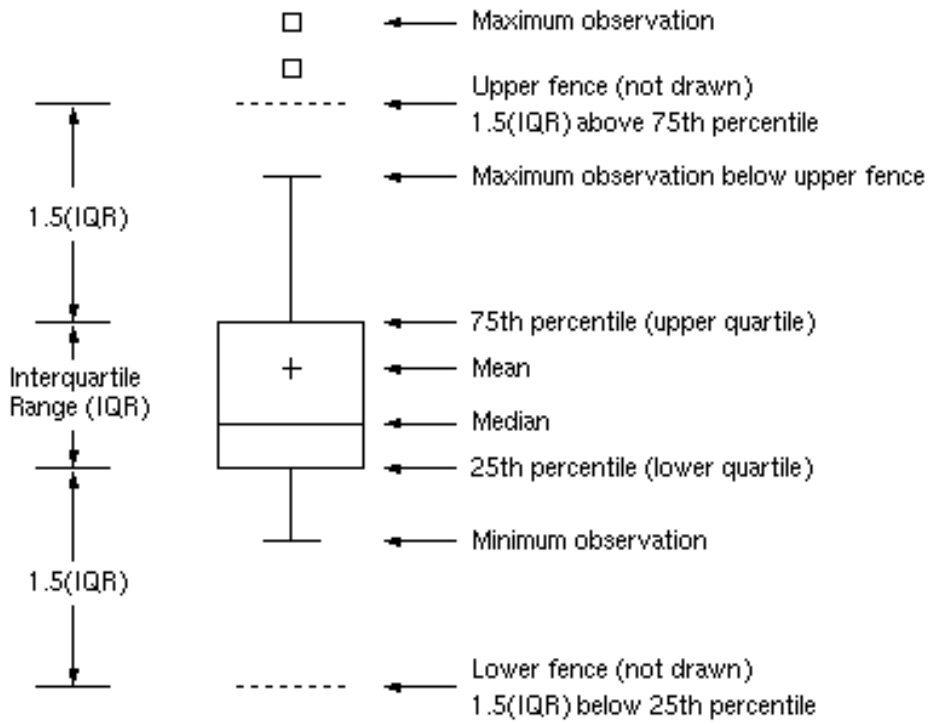
Negotiators expressed interest in using repayment rate as a metric for identifying programs that do not meet Department benchmarks. This report provides one option for identifying outliers[1,2]. The Department invites discussion and recommendations.

Open questions:

1. Should outliers be identified at the individual CIP level, among institutions of similar size and scope that serve a demographically matched student population, or across the universe?
2. How should thresholds be selected?
3. How frequently should thresholds be re-assessed?
4. How should the methods and thresholds be made available to the public?

Recommended Method: Box Plots[3,4]

Threshold set at first quartile reduced by 1.5 times the interquartile range (IQR). The IQR is the third quartile minus the first quartile (Threshold = Q1 – 1.5(Q3 – Q1)).

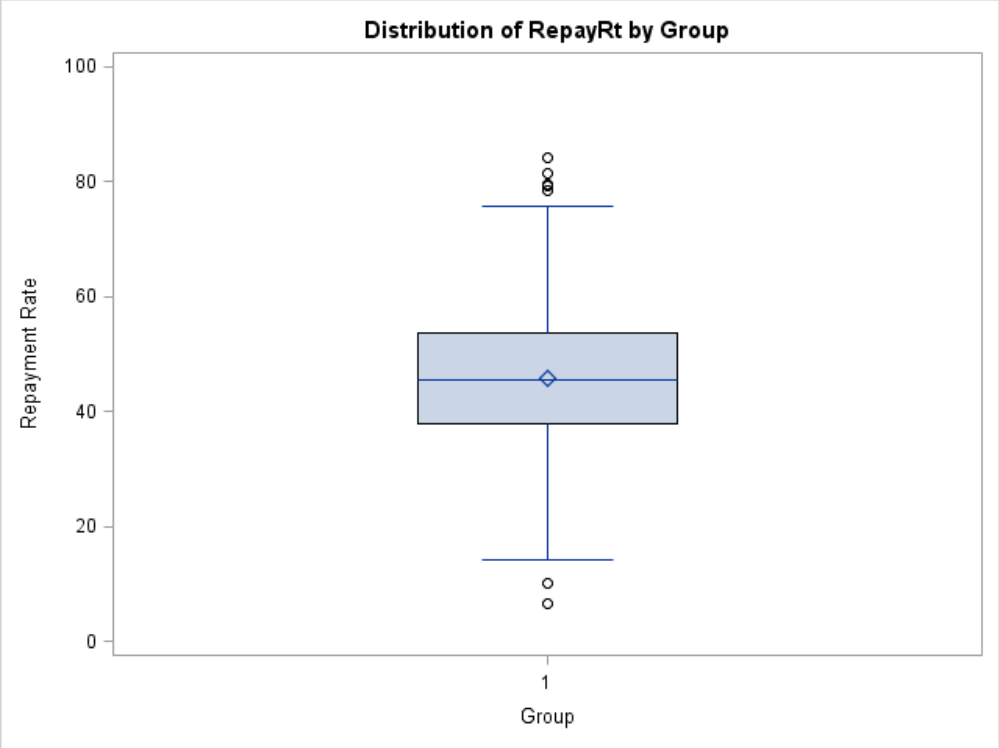


Source: The SAS System Documentation. SAS Institute Inc., Cary, NC, USA.

[1] Iglewicz, Boris and Hoaglin, David (1993), "Volume 16: How to Detect and Handle Outliers", *The ASQC Basic References in Quality Control: Statistical Techniques*, ASQC Quality Press. Milwaukee, WI.
[2] Barnett and Lewis (1994), *Outliers in Statistical Data*, 3rd. Ed., John Wiley and Sons.
[3] Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey, (1983), *Graphical Methods for Data Analysis*, Wadsworth.
[4] Dawson, Robert. (2011). "Volume 19, Number 2: How Significant is a Boxplot?" *Journal of Statistics Education*. Accessed from: http://ww2.amstat.org/publications/jse/v19n2/dawson.pdf

The benchmark for repayment rate would be set at the lower fence (Q2 – 1.5*IQR). Therefore, programs or institutions with repayment rates less than the lower fence would not meet benchmarks. If the lower fence is less than zero, the threshold would be set at 0 percent.

The example below uses randomly generated values that are normally distributed. We assumed a mean of 46 and a standard deviation of 12. To mimic repayment rate, the data were created to range from 0% to 100%. Note that there are two elements identified as "outliers" for repayment rate. That is, they fall below the lower fence.



Source: generated using SAS software, Version 9.4. Copyright 2012. SAS Institute Inc.

Other methods considered:

- Percentiles—Choosing a specific percentile as a threshold assumes that outliers exist and that elements beyond that threshold are outliers.
- Statistical test[5,6]—Most tests for outliers require selection of number of outliers to identify or an upper limit on number of outliers to identify. Tests would be run annually on the data, no thresholds would be made available in advance, and programs identified as outliers would be notified after statistical testing is complete.

[5] Tietjen and Moore (August 1972), Some Grubbs-Type Statistics for the Detection of Outliers, *Technometrics*, 14(3), pp. 583-597.
[6] Rosner, Bernard (May 1983), Percentage Points for a Generalized ESD Many-Outlier Procedure, *Technometrics*, 25(2), pp. 165-172.