

Analysis 5 – LOC effect on APFD.

1 factor (Lines of Code metric)

As the variable “LOC” is continuous, we establish categories to divide them. We base our categories on Do and Rothermel, 2006. They divide programs in small, medium and large according to LOC. They consider small programs those with LOC smaller than 10k, medium as larger than 10k and smaller than 100k and large as larger than 100k.

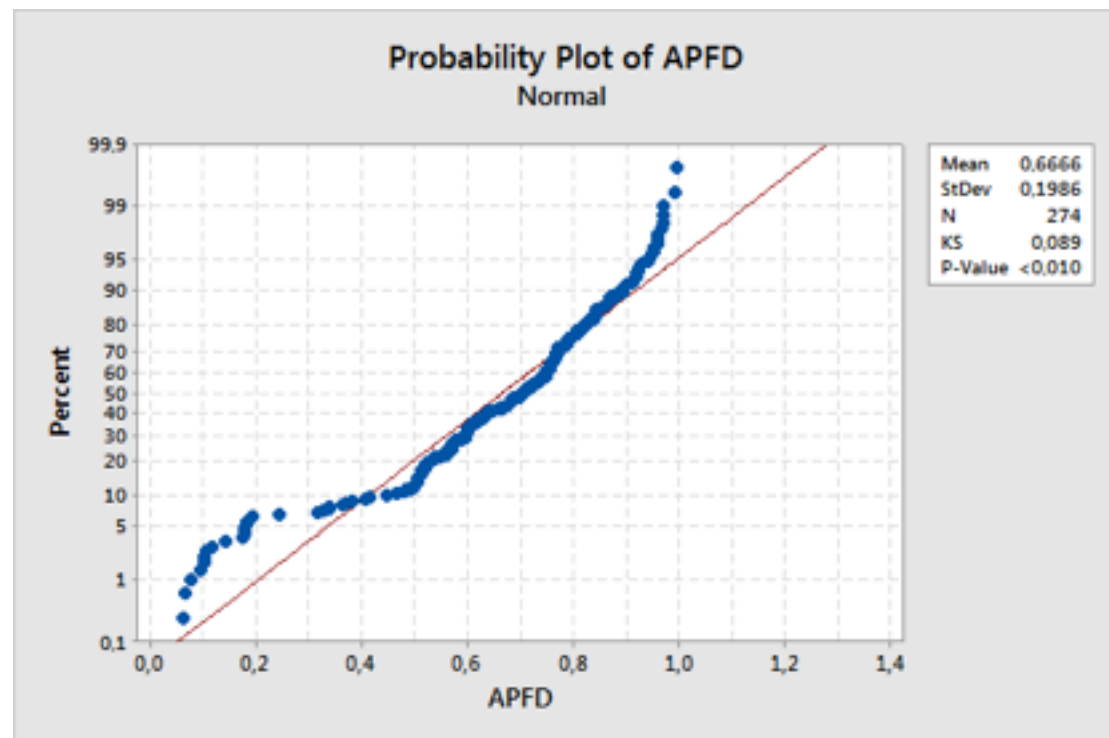
H0 – The means of TCP techniques execution results obtained using small, medium and large programs are equal.

H1 – The means of TCP techniques execution results obtained using small, medium and large programs are significantly different.

Data is available [here](#).

Normality test:

As the sample has 274 values, Kolmogorov-Smirnov test is used.



Given that the p-value is  $< 0.010$ , which is less than the established level of significance 0.05, the sample has a non normal distribution.

As the distribution is not normal, a non-parametric hypothesis test is used. In this case, we use Kruskal-Wallis test.

Kruskal-Wallis Test on APFD

LOC_CATEGORY	N	Median	Ave Rank	Z
large	2	0,6605	126,3	-0,20
medium	114	0,6725	136,4	-0,20
small	158	0,7200	138,5	0,23

Overall                      274                      137,5

H = 0,09    DF = 2    P = 0,958  
H = 0,09    DF = 2    P = 0,958    (adjusted for ties)

\* NOTE \* One or more small samples

A p-value of 0,958, which is bigger than the established significance level of 0.05, indicates that we don't have enough evidence to reject the null hypothesis. Thus, we can't say that LOC amount has a significant effect on APFD results, based on the analyzed data.

Regression analysis for continuous values of LOC on APFD results.

