

The graph displays the performance of three methods (GMM, EM, GEM) across different cluster sizes (5 to 10). The y-axis represents a metric ranging from 0.0 to 1.0. For each method, three lines are plotted: a solid line (GMM), a dashed line (EM), and a dotted line (GEM). GMM (solid lines) shows a slight increase in the metric as clustersize increases. EM (dashed lines) shows a decrease in the metric as clustersize increases. GEM (dotted lines) shows a slight increase in the metric as clustersize increases, but remains the lowest overall.

Clustersize	GMM (solid)	EM (dashed)	GEM (dotted)
5	0.25	0.85	0.10
6	0.28	0.78	0.12
7	0.30	0.72	0.13
8	0.32	0.65	0.14
9	0.34	0.58	0.15
10	0.36	0.52	0.16

The graph displays the Type-I error rate for three statistical procedures (ANOVA, MCTP, and Wald) across different cluster sizes (5 to 10). The y-axis represents the Type-I error, ranging from 0.05 to 0.10. The x-axis represents the Clustersize, ranging from 5 to 10. The legend indicates that solid lines represent ANOVA, dashed lines represent MCTP, and dash-dot lines represent Wald.

Clustersize	ANOVA (Solid)	MCTP (Dashed)	Wald (Dash-dot)
5	0.050	0.060	0.060
6	0.055	0.065	0.065
7	0.060	0.070	0.070
8	0.065	0.080	0.080
9	0.070	0.090	0.090
10	0.075	0.100	0.100

The graph illustrates the relationship between the number of clusters and the Clustersize parameter for three clustering methods: K-means, PAM, and CLARA. The x-axis represents Clustersize (ranging from 5 to 10), and the y-axis represents the number of clusters (ranging from 0 to 10). Solid lines indicate the mean number of clusters, while dashed lines represent the standard deviation.

- K-means (Green):** The mean number of clusters increases sharply as Clustersize increases, starting around 2.5 at Clustersize 5 and reaching approximately 8.5 at Clustersize 10. The standard deviation is also high, with the dashed line reaching nearly 10 at Clustersize 10.
- PAM (Blue):** The mean number of clusters remains relatively stable, around 2.5, across all Clustersize values. The standard deviation is also low, with the dashed line staying around 2.5.
- CLARA (Red):** The mean number of clusters increases gradually as Clustersize increases, starting around 1.5 at Clustersize 5 and reaching approximately 3.5 at Clustersize 10. The standard deviation is also low, with the dashed line staying around 3.5.

Overall, K-means shows the most significant increase in the number of clusters as Clustersize increases, while PAM and CLARA show more stable and gradual increases, respectively.

Figure 1: Type-I error vs Clustersize for sample sizes 12, 15, and 20. The graph shows that for sample size 12, the Type-I error increases with clustersize, while for sample size 15, it decreases. For sample size 20, the error remains relatively constant. Dashed lines represent the nominal significance levels (0.07, 0.15, and 0.19).

Clustersize	Sample Size 12 (Solid)	Sample Size 12 (Dashed)	Sample Size 15 (Solid)	Sample Size 15 (Dashed)	Sample Size 20 (Solid)	Sample Size 20 (Dashed)
5	0.070	0.190	0.130	0.210	0.060	0.090
6	0.078	0.190	0.112	0.150	0.060	0.092
7	0.085	0.190	0.095	0.150	0.060	0.094
8	0.095	0.190	0.078	0.150	0.060	0.096
9	0.102	0.190	0.058	0.150	0.060	0.098
10	0.110	0.190	0.040	0.150	0.060	0.100