in deviance ΔD_i
$dr_i = \text{sign}(y_i - \hat{y}_i)\sqrt{d_i}$, $s\Delta D_i = \frac{dr_i}{\sqrt{1-h_i}}$
$s\Delta D_i$
probability P_i
0
1
2
3
4
5
6
7
0.1
0.2
0.3
0.4
0.5
0.6
0.7

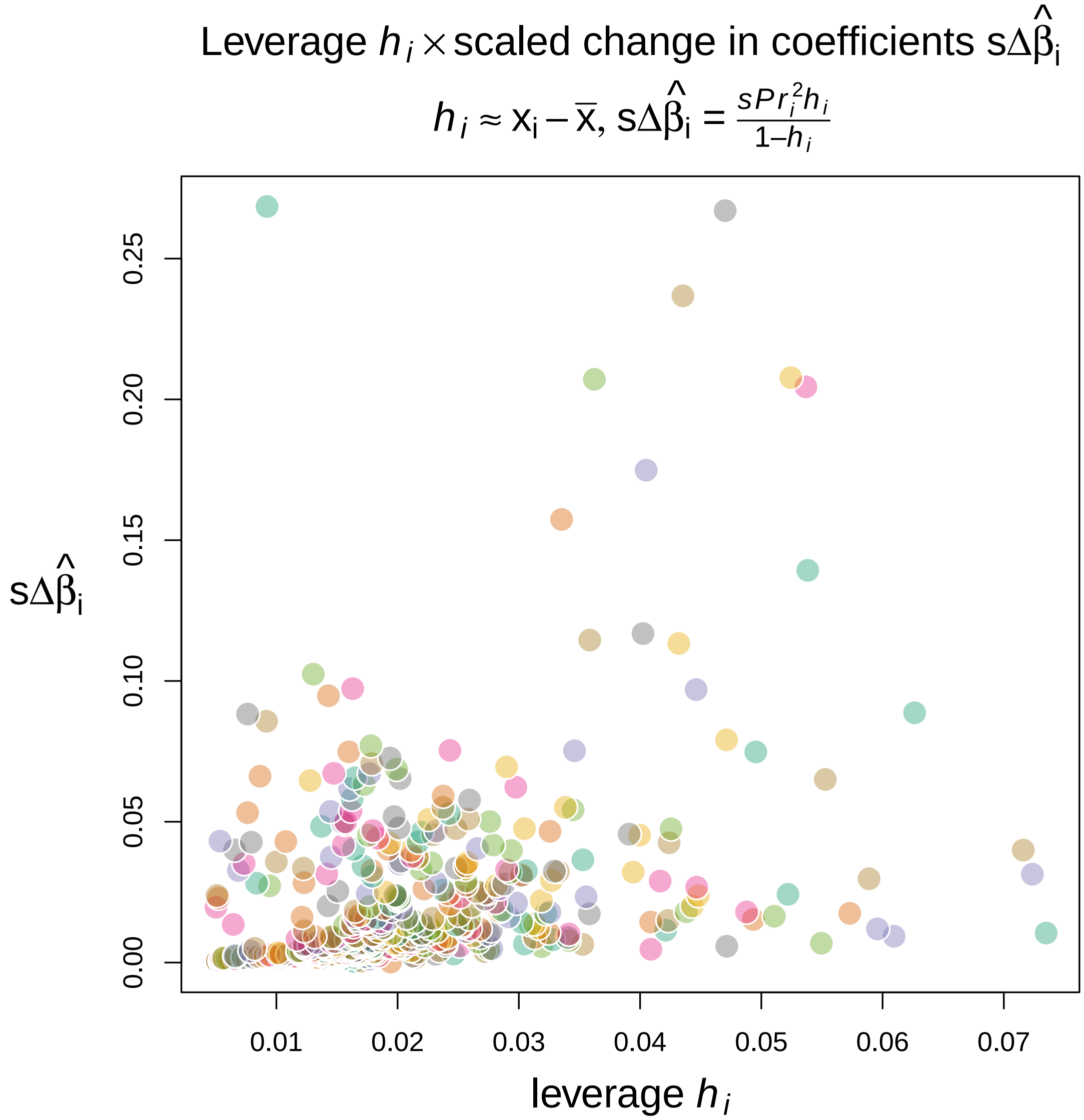

Probability $P_i \times$ scaled change in coefficients $s\Delta\hat{\beta}_i$
$s\Delta\hat{\beta}_i = \frac{sPr_i^2 h_i}{1-h_i}$
$s\Delta\hat{\beta}_i$
0.00
0.05
0.10
0.15
0.20
0.25
0.1
0.2
0.3
0.4
0.5
0.6
0.7
probability P_i

