

# Evaluating Model Fit of Logistic Regression for Propensity Score Analysis

Jason Bryer  
Excelsior College

---

## Abstract

Logistic regression is a common method used for the estimation of propensity scores in observational studies. In many observational studies the ratio of treatment to comparison group subjects can be quite large. We show that, in many instances, as the ratio increases the range of propensity scores (i.e. fitted values from a logistic regression model), shrinks. This has important implications for ensuring adequate overlap in propensity scores between treatment and comparison group subjects. Several common instances are explored using simulated data and then applied to the Programme of International Student Assessment (PISA) for comparing public and private school students.

*Keywords:* propensity score analysis, logistic regression, graphics, R.

---

## 1. Introduction

```
> getSimulatedData <- function(nvars=3,
+   ntreat=100, treat.mean=.6, treat.sd=.5,
+   ncontrol=1000, control.mean=.4, control.sd=.5) {
+   if(length(treat.mean) == 1) { treat.mean = rep(treat.mean, nvars) }
+   if(length(treat.sd) == 1) { treat.sd = rep(treat.sd, nvars) }
+   if(length(control.mean) == 1) { control.mean = rep(control.mean, nvars) }
+   if(length(control.sd) == 1) { control.sd = rep(control.sd, nvars) }
+   df <- c(rep(0, ncontrol), rep(1, ntreat))
+   for(i in 1:nvars) {
+     df <- cbind(df,
+                 c(rnorm(ncontrol, mean=control.mean[1], sd=control.sd[1]),
+                   rnorm(ntreat, mean=treat.mean[1], sd=treat.sd[1])))
+   }
+   df <- as.data.frame(df)
+   names(df) <- c('treat', letters[1:nvars])
+   return(df)
+ }
```