

# Loess Graphic for Propensity Score Analysis with ggplot2

Jason Bryer

March 20, 2011

## Abstract

This paper describes how to recreate the `loess.psa` function in the `PSAgraphics` package (Helmreich & Pruzek, 2009) with `ggplot2`. The result is a loess-based graphic to estimate the effects of a treatment after propensity score adjustment. An example using the birth weight data from the `MASS` package is provided.

## 1 Load Data

```
> library(MASS)
> library(ggplot2)
> data(birthwt)
> attach(birthwt)
> race <- factor(race, labels = c("white", "black", "other"))
> ptd <- factor(pt1 > 0)
> ftv <- factor(ftv)
> levels(ftv)[-1:2] <- "2+"
> bwt <- data.frame(low = factor(low), age, lwt, race, smoke = (smoke >
+ 0), ptd, ht = (ht > 0), ui = (ui > 0), ftv, bwt = birthwt$bwt)
> detach("birthwt")

> str(bwt)

'data.frame':      189 obs. of  10 variables:
 $ low  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
 $ age  : int   19 33 20 21 18 21 22 17 29 26 ...
 $ lwt  : int   182 155 105 108 107 124 118 103 123 113 ...
 $ race : Factor w/ 3 levels "white","black",...: 2 3 1 1 1 3 1 3 1 1 ...
 $ smoke: logi  FALSE FALSE TRUE TRUE TRUE FALSE ...
 $ ptd  : Factor w/ 2 levels "FALSE","TRUE": 1 1 1 1 1 1 1 1 1 1 ...
 $ ht   : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ ui   : logi  TRUE FALSE FALSE TRUE TRUE FALSE ...
 $ ftv  : Factor w/ 3 levels "0","1","2+": 1 3 2 3 1 1 2 2 2 1 ...
 $ bwt  : int   2523 2551 2557 2594 2600 2622 2637 2637 2663 2665 ...
```

```

> names(bwt)
[1] "low"  "age"  "lwt"  "race" "smoke" "ptd"  "ht"   "ui"   "ftv"
[10] "bwt"
> nrow(bwt)
[1] 189

```

## 2 Estimate the Propensity Scores

```

> lr = glm(smoke ~ age + race + lwt + ftv, data = bwt, family = "binomial")
> summary(lr)

```

```

Call:
glm(formula = smoke ~ age + race + lwt + ftv, family = "binomial",
    data = bwt)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6429  -0.8467  -0.5909   0.9667   2.5093

```

```

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.326224  1.049813   2.216  0.02670 *
age          -0.014156  0.033474  -0.423  0.67236
raceblack    -0.696776  0.491252  -1.418  0.15608
raceother    -2.149384  0.423253  -5.078 3.81e-07 ***
lwt          -0.009677  0.005913  -1.636  0.10175
ftv1         -1.359942  0.446667  -3.045  0.00233 **
ftv2+        -0.465798  0.424397  -1.098  0.27240
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

(Dispersion parameter for binomial family taken to be 1)

```

```

Null deviance: 253.04  on 188  degrees of freedom
Residual deviance: 215.85  on 182  degrees of freedom
AIC: 229.85

```

```

Number of Fisher Scoring iterations: 4

```

## 3 Data Preparation

```

> birthwt.psa = data.frame(treatment = bwt$smoke, response = bwt$bwt,
+   ps = fitted(lr))

```

## 4 The Graphic

```
> p = ggplot(birthwt.psa, aes(x = ps, y = response, colour = treatment)) +  
+   geom_point() + geom_smooth(method = "loess", size = 0.7,  
+   span = 0.8, alpha = 0.3) + xlab("Propensity Score") + ylab("Birth Weight") +  
+   scale_colour_hue("Smoked")  
> print(p)
```

