Methodology Report for Fleet Socio Economic Variables National Fisheries Data Collection Program

Version 3

Athens, 2020

AGRICULTURE ECONOMICS RESEARCH INSTITUTE (AGR.E.R.I)

Alkmanos str., 115 28, Ilisia, Athens / T.: +30 210 27 55 086 / E-mail: info@agreri.gr

Scientific coordination : Dr Irene Tzouramani

Research Associates: Dr Alexandra Sintori

Dr Angelos Liontakis

Dr Georgios Valakas

Stamatis Mantziaris, MSc

Ass. Prof. Petros E. Maravelakis

AGR.E.R.I partner: Magdalene Bardounioti

Contents

List of Tables

List of Abbreviations

Abbreviation	Term
CNR	Complete non-response errors
CV	Coefficient of variation
DCF	Data Collection Framework
DFN	Drift and/or fixed netters
DRB	Boat dredges
DRH	Hand dredges
DTS	Demersal trawlers and/or demersal seiners
FPO	Vessels using Pots and/or traps
GNC	Encircling gillnets
GTR	Trammel nets
HOK	Vessels using hooks
LHM	Handlines and pole-lines (mechanised)
LHP	Handlines and pole-lines (hand-operated)
LLD	Drifting longlines
LLS	Set longlines
LTL	Troll lines
MAR	Missing at random
MCAR	Missing completely at random
MGO	Vessels using other active gears
MGP	Vessels using Polyvalent 'active' gears only
MNAR	Missing not at random
PGO	Vessels using other Passive gears
PGP	Vessels using Polyvalent 'passive' gears only
PMP	Vessels using active and passive gears
PNR	Partial non-response errors
PS	Purse seiners
SB	Beach seines
TBB	Beam trawlers
TM	Pelagic trawlers
H.C.M.R.	Hellenic Centre for Marine
HAO-DEMETER	Hellenic Agricultural Organization – Demeter
I.M.B.R.I.W	Institute of Marine Biological Resources & Inland Waters
AGR.E.R.I	Agricultural Economics Research Institute
F.R.I	Fisheries Research Institute

Abstract

This report presents the methodological framework developed for the collection and analysis of socio-economic variables, in the context of the National Fisheries Data Collection Programme for the periods 2017-2019 and 2020-2021. It explains the data frame and describes the probability sampling survey for data collection. In addition, it presents the target and frame population as well as the main sources of information utilized. Also, the report explains the ways used to determine the sample size and how the sample is selected. The fishing gears, the vessel length, the importance of each fleet segment as well as the geographical location of the vessels are used to determine the sample size and the sample distribution. In the majority of cases where the statistical inference about the population cannot be derived by census data, it is based on an inference to the active population of the fishing vessels. Afterwards, the report describes the methods that ensure the quality of both the raw data and the statistical analysis, such as the unbiased indicator (coverage rate) and the indicators of variability (coefficient of variation and confidence intervals). Finally, it presents information on the availability, cohesion and comparability of data, along with the procedures applied to ensure confidentiality of data.

1. Type of Data Collection

1.1 Introduction

The present report describes the methodology adopted to collect socioeconomic data in the fisheries sector in accordance with Council Regulation 1004/2017 (EC) and Commission Implementing Decision (EU) 2019/909). Also, the report presents the methods used to ensure the data quality according to the European Statistical System (ESS) (Eurostat, 2014). The report includes the description of data accessibility, coherence and comparability as well as the principles of data confidentiality and security.

Hellenic Agricultural Organization – Demeter (**HAO-DEMETER**) and Hellenic Centre for Marine Research (**H.C.M.R.**) are the scientific partners of the Programme. More specifically, the Agricultural Economics Research Institute (AGR.E.R.I) and the Fisheries Research Institute (F.R.I) of the HAO-DEMETER organization participate in the Programme. AGR.E.R.I is responsible for the collection and evaluation of socioeconomic data in the fisheries sector. F.R.I is responsible for collecting scientific data on fisheries in the North and Central Aegean. The Institute of Marine Biological Resources & Inland Waters of the H.C.M.R. contributes also to the implementation of the National Programme and is responsible for collecting scientific data on fisheries in the South Aegean, the Ionian Sea and the Cretan Sea.

1.2 Data Frame

Even though the ideal survey method is the census, the special characteristics of the Greek fishing fleet such as the large number of vessels and the large geographical dispersion of ports, prevent the implementation of such a survey. Particularly, the Greek fleet consists of more than 14,000 vessels, 94% of which are smaller than 12 meters. In addition, the fishing fleet is distributed across more than 200 ports. Therefore, the majority of the economic and social variables of the fleet is collected by **Probability Sample Survey**, using face to face interviews and structured questionnaires. Moreover, other data sources are used as it is explained on section 3. The sampling frame presented below takes into account the classification of the national fleet according to vessel length and fishing gear, following the Commission Implementing Decision (EU) 2016/1251 (European Commission, 2016).

