한국항해항만학회지 제38권 제6호
J. Navig. Port Res. Vol. 38, No. 6 : 689-694, December 2014 (ISSN:1598-5725(Print)/ISSN:2093-8470(Online))
DOI : http://dx.doi.org/10.5394/KINPR.2014.38.6.689

A Study on Minimum Number of Ship-handling Simulation Required for Evaluating Vessel's Proximity Measure

† *Tae-Gweon JEONG · Bao-Feng PAN**

† *Professor, Division of Navigation Science, Korea Maritime and Ocean University, Busan 606-791, Korea*
* *Master Course, Graduate School, Korea Maritime and Ocean University, Busan 606-791, Korea*

Abstract : *The Korean government has introduced and enforced maritime traffic safety assessment to secure traffic safety since 2010. The maritime traffic safety assessment is needed by law to design a new port or modify an existing one. According to Korea Maritime Safety Act, in the assessment the propriety of marine traffic system consists of the safety of channel transit and berthing/unberthing maneuver, safety of mooring, and safety of marine traffic flow. The safety of channel transit and berthing/unberthing maneuver can be evaluated only by ship-handling simulation. The ship-handling simulation is carried out by sea pilots working with the port concerned. The vessel's proximity measure is an important factor to evaluate traffic safety. The proximity measure is composed of vessel's closest distance to channel boundary and probability of grounding/collision. What is more, the probability of grounding becomes important. According to central limit theorem, a sample has a normal distribution on condition that its size is more than 30. However, more than 30 simulation runs bring about the increase of assessment period and difficulty of employing sea pilots. Therefore this paper is to find out minimum sample size for evaluating vessel's proximity. First sample sets of size of 3, 5, 7, 9 etc. are selected randomly on the basis of normal distribution. And then KS test for goodness of fit and t-test for confidence interval are applied to each sample set. Finally this paper decides the minimum sample size. As a result this paper suggests the minimum sample size of 5, that is, the simulation of more than five times.*

Key words : *minimum sample size, vessel's proximity measure, maritime traffic safety assessment, Korea maritime safety act, propriety of marine traffic system, ship-handling simulation*

1. Introduction

In 2010 the Korean government determined that the maritime traffic safety assessment was enforced in order to improve the safety of the sea transportation in the harbor and harbor approaches (MOF. 2013a). The maritime traffic safety assessment is needed by law to design a new port or modify an existing one. According to the Korea Maritime Safety Act, the assessment is composed of five items such as investigation of marine traffic environment, measurement of marine traffic, propriety of marine traffic system, safety measure of marine traffic and comprehensive evaluation (MOF, 2013b). The propriety of marine traffic system consists of three sub-items, that is, the safety of channel transit and berthing/unberthing maneuver, safety of mooring, and safety of marine traffic flow. The safety of channel transit and berthing/unberthing maneuver can be evaluated only by ship-handling simulation. The ship-handling simulation is carried out by sea pilots working with the port concerned. In the result of the ship-handling simulation the vessel's proximity measure is an important factor to evaluate traffic safety. The proximity measure is composed of vessel's closest distance to channel boundary and probability of grounding/collision. According to the act the probability of grounding should be less than 0.0001 or 10^{-4}. And also because the simulation run should be more than three times, the assessment may be carried on the basis of the number of three times.

According to central limit theorem, a sample has a normal distribution on condition that its size is more than 30 (Kim et al, 1999). In practice, more than 30 simulation runs bring about the increase of assessment period and difficulty of employing sea pilots. Jeong (2014) presented the outline of the minimum sample in ship-handling simulation, which was not fully based on the statistics.

Therefore this paper is to find out minimum sample size for evaluating vessel's proximity on the basis of statistics. At first sample sets of size of 3, 5, 7, 9 etc. are selected randomly on the basis of normal distribution. And then the *KS* test for goodness of fit and confidence interval of the t-test are applied to each sample set. Finally this paper decides minimum sample size.

† Corresponding Author: tgjeong@kmou.ac.kr 051)410-4246
* pbf9527@gmail.com 051)410-4856

Note) This paper was presented on the subject of "A Study on Minimum Sample size for Vessel's Proximity Measure in Ship-handling Simulation" in Asia Navigation Conference 2014 proceedings(Xiamen, China November 6th-8th, 2014, pp. 426-433).

