

# Outlier Analysis

Outlier detection is an extremely useful tool. There are many ways to identify an outlier. This example will discuss one univariate approach and one multivariate approach. There are many uses for outlier detection. One use can be to inspect a dataset prior to analysis to ensure accurate analysis. It can also be used to validate data during data entry to help prevent data entry errors. If a researcher has a simple univariate dataset then something like the Grubbs test for outliers would work fine. The approach taken here to identify outliers is an approach known as Local Outlier Factor (LOF). The R package is known as *lofactor* and it replaces the *dprep* package. The *lofactor* can help identify multivariate outliers.

The below dataset creates an artificial outlier and can be seen in the multivariate k-means clustering. With the LOF, the density of a point is compared to each of its neighbors. This example uses two packages: the *DMwR* for the *LOF* function and the *outlier* package for the grubbs test for outliers.

```
library(DMwR);
library(outliers);

set.seed(1234)
gen.xyz <- function(n, mean, sd) {
  cbind(rnorm(n, mean[1], sd[1]),
        rnorm(n, mean[2], sd[2]),
        rnorm(n, mean[3], sd[3])
  );
}

xyz <- rbind(gen.xyz(150, c(0,0,0), c(.2,.2,.2)),
            gen.xyz(150, c(2.5,0,1), c(.4,.2,.6)),
            gen.xyz(150, c(1.25,.5, .1), c(.3,.2, .5)));

xyz[1,] <- c(0,2,1.5);
km.3 <- kmeans(xyz, 3);

outlier.scores <- lofactor(xyz, k=5)
plot(density(outlier.scores));

outliers <- order(outlier.scores, decreasing=T)[1:5]
print(outliers);

grubbs.test(xyz[,1], type = 10, opposite = FALSE, two.sided = FALSE)
grubbs.test(xyz[,2], type = 10, opposite = FALSE, two.sided = FALSE)
grubbs.test(xyz[,3], type = 10, opposite = FALSE, two.sided = FALSE)

pch <- rep(".", n)
pch[outliers] <- "+"
col <- rep("black", n)
col[outliers] <- "red"
pairs(xyz, pch=pch, col=col)

my.cols = km.3$cluster;
plot(xyz[,c(1,2)], col=my.cols);
plot(xyz[,c(1,3)], col=my.cols);
plot(xyz[,c(2,3)], col=my.cols);
```