The chosen method of sampling is **the stratified random sampling** to take into consideration additional features of the Greek fleet, such as individual fishing gear (type) and geographical distribution. Tables A.1, A.2, and A.3 in the Annex present the collection method (census or probability sample survey) and the data sources for the activity, economic and social variables. Since the Greek management system does not involve individual quotas or other fishing rights, the following economic variables are not collected.

- *Income from leasing out quota or other fishing rights*
- *Value of quotas or other fishing rights*
- *Lease/rental payments for quota or other fishing rights*

The sample unit is the vessel and it is selected from the Greek vessel registry that corresponds to December 31th of the reference year. Moreover, the sample unit and, therefore, the sampling frame are common for the economic and effort variables following the Commission Decision 2010/93/EU (European Commission, 2010 - section A.1.1).

2. Target and Frame Population

The target population is the total number of registered vessels in the Greek fleet. The fishing gears are divided into active, passive and polyvalent gears (active and passive), as Table 2.1 presents. According to Data Collection Framework (DCF), the vessel length (in meters) in the Mediterranean Sea is categorized into the following groups: 0-6m, 6-12m, 12-18m, 18-24m, 24-40m and the respective group names used in this report are VL0006, VL0612, VL1218 and VL1824 and VL2440. Table 2.1 shows the total number of fishing vessels of the Greek fishing fleet by basic gear and length category, features that determine the segments of fishing fleet. It is worth mentioning that the values of Table 2.1 may diverge from the respective data of each year Annual Report. This may happen for the following reasons:

- Table 2.1 includes values of the national fleet register posted on 13/06/2018, while the corresponding table in each annual report uses the national fleet register that corresponds to the December 31th of the reference year.
- The characteristics of the population may be adjusted according to the sampling results when a different main gear from fleet register is declared in the field research.

Table 2.2 and Figure 2.1 show the fleet segments that are merged either for the presentation of results or for the sampling design. The clustering refers to the vessel length variable per main gear. In order to categorize the fishing vessels, the clustering is necessary for both sampling and confidentiality purposes[1] following the guidelines of STECF (2009). In most cases, clustering does not affect the aggregated segment, due to the small number of clustered vessels per segment. Table 2.3 presents the total number of vessels per fishing gear after clustering. The number of inactive vessels is not taken into account *a priori* as no information is available. Thus, the target population is the same with the frame population, as already mentioned.

[1] See COMMISSION DECISION of 18 December 2009 (adopting a multiannual Community Programme for the collection, management and use of data in the fisheries sector for the period 2011-2013).

		Length Category				
	Gear Classes	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
Active Gears	Beam trawlers (**TBB**)	0	0	0	0	0
	Demersal trawlers and/or demersal seiners (**DTS**)	1	109	28	98	146
	Pelagic trawlers (**TM**)	0	0	0	0	0
	Purse seiners (**PS**)	0	2	84	136	26
	Dredgers (**DRB/DRH**[2])	11	35	0	0	0
	Vessel using other active gears (**MGO**)	0	0	0	0	0
	Vessels using Polyvalent 'active' gears only (**MGP**)	0	0	0	0	0
Passive Gears	Vessels using hooks (**HOK**)	1551	2117	181	8	0
	Drift and/or fixed netters (**DFN**)	3719	6155	130	2	0
	Vessels using Pots and/or traps (**FPO**)	68	321	5	1	0
	Vessels using other Passive gears (**PGO**)	0	0	0	0	0
	Vessels using Polyvalent 'passive' gears only (**PGP**)	0	0	0	0	0
Vessels using active and passive gears (PMP)		0	0	0	0	0

Table 2.1: The segmentation of the Greek fleet according to the European Union Data Collection Framework (DCF) and the national fleet register data (using the main gear class) posted on 13/06/2018.

	Merged Class		Clustered Class		
Gear Class	**Length Category**	**Number of Vessels**	**Length Category**	**Number of Vessels**	**Total vessel number after clustering**
DFN	VL1824	2	VL1218	130	132
DRB	VL0006	11	VL0612	35	46
DTS	VL0006	1	VL0612	109	110
FPO	VL1218 VL1824	5 1	VL0612	321	327
HOK	VL1824	8	VL1218	181	189
PS	VL0612	2	VL1218	84	86

Table 2.2: The clustering scheme of Greek fleet segments* (based on the 13/06/2018 Fleet Registry)

*The first column indicates the main fishing gear for each cluster. The second column indicates the vessel length merged into the segment of the fourth column. The third column indicates the number of vessels merged, the fifth column includes the number of vessels per segment before the clustering and the last column presents the total number of vessels after the clustering per segment.

[2] There is no clear community decision on the classification of the hand dredgers (DRH) in the categories of main gears. Therefore, DRH fishing gear has been classified as boat dredgers (DRB) due to their common features.