2. The current status of local pilot districts in Korea

Table 1 shows the current situation of local pilot districts in Korea (KMPA, 2014). The number of pilot can indicate how small the sample size is. The smallest number of the pilots is 5 in the port of Donghae, while the largest number is 51 in the port of Busan.

Assuming that ship-handling simulation is carried out by sea pilots only, more than five runs per simulation scenario cannot be done in the port of Donghae.

Table 1 Current status of pilot districts in Korea

Pilot district	Number of Pilot
Gunsan	8
Daesan	17
Donghae	5
Masan	16
Mokpo	6
Busan	51
Yeosu	44
Ulsan	29
Incheon	42
Pyongtaek	20
Pohang	9

3. The determination of the minimum sample size of ship-handling simulation

3.1 The random sample set

At first in order to decide the minimum sample size or minimum simulation run, this paper generates the random sample set on the condition that its population is normally distributed. At the same time the collision or grounding probability is less than 10^{-4}. The following indicates an example of population, which is obtained from the latest maritime traffic safety assessment.

Average : $\mu = 79.35(m)$

Variance : $\sigma^2 = 20.29^2$

Collision probability : $0.000046 < 10^{-4}$

Using the above parameters the paper obtains the random sample sets with the size of 3, 4, 5, 6, 7, and so on. Each random sample set will be composed of 20 simulation sets. Each sample set is tested by KS test and t-test. And the confidence interval of each sample set is given (MathWorks, 2014).

Table 2, Table 3, Table 4 and Table 5 depict the samples, the result of inference(KS test), and the confidence interval of each sample set of the size of 3, 4, 5, and 6 respectively. According to the KS test, all of sample sets indicate that $h = 0$ at a confidence level of $\alpha = 0.05$. It means that the null hypothesis of the normal distribution cannot be rejected.

Therefore the paper uses the confidence interval of each sample set given by the t-test.

Table 2 Sample sets of three(3) runs

Run \ Set	1	2	3	4	5	6	7	8	9	10
1	98.206	89.715	71.326	96.394	59.907	90.71	69.815	46.456	62.482	74.314
2	72.272	29.413	56.123	66.896	98.689	58.289	59.874	76.157	92.282	90.001
3	84.817	66.892	34.055	103.13	93.623	75.732	70.32	75.344	58.074	83.676
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	52.881	-13.627	7.2806	40.932	31.705	34.602	52.036	23.96	24.72	63.059
	117.32	137.64	100.39	136.68	136.44	115.22	81.303	108.01	117.17	102.27
	(64.439)	(151.267)	(93.109)	(95.748)	(104.74)	(80.618)	(29.267)	(84.050)	(92.450)	(39.211)

Run \ Set	11	12	13	14	15	16	17	18	19	20
1	30.707	68.123	75.949	93.664	49.457	91.044	84.541	68.109	83.853	122.3
2	63.678	51.898	105.31	76.704	64.104	62.419	67.946	65.547	90.427	81.564
3	71.11	78.475	81.627	42.311	70.55	66.592	54.425	93.374	70.022	60.687
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	1.7446	32.888	48.939	5.8959	34.52	34.938	31.5	37.471	55.56	10.344
	108.59	99.443	126.32	135.89	88.221	111.77	106.44	113.88	107.31	166.02
	(106.85)	(66.555)	(77.381)	(129.99)	(53.701)	(76.832)	(74.940)	(76.409)	(51.750)	(155.68)

Table 3 Sample sets of four(4) runs

Run \ Set	1	2	3	4	5	6	7	8	9	10
1	52.494	69.817	94.369	34.327	76.017	106.79	60.465	122.6	73.09	53.229
2	81.352	46.318	72.793	76.416	103.16	74.343	80.895	72.739	82.315	71.011
3	71.099	93.196	106.94	80.947	52.397	85.194	99.349	74.825	80.116	124
4	51.92	118.79	85.255	90.757	129.25	61.643	98.135	23.871	66.209	60.5
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	41.165	32.554	66.881	30.959	37.248	51.556	55.704	9.3575	63.817	26.183
	87.268	131.51	112.8	110.26	143.17	112.43	113.72	137.66	87.048	128.19
	(46.103)	(98.956)	(45.919)	(79.301)	(105.92)	(60.874)	(58.016)	(128.30)	(23.231)	(102.01)