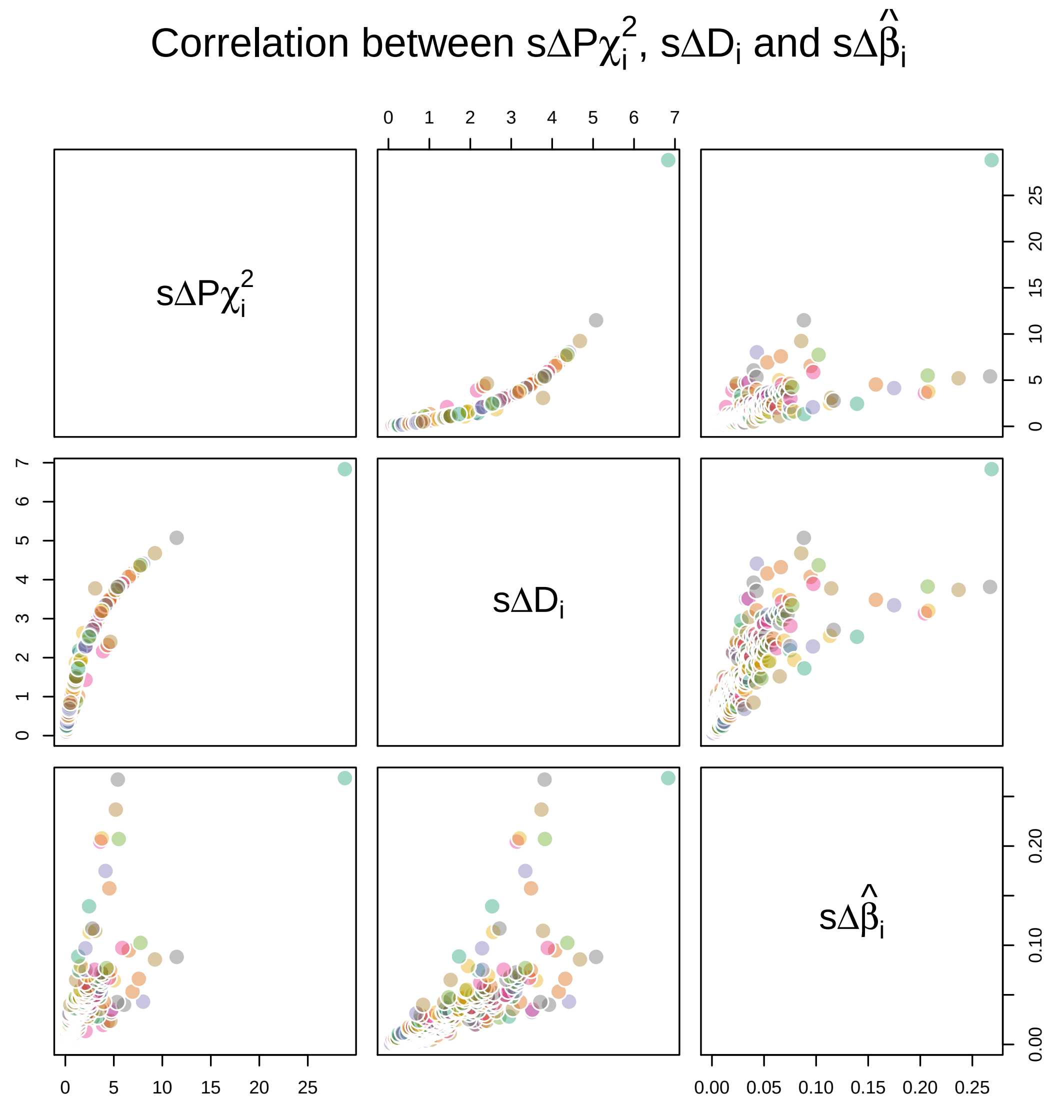


Probability $P_i \times$ scaled change in Pearson chi−sq $s\Delta P$
area $\propto s\Delta\hat{\beta}_i$, radius $= \sqrt{\frac{s\Delta\hat{\beta}_i}{P_i}}$
$s\Delta P\chi^2_i$
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
probability P_i


Leverage $h_i \times$ scaled change in Pearson chi−sq sΔPχ
$h_i \approx x_i - \mu_x$, $s\Delta P\chi_i^2 = sPr_i^2$
$s\Delta P\chi_i^2$
0
5
10
15
20
25
0.01
0.02
0.03
0.04
0.05
0.06
0.07
leverage h_i


Leverage $h_i \times$ scaled change in deviance $s\Delta D_i$
$dr_i = \text{sign}(y_i - \hat{y}_i)\sqrt{d_i}$, $s\Delta D_i = \frac{dr_i}{\sqrt{1-h_i}}$
$s\Delta D_i$
0
1
2
3
4
5
6
7
0.01
0.02
0.03
0.04
0.05
0.06
0.07
leverage h_i


Leverage $h_i \times$ scaled change in coefficients $s\Delta\hat{\beta}_i$
$h_i \approx x_i - \bar{x}$, $s\Delta\hat{\beta}_i = \frac{sPr_i^2 h_i}{1-h_i}$
$s\Delta\hat{\beta}_i$
0.00
0.05
0.10
0.15
0.20
0.25
0.01
0.02
0.03
0.04
0.05
0.06
0.07
leverage h_i


Correlation between $s\Delta P\chi_i^2$, $s\Delta D_i$ and $s\Delta\hat{\beta}_i$
$s\Delta P\chi_i^2$
$s\Delta D_i$
$s\Delta\hat{\beta}_i$
0 1 2 3 4 5 6 7
0 5 10 15 20 25
0.00 0.05 0.10 0.15 0.20 0.25
0.00 0.10 0.20