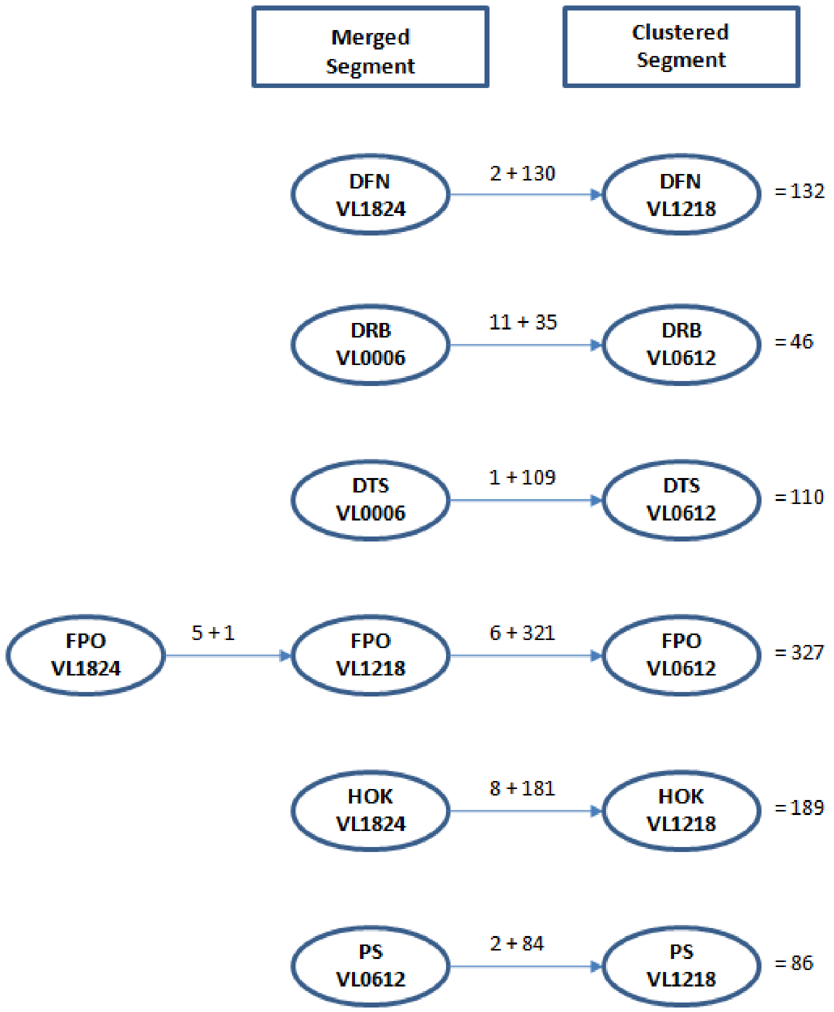

Figure 2.1: The process of clustering and the corresponding number of vessels per cluster.

Gear Class	VL0006	VL0612	VL1218	VL1824	VL2440	Total
DFN	3719	6115	132	0	0	**10006**
DRB	0	46	0	0	0	**46**
DTS	0	110	28	98	146	**382**
FPO	68	327	0	0	0	**395**
HOK	1551	2117	181	8	0	**3857**
PS	0	0	86	136	26	**248**

Table 2.3: The total number of vessels of Greek fleet after the clustering of segments (based on the 13/06/2018 Fleet Registry).

3. Data Sources

3.1 National fleet register

The Greek National Fleet Register (NFR) is held by the Hellenic Ministry of Maritime Affairs and Insular Policy and it is the primary source of information for the categorization of fleet segments per gear class and vessel length. The NFR is frequently updated at regular intervals and is available at http://ec.europa.eu/fisheries/fleet/index.cfm. The NFR includes information on the capacity and some other characteristics of each vessel, such as the port of registry, the vessel length, the year of construction, the capacity, the horsepower as well as the main and secondary fishing gear.

3.2 Questionnaires

For the sample probability survey of the no census variables, a structured questionnaire drawn up by AGR.E.R.I has been used from the beginning of the Programme and is continually updated to better serve the purposes of the survey. The questionnaire is filled out by face-to-face interviews with fishermen, while correspondents are fisheries scientists who receive the appropriate yearly training for the correct completion of the questionnaires. Moreover, correspondents receive written instructions and they are in contact with AGR.E.R.I research team for questions and clarifications. H.C.M.R and F.R.I are the responsible institutes for the collection of the questionnaires.

3.3 Other Sources

Data of electronic report system (ERS) from vessels and satellite-based Vessel Monitoring System (VMS) are additional sources of information. The Ministry of Rural Development and Food has an internal agreement with HAO-Demeter and H.C.M.R. to provide ERS data exclusively for the proper implementation of the Multi-annual Programme. These data refer to the logbook of all fishing vessels with length more than ten meters, the beach seiners (SB) with length less than 10 meters and the vessels which are larger than six meters and utilize the following fishing gears: handlines and pole-lines (mechanised) (LHM) and drifting longlines (LLD). Thus, these data can be utilized to estimate activity variables, such as: *days at sea*, *landing weight* and *value of landings*. The transmission of the above information excludes vessel identity for confidentiality reasons.

Finally, it should be noticed that the Ministry of Rural Development and Food provides HAO-Demeter with the list of vessels that utilize fishing gears that need a special fishing permit, i.e. SB, LHM LLD. This information is also used to categorize the Greek fleet by basic tool.