Run \ Set	11	12	13	14	15	16	17	18	19	20
1	79.404	82.603	79.278	84.805	86.809	65.257	87.945	68.03	104.84	103.8
2	55.862	78.385	82.123	90.886	46.434	84.831	69.186	85.129	94.388	88.395
3	102.98	109.36	110.58	77.35	65.266	74.982	36.243	99.023	81.833	90.151
4	95.307	91.073	89.39	59.319	100.82	84.98	100.07	57.388	72.07	133.99
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	50.272	68.519	67.833	56.316	36.768	62.531	29.099	48.116	65.485	70.522
	116.5	112.19	112.85	99.864	112.89	92.493	117.62	106.67	111.08	137.65
	(66.228)	(43.671)	(45.017)	(43.548)	(76.122)	(29.962)	(88.521)	(58.554)	(45.595)	(67.128)

Table 4 Sample sets of five(5) runs

Run \ Set	1	2	3	4	5	6	7	8	9	10
1	90.604	59.162	85.065	78.233	96.687	65.158	84.42	86.866	124.84	71
2	103.22	79.012	97.902	98.228	73.779	53.664	59.759	66.031	75.966	68.392
3	78.322	44.531	89.145	63.857	83.872	103.17	75.987	96.219	70.036	107.08
4	115.92	104.87	86.929	54.278	43.947	100.45	80.277	97.124	90.432	93.692
5	77.421	55.324	61.81	65.174	63.673	47.551	114.04	82.156	52.926	91.678
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	72.567	39.011	67.492	50.866	47.488	41.491	58.359	69.949	49.247	66.06
	113.63	98.147	100.85	93.041	97.295	106.51	107.43	101.41	116.44	106.67
	(41.063)	(59.136)	(33.358)	(42.175)	(49.807)	(65.019)	(49.071)	(31.461)	(67.193)	(40.610)

Run \ Set	11	12	13	14	15	16	17	18	19	20
1	98.013	80.117	69.507	97.421	103.06	91.042	99.395	80.079	86.384	57.833
2	97.599	113.1	66.494	66.457	86.248	90.957	64.906	84.531	63.798	102.78
3	80.014	96.211	101.71	71.772	58.761	82.326	52.277	93.495	69.338	85.828
4	50.622	100.61	64.802	89.939	81.24	60.567	99.726	70.601	109.88	124.06
5	45.245	115.75	123.99	93.815	48.18	105.18	91.697	110.88	66.164	73.914
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	42.994	83.346	52.511	66.665	48.192	65.629	54.641	68.973	55.091	57.056
	105.6	118.97	118.09	101.1	102.8	106.4	108.56	106.86	103.13	120.71
	(62.606)	(35.624)	(65.579)	(34.435)	(54.608)	(40.771)	(53.919)	(37.887)	(48.039)	(63.654)

Table 5 Sample sets of six(6) runs

Run \ Set	1	2	3	4	5	6	7	8	9	10
1	92.296	79.234	90.522	101.35	33.898	75.1	89.644	119.1	83.029	70.26
2	78.957	72.052	92.634	79.055	74.569	68.815	54.552	70.178	67.238	74.203
3	88.121	79.876	85.169	92.395	53.923	115.92	79.788	90.17	99.685	81.479
4	47.094	36.959	108.13	46.483	58.447	77.603	61.379	70.076	98.114	67.98
5	84.415	104.55	56.647	86.426	91.379	87.073	75.462	81.165	48.475	88.145
6	94.834	86.292	77.835	101.33	101.47	72.291	93.34	135.81	66.363	88.902
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	62.565	53.099	67.108	62.958	42.578	64.578	59.609	65.883	56.064	69.025
	99.341	99.887	103.2	106.06	95.318	101.02	91.78	122.95	98.238	87.964
	(36.776)	(46.788)	(36.092)	(43.102)	(52.740)	(36.442)	(32.171)	(57.067)	(42.174)	(18.939)

Run \ Set	11	12	13	14	15	16	17	18	19	20
1	62.254	59.193	115.93	77.578	86.047	59.583	68.085	113.73	58.78	54.152
2	86.977	95.214	97.532	98.465	50.941	59.476	76.821	61.836	89.677	87.435
3	56.588	73.745	79.948	77.575	68.761	60.654	96.332	88.263	105.01	86.236
4	96.087	77.876	64.182	103.55	99.74	96.052	53.227	54.051	76.152	63.451
5	75.011	73.837	81.055	69.242	56.264	79.483	63.96	92.695	61.117	94.649
6	93.959	92.97	59.108	88.228	76.241	86.941	100.06	82.051	56.667	57.729
h	0	0	0	0	0	0	0	0	0	0
Confidence Interval	61.068	64.682	60.779	71.765	53.76	56.906	56.961	59.362	54.067	55.605
	95.891	92.93	105.14	99.781	92.238	90.49	95.867	104.85	95.067	92.279
	(34.823)	(28.248)	(44.361)	(28.016)	(38.478)	(33.584)	(38.906)	(45.488)	(41.000)	(36.674)