4. Sampling Frame

4.1 Sampling Strategy

A sampling scheme of stratified random sampling without replacement is chosen for the probability sample survey. The sample unit is the vessel and it is selected randomly from each segment as it is described below. The stratified random sampling of the Greek fleet is considered the most proper sampling technique due to the heterogeneous features of segments. It is noted that the stratified random sampling was also applied in the previous years of the National Fisheries Data Collection Program implementation.

4.2 Stratification

The stratification of Greek fleet is done by the variables: gear class and vessel length. Additional stratums (gear type level 4), are considered for vessels using drift and/or fixed netters (DFN): GNS, GTN and GTR). Also, two stratums are considered for the gear class Hook: Hook A, includes the gear types (level 4): LHP, LLS and LHM. Hook B includes the gear types (level 4): LHM and LLD. Table 4.1 shows the gear types (1st column) that have been merged to the strata GTR, HOK_A και HOK _B (4th column). The gear type GNC has been categorized into GTR strata due to statistical and confidential reasons.

Gear Type	Length Category	Number of Vessels	Sampling Stratum	Length Category
GNC	VL0006	3	**GTR**	VL0006
GNC	VL0612	1	**GTR**	VL0612
GTR	VL1824	2	**GTR**	VL1218
LHP	VL0006	117	**HOK_A**	VL0006
LHP	VL0612	71	**HOK_A**	VL0612
LHP	VL1218	3	**HOK_A**	VL1218
LLS	VL0006	1418	**HOK_A**	VL0006
LLS	VL0612	1890	**HOK_A**	VL0612
LLS	VL1218	45	**HOK_A**	VL1218
LTL	VL0006	15	**HOK_A**	VL0006
LTL	VL0612	19	**HOK_A**	VL0612
LHM	VL0006	1	**HOK_B**	VL0612
LHM	VL0612	47	**HOK_B**	VL0612
LHM	VL1218	43	**HOK_B**	VL1218
LLD	VL0612	90	**HOK_B**	VL0612
LLD	VL1218	90	**HOK_B**	VL1218
LLD	VL1824	8	**HOK_B**	VL1218

Table 4.1: List of gear types that have been merged to the strata GTR, HOK_A and HOK _B (based on the 13/06/2018 Fleet Registry)

Furthermore, the geographical distribution of vessels is taken into consideration for the construction of strata. As it is explained below, the features *gear class* and *vessel length* are accounted for the determination of sample size, while the registry port is used for the distribution of the sample over geographical areas. Following the sampling scheme of the biological data collection, the Hellenic coastline and marine area is divided in 12 major areas and stratified sampling will be carried out using these areas as -geographic- strata. These areas are the following: Argosaronikos (ARGSAR), Chios - Mytilene (CH-MIT), Central Ionion (C-ION), Crete (CRETE), Cyclades (CYCL), Dodecanese (DODEC), Evia (EVIA), North Ionion (N-ION), South Ionion (S-ION), Thermaikos Gulf (THERM), Thracian Sea-Limnos (THR-LIM) and Volos-Sporades (VOL-SPOR). F.R.I is responsible for the coordination and collection of questionnaires in the regions: Chios - Mytilene, Thermaikos Gulf, Thracian Sea-Limnos and Volos-Sporades. H.C.M.R. is responsible for the remaining areas. The specific fishing gear, the population size for each area and the total population for each stratum are presented in Table 4.2.

4.3 Sample Size

The sample size is determined by the specific gears and the length category as it is presented in the first column of Table 4.2. The registry port of each vessel is accounted for the distribution of sample over geographical areas. The variable *days at sea* was selected among the activity variables in order to determine the sample size in each segment (auxiliary variable) and the margin of error affecting the sample size was determined by the importance of each individual fishing gear (see Table 4.3). The level of statistical significance for all segments is set at 10% (z = 1.64). The sample size of each segment is estimated using the following equation (Dattalo, 2008; Moura, 2016):

$$\mathrm{n} = \frac{\mathrm{n}_0 \cdot \mathrm{N}}{\mathrm{n}_0 + (\mathrm{N} - 1)} \quad (4.1)$$

where N is the population size of each segment and

$$\mathrm{n}_0 = \frac{z^2 \cdot s^2}{\mathrm{e}^2 \cdot \bar{\mathrm{x}}} \quad (4.2)$$

where s is the standard deviation και $\bar{\mathrm{x}}$ is the average of the auxiliary variable (*days at sea*), based on the previous reference year data (see Annex Tables A.4 and A.5).

The sample size of each segment, as it was calculated by equation 4.2, is adjusted according to equation 4.3 when the population size of a segment is very small and the sample size is relatively large (n/N>0.05) (see e.g. Thomson, 2002).

$$\mathrm{n}_{\mathrm{adj}} = \frac{\mathrm{n}}{1 + \mathrm{n}/\mathrm{N}} \quad (4.3)$$

	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0612	27	2	1	0	1	0	0	1	0	4	9	1	**46**
DTS_0612	20	9	9	1	13	19	7	21	7	3	1	0	**110**
DTS_1218	5	0	0	3	9	3	1	0	0	0	3	0	**24**
DTS_1824	16	7	9	1	6	6	8	6	2	11	26	4	**102**
DTS_2440	29	12	5	7	4	4	13	2	2	39	25	4	**146**
FPO_0006	5	8	3	0	0	4	4	5	0	17	21	1	**68**
FPO_0612	11	12	2	0	0	16	13	7	0	166	93	7	**327**
GNS_0006	85	260	83	41	21	49	60	50	28	74	112	70	**933**
GNS_0612	178	241	143	46	23	40	52	123	31	149	133	78	**1237**
GNS_1218	10	0	3	2	3	1	0	1	1	2	2	0	**25**
GTN_0006	18	20	28	8	54	27	8	12	12	10	7	5	**209**
GTN_0612	72	23	50	8	100	30	25	55	33	20	38	20	**474**
GTN_1218	0	0	0	0	7	0	1	0	0	0	2	0	**10**
GTR_0006	231	258	318	118	85	155	244	371	150	262	206	179	**2577**
GTR_0612	583	303	471	276	255	322	388	626	247	402	403	168	**4444**
GTR_1218	13	9	5	7	27	15	5	2	1	1	10	2	**97**
HOK_A_0006	230	175	138	152	63	129	153	114	98	121	90	87	**1550**
HOK_A_0612	359	154	185	185	120	240	206	150	128	117	58	78	**1980**
HOK_A_1218	9	2	1	3	8	7	6	0	5	0	5	2	**48**
HOK_B_0612	31	8	17	15	4	20	3	13	0	6	9	12	**138**
HOK_B_1218	12	5	11	18	13	37	4	2	4	6	17	12	**141**
PS_1218	21	5	12	3	5	4	19	4	0	4	7	2	**86**
PS_1824	27	4	12	4	4	1	24	8	1	13	24	14	**136**
PS_2440	1	0	0	1	1	0	6	0	0	12	5	0	**26**
Total	**1993**	**1517**	**1506**	**899**	**826**	**1129**	**1250**	**1573**	**750**	**1439**	**1306**	**746**	14934

Table 4.2: The population of fishing vessels per stratum (based on the 13/06/2018 Fleet Registry)

Sampling Strata	Margin of Error
DTS_1824, DTS_2440, GNS_0612, GTR_0612, HOK_A_1218 PS_1218, PS_1824, PS_2440	5%
FPO_0006, FPO_0612, GNS_0006, GNS_1218, GTR_0006, GTR_1218, HOK_A_0006, HOK_A_0612, LLS_0006, LLS_0612	15%
DTS_0612, DTS_1218, GTN_0006, GTN_0612, GTN_1218, DRB_0612	20%

Table 4.3: The fleet strata with the respective margin of errors used to determine the sample size.

After the determination of the sample size of the fleet segments, the sample size per geographical stratum is determined using the method of proportional allocation (Eurostat, 2008):

$$n_g = \frac{n \cdot N_g}{N} \qquad (4.4)$$

where n is the sample size per fleet segment, as it was calculated by equation 4.3, N_g is the sum of vessels of the geographical area per fleet segment and N is the population size for each segment. Using the proportional allocation, the proportions are respected and, consequently, the share of a geographic stratum population to the total population will be similar to the share of this stratum to the sample size. Decimal values of sample size were rounded up to the nearest integer. The sample size per stratum is presented in Table 4.4.

In conclusion, the determination of sample size and the sample distribution was based on the specific fishing segments and their importance by taking into account the most representative auxiliary variable (days at sea) and the 12 major geographic areas. The total sample size for the reference year 2018 is set to 861 vessels, divided according to the main fishing gear and the length category, as shown in Table 4.5. Table 4.6 presents the sample ratio to the population, per main gear and vessel length.

	ARGSAR	CHI-MIT	C-ION	CRETE	CYCL	DODEC	EVIA	N-ION	S-ION	THERM	THR-LIM	VOL-SPOR	Total
DRB_0612	1	1	1	0	1	0	0	1	0	1	1	1	**8**
DTS_0612	3	2	2	1	2	3	1	3	1	1	1	0	**20**
DTS_1218	2	0	0	1	3	1	1	0	0	0	1	0	**9**
DTS_1824	7	3	4	1	3	3	4	3	1	5	10	2	**46**
DTS_2440	8	3	2	2	1	1	4	1	1	10	7	1	**41**
FPO_0006	2	2	1	0	0	1	1	2	0	4	5	1	**19**
FPO_0612	1	1	1	0	0	1	1	1	0	8	5	1	**20**
GNS_0006	2	5	2	1	1	1	2	1	1	2	2	2	**22**
GNS_0612	11	14	9	3	2	3	4	8	2	9	8	5	**78**
GNS_1218	5	0	2	1	2	1	0	1	1	1	1	0	**15**
GTN_0006	1	1	2	1	3	2	1	1	1	1	1	1	**16**
GTN_0612	1	1	1	1	2	1	1	1	1	1	1	1	**13**
GTN_1218	0	0	0	0	3	0	1	0	0	0	1	0	**5**
GTR_0006	2	2	3	1	1	2	2	3	1	2	2	2	**23**
GTR_0612	11	6	9	5	5	6	7	12	5	8	8	3	**85**
GTR_1218	4	3	2	2	8	5	2	1	1	1	3	1	**33**
HOK_A_0006	4	3	3	3	2	3	3	2	2	3	2	2	**32**
HOK_A_0612	23	10	12	12	8	15	13	10	8	8	4	5	**128**
HOK_A_1218	5	1	1	2	4	4	3	0	3	0	3	1	**27**
HOK_B_0612	10	3	6	5	2	7	1	5	0	2	3	4	**48**
HOK_B_1218	5	2	5	8	6	15	2	1	2	3	7	5	**61**
PS_1218	9	3	6	2	3	2	9	2	0	2	3	1	**42**
PS_1824	10	2	5	2	2	1	9	3	1	5	9	6	**55**
PS_2440	1	0	0	1	1	0	3	0	0	6	3	0	**15**
Total	**128**	**68**	**79**	**55**	**65**	**78**	**75**	**62**	**32**	**83**	**91**	**45**	**861**

Table 4.4: Sample size of fishing vessel per stratum (based on the 13/06/2018 Fleet Registry).

	Length category				
Fishing gear	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
DFN	61	176	53	0	0
DRB	0	8	0	0	0
DTS	0	20	9	45	41
FPO	19	20	0	0	0
HOK	32	176	88	0	0
PS	0	0	42	55	15

Table 4.5: The sample size of fleet segments. The sample size is zero when there are no vessels on a population segment after the clustering.

	Length category				
Fishing gear	**VL0006**	**VL0612**	**VL1218**	**VL1824**	**VL2440**
DFN	1,64%	2,86%	40,15%	-	-
DRB	-	17,39%	-	-	-
DTS	-	18,18%	37,50%	45,10%	28,08%
FPO	27,94%	6,12%	-	-	-
HOK	2,06%	8,31%	46,56%	-	-
PS	-	-	48,84%	40,44%	57,69%

Table 4.6: The coverage rate per main gear and vessel length.

4.4 Sample Selection

The sample selection, as already mentioned, is random, producing one random number for each fishing vessel of the Fleet registry. The random numbers are produced by computerized routine and receive values from 0 to 1. For each stratum of sampling, vessels are sorted from the lowest to the highest random number. Then, the number of vessels is selected according to the sample size in each stratum. The sample of each stratum includes the vessels with the smallest random values. In this way the selection of sample is achieved without replacement and each vessel has an equal probability to be selected.

A complementary sample is selected according to the above classification, in order to address non-response of the sampling unit. If the complementary sample is exhausted in the defined geographical areas, sampling shall be carried out in adjacent areas and in vessels with similar characteristics in terms of vessel length and main fishing gear.

4.5 Sample Evolution

The main differences of proposed sample size during the years occur due to the changes of the segments' significance. Table 4.7 shows the coverage rate of proposed sample in relation to the population from 2012 to 2018 for each fleet segment. Over the years, there are small differences in coverage rates between fleet segments. These small differences are due to the population size and not due to the sample size. Also, in some cases, as in the DTS segment, even greater significance is attributed to larger vessels. For example, DTS segment with length 18-24 m has 45.10% coverage rate compared to 28.57% of the year 2017. Moreover, the previous sampling schemes classified some vessels to the PGP segment, while the present does not. The main reason for this decision is that vessels previously classified to the PGP segment have high variability of the second gear type, while using a main gear type systematically.

Gear Segment	2012	2013	2014	2015	2016	2017	2018
DFN_VL0006	-	-	5.04%	5.02%	2.15%	2.11%	1.64%
DFN_VL0612	-	-	5.43%	5.01%	1.86%	1.86%	2.86%
DFN_VL1218	-	-	30.40%	30.30%	28.65%	28.78	40.15%
DRB_VL0612	-	-	-	-	-	-	17.39%
DTS_VL0612	31.75%	30.14%	-	-	-	28.43%	18.18%
DTS_VL1218	-	52.29%	-	-	-	51.43%	37.50%
DTS_VL1824	33.78%	34.17%	30.78%	30.77%	30.00%	28.57%	45.10%
DTS_VL2440	10.87%	21.56%	21.69%	23.03%	20.00%	20.00%	28.08%
FPO_VL0006	-	-	31.82%	30.07%	40.00%	38.36%	27.94%
FPO_VL0612	-	-	30.14%	30.23%	16.72%	16.47%	6.12%
FPO_VL1218	-	-	50.00%	53.84%	-	-	-
HOK_VL0006	-	-	15.35%	15.04%	6.28%	6.15%	2.06%
HOK_VL0612	-	-	5.16%	5.01%	9.14%	8.90%	8.31%
HOK_VL1218	35.07%	32.08%	33.05%	30.39%	43.56%	43.44%	46.56%
PGP_ V0006	8.43%	8.30%	-	-	37.50%	33.33%	-
PGP_ V0612	6.28%	6.00%	-	-	38.46%	35.71%	-
PGP_ V1218	34.93%	32.54%	-	-	-	-	-
PS_VL1218	36.60%	35.63%	30.69%	31.03%	36.14%	36.14%	48.84%
PS_VL1824	33.93%	33.58%	30.36%	30.60%	22.22%	22.22%	40.44%
PS_VL2440	50.32%	51.61%	51.61%	51.61%	50.00%	50.00%	57.69%

Table 4.7: Evolution of proposed sample coverage rate during the years.

5. Estimation Procedure and Modelling

5.1 Estimation

Statistical inference of population for the economic and social variables is based on the inference of the sample to the active population of fishing vessels, where applicable. No estimation is made for the variables that result from census. The Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) is used to allow correct generalization of the sample statistics to the population parameters per fleet segment:

$$\hat{Y} = \sum_{i=1}^{n} y_i {\pi_i}^{-1} \qquad (5.1)$$

where $\pi_i = \mathrm{n_{act}}/\mathrm{N_{act}}$ is the inclusion probability for each stratum, $\mathrm{n_{act}}$ the number of active vessels in sample per stratum and $\mathrm{N_{act}}$ is the number of active vessels in population per stratum. At a pilot stage, the population features will be estimated using linear regression analysis, to explore possible estimation models that can be used in the future.

Using the above formula, the number of inactive vessels of the population per stratum is also estimated, before the estimation of other variables[3]. The estimation is done for the stratums less than 12 meters in length, except for the stratums using the gears SB, LLD and LHM, where inactivity can be based on ERS information.

The economic variables: *Consumption of fixed capital* and *Value of physical capital* are estimated using data from both the probability sample survey (replacement value) and the NFR (mean LOA and number of vessels per fleet segment), as it is suggested by Perpetual Inventory Methodology (PIM) (European Commission, 2006). More specifically, the "degressive" depreciation function is used for the variable *Consumption of fixed capital* and the capital values are determined assuming that the engine is renovated every 10 years, electronics and other equipment every 5 and 7 years respectively, while the hull is never renovated. The share of each asset item in the total vessel price is:

- Hull – 60%
- Engine – 20%
- Electronics – 10%
- Other equipment – 10%

For the variable *Value of the physical capital*, the unit price is determined by direct survey. The selling prices of second-hand vessels and their insurance costs are also considered to evaluate the results of the survey.

5.2 Modelling

Statistical modelling is able to explain the variability of crucial socio-economic variables. Standard statistical techniques like linear regression provide a flexible tool to address such problems. Moreover, linear regression can be used as an alternative method to handle outliers and non-response errors. The use of a powerful model can give the researcher the ability to predict values of the dependent variable based on specific value(s) of the independent(s) variable(s). Apparently, a model that is able to fit the data neatly offers the researcher the ability to draw conclusions about different policies and future change scenarios (e.g. environmental risks). Under this perspective we present in the following sub paragraphs linear regression models to aid the researcher in explaining the inner relationships between variables using IBM SPSS Statistics 26.

[3] According to Commission Decision 2010/93/EU (European Commission, 2010), inactive are the vessels that have not been engaged in fishing operations during the reference year.

5.2.1. Full-Time Equivalents (FTEs)

The level of Full-time equivalents (FTE) is modelled using segment, length, days at sea and fishing time. The results are given in Table 5.1.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,912[a]	,831	,829	1,10416

a. Predictors: (Constant), DAYS_AT_SEA, OTB, DFN, LLD, FPO, , LLS, VL0006, VL0612, VL1218, VL1824, , FISHING_TIME

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4288,201	11	389,836	319,754	,000[b]
	Residual	870,493	714	1,219		
	Total	5158,694	725			

a. Dependent Variable: FTE

b. Predictors: (Constant), DAYS_AT_SEA, OTB, DFN, LLD, FPO, , LLS, VL0006, VL0612, VL1218, VL1824, , FISHING_TIME, ,

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8,410	,296		28,449	,000
	DAYS_AT_SEA	,004	,001	,094	4,250	,000
	FISHING_TIME	,0003494	,000	,131	5,561	,000
	VL0006	-3,783	,379	-,554	-9,975	,000
	VL0612	-3,749	,365	-,700	-10,283	,000
	VL1218	-2,898	,325	-,371	-8,918	,000
	VL1824	-1,398	,247	-,140	-5,657	,000
	DFN	-4,921	,260	-,923	-18,899	,000
	FPO	-5,139	,298	-,514	-17,249	,000
	LLD	-3,924	,297	-,283	-13,196	,000
	LLS	-4,901	,277	-,743	-17,688	,000
	OTB	-4,198	,242	-,395	-17,339	,000

a. Dependent Variable: FTE

Table 5.1: Linear Regression between Full-time equivalents and segment, length, days at sea, fishing time.