In these tables the parentheses in the column of confidence interval describe the range of the interval. In Table 2 of three simulation runs, 14 sets of 20 have the range of confidence interval of more than 70. They are underlined and shaded. In Table 3 of four runs, 7 sets of 20 have the range of confidence interval of more than 70.

However, in Table 4 of five runs and Table 5 of six runs, no set of 20 is over the range of confidence interval of 70. In view of the result we can conclude that the larger the sample size, the smaller the range of confidence interval is.

3.2 Confidence interval of sample set

Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7 show the confidence intervals and box plots of the sample sets obtained from Table 2 to Table 5. The confidence interval of each sample set are given by the symbols of '×' above and below each box. The tops and bottoms of boxes are the 25th and 75th percentiles of sample sets. In the box plots the central marks of a symbol of '—' are the medians. And the whiskers of '┴' or '┬' extend to the most extreme data points.

As shown in Fig. 1, in the sample sets of 3 runs, 6 sets of 20 are out of the mean of population and 4 sets of 20 are also outside of the interval of 40 to 120. In Fig. 2 of the sample sets of 4 runs, 4 sets of 20 are also out of the mean and 5 sets of 20 outside of the interval of 40 to 120. Meanwhile in Fig. 3 of the sample sets of 5 runs, 1 set of 20 is out of the mean and 1 set of 20 outside of the interval of 40 to 120. In Fig. 4 of the sample sets of 6 runs no run is out of the mean and no run is also outside of the interval of 40 to 120. In the sample sets of more than 6 runs show the same result as shown in Fig. 4 of 6 runs.

Considering the above result this paper suggests the minimum simulation run of 5 times.

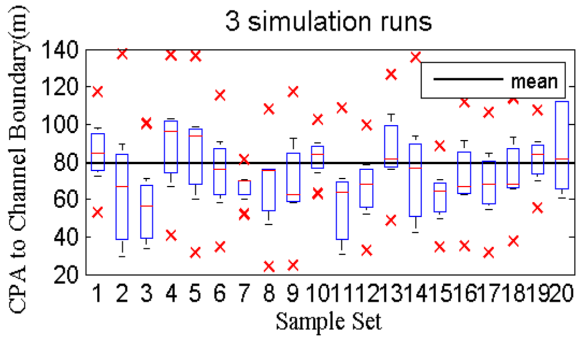

Fig. 1 Sample sets of 3 simulation runs

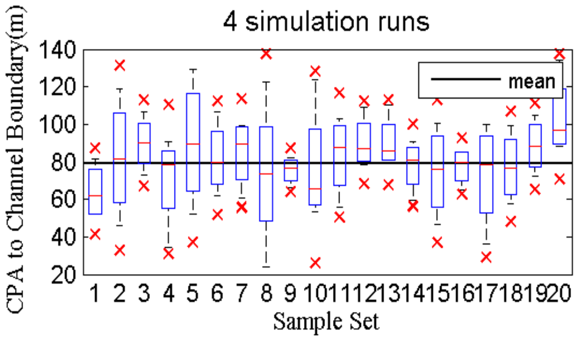

Fig. 2 Sample sets of 4 simulation runs

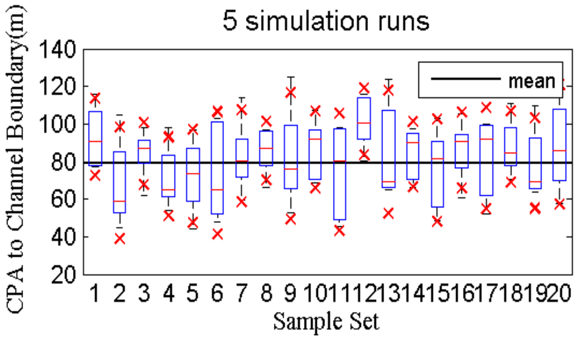

Fig. 3 Sample sets of 5 simulation runs

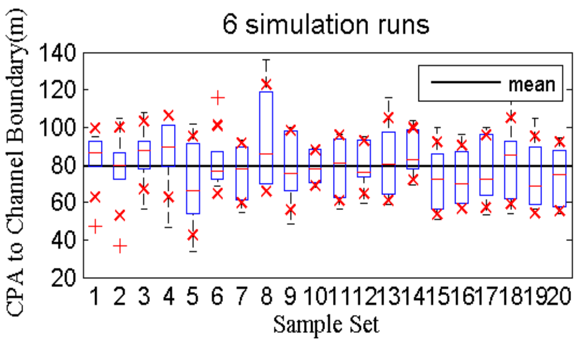

Fig. 4 Sample sets of 6 simulation runs

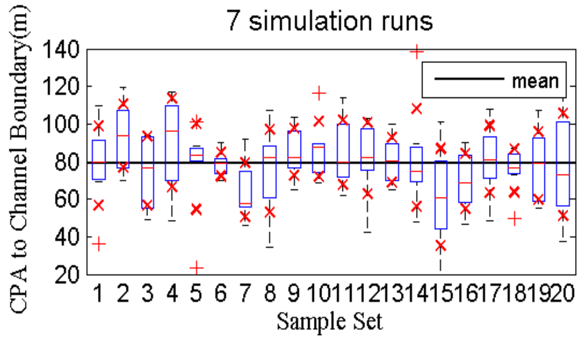

Fig. 5 Sample sets of 7 simulation runs

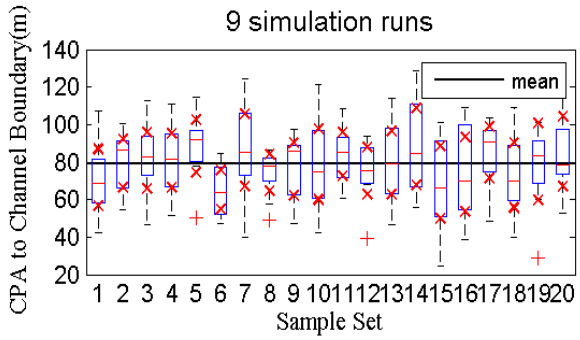

Fig. 6 Sample sets of 9 simulation runs

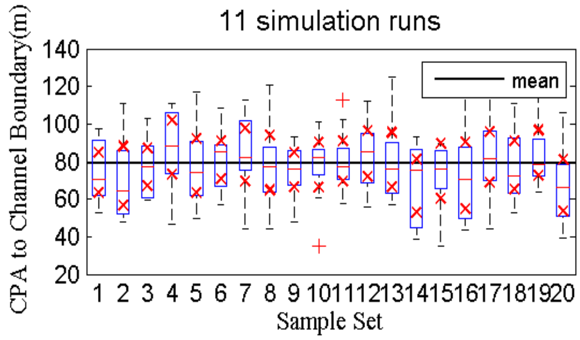

Fig. 7 Sample sets of 11 simulation runs

4. Conclusion

For the purpose of obtaining the minimum simulation run this paper generated the random sample sets, assuming that the population is distributed normally. And the paper carried out the KS test for goodness of fit, and t-test for confidence interval. As a result conclusions are the following.

(1) When the size of the sample or simulation run is more

than 3, the sample distribution follows the normal distribution under *KS* test.

(2) The confidence interval of less than 5 simulation runs is much larger than that of 5 simulation runs and more.

(3) In the box of 25th and 75th percentiles of less than 5 simulation runs, 4 sample sets or more of 20 are outside the mean of population.

In view of the above this paper suggests the minimum simulation runs of more than 5 times.

In the future the tests other than KS test will be applied to goodness of fit for the sample distribution.

References

[1] Jeong, T.G. (2014), 'A Study on Determination of Minimum Sample Size for Vessel Proximity Measure in Ship-handling simulation', Proceedings of Spring Conference of Korea Institute of Navigation and Port Research, pp. 41-42.

[2] Kim,Y.R., J.S. Lee & S.Y. Hwang (1999), Intelligible Statistics by using Excel, Myunkyung Publisher, p. 160

[3] KMPA (2014), Introduction of Pilots, http://kmpilot.or.kr/portal/

[4] MathWorks (2014), Statistics Toobox.

[5] MOF(2013a), Regulation 11.1 of Maritime Safety Act

[6] MOF(2013b), Guidelines of Maritime Traffic Safety assessment

Received 30 September 2014
Revised 27 December 2014
Accepted 27 December 2014