The equation is:

$$\widehat{FTE} = 8.41 + 0.004\, days\ at\ sea + 0.00035\, Fishing\ Time - 3.783\, VL0006 - 3.749\, VL0612 - 2.898\, VL1218 - 1.398\, VL1824 - 4.921\, DFN - 5.139\, FPO - 3.921\, LLD - 4.901\, LLS - 4.198\, OTB.$$

The interpretation of the coefficient of Days at Sea (0.004) is that for every unit increase in days at sea, FTE increases on average by 0.004 given that the remaining variables remain constant. Using the same way of thinking we are able to interpret the remaining coefficients of the model. This model is statistically significant since p-value is approaching zero (smaller than α=0.05, and also smaller than all the usually used significance levels). This model is able to explain R^2=83.1% of the total variability of FTE.

5.2.2 Fuel Consumption

Fuel Consumption is modelled using segment, capacity unit(s), days at sea and trip duration. The results are given in Table 5.2.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,934[a]	,873	,871	17512,22892

a. Predictors: (Constant), Trip_duration, Days_at_Sea, FPO, LLD, LLS, OTB, Capacity_Units, DFN

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1490825788297	8	186353223537	607,651	,000[b]
	Residual	217434816670	709	306678161		
	Total	1708260604968	717			

a. Dependent Variable: Fuel_Consumption

b. Predictors: (Constant), Trip_duration, Days_at_sea, FPO, LLD, LLS, OTB, DFN, Capacity_Units

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3968,542	3512,849		1,130	,259
	DFN	-16817,018	2753,702	-,172	-6,107	,000
	FPO	-15151,766	3560,297	-,082	-4,256	,000
	LLD	-5743,486	4041,399	-,022	-1,421	,156

	LLS	-14583,416	3003,954	-,121	-4,855	,000
	OTB	72119,082	3841,424	,373	18,774	,000
	Capacity_Units	855,851	42,000	,495	20,377	,000
	Days_at_sea	55,248	10,578	,072	5,223	,000
	Trip_duration	370,098	130,249	,042	2,841	,005

a. Dependent Variable: Fuel_Consumption

Table 5.2: Linear Regression between Fuel Consumption and segment, capacity unit(s), days at sea, trip duration.

The equation is equal to:

$$\widehat{Fuel\ Consumption} = 3968.54 + 55.248\ days\ at\ sea + 855.851\ Capacity\ Units + 370.098\ Trip\ duration - 16817.018\ DFN - 15151.77\ FPO - 5743.49\ LLD - 14583.42\ LLS + 72119.08\ OTB.$$

The interpretation of the coefficient of Days at Sea (55.248) is that for every unit increase in days at sea, Fuel Consumption increases on average by 55.248 given that the remaining variables remain constant. Using the same way of thinking we are able to interpret the remaining coefficients of the model. This model is statistically significant since p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). This model is able to explain R^2=87.3% of the total variability of Fuel Consumption.

5.2.3 Fuel Cost

Fuel Cost is modelled using the product Fuel consumption by segment and segment. The results are given in Table 5.3.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,968[a]	,938	,937	6220,72069

a. Predictors: (Constant), OTB, LLD, FPO, LLS, FuelConsSEG, DFN

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	417355602300,077	6	69559267050,013	1797,519	,000[b]
	Residual	27823406075,844	719	38697365,891		
	Total	445179008375,921	725			

a. Dependent Variable: fuel_cost

b. Predictors: (Constant), OTB, LLD, FPO, LLS, FuelConsSEG, DFN

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	115,653	999,055		,116	,908
	FuelConsSEG	,061	,001	,876	43,184	,000
	DFN	3384,408	1046,431	,068	3,234	,001
	FPO	3655,978	1289,744	,039	2,835	,005
	LLD	12370,474	1471,908	,096	8,404	,000
	LLS	2277,799	1106,983	,037	2,058	,040
	OTB	15309,868	1510,385	,155	10,136	,000

a. Dependent Variable: fuel_cost

Table 5.3: Linear Regression between Fuel Cost and the product of Fuel consumption by segment and segment

The equation is:

$$\widehat{Fuel\ Cost} = 115.65 + 0.061\ Fuel\ Cons * Segm + 3384.41\ DFN + 3655.98\ FPO + 12370.47\ LLD + 2277.8\ LLS + 15309.87\ OTB$$

The interpretation of the coefficient of Fuel Cons*Segm 0.061 is that for every unit increase in the product Fuel Consumption by Segment, Fuel Cost increases on average by 0.061 given that the remaining variables remain constant. Using the same way of thinking we are able to interpret the remaining coefficients of the model. This model is statistically significant since p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). The model is able to explain R^2=93.8% of the total variability of Fuel Cost.

5.2.4 Other Variable Costs

“Other variable costs” variable is modelled by Commercial costs, segment and length. The results are given in Table 5.4.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,949[a]	,901	,899	11097,32276

a. Predictors: (Constant), OTB, DFN, LLD, FPO, , LLS, Commercial_costs, VL0006, VL0612, VL1218, VL1824,

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	768304531317,844	10	76830453131,784	623,874	,000[b]
	Residual	84850744338,419	689	123150572,334		
	Total	853155275656,263	699			

a. Dependent Variable: Other_variable_cost

b. Predictors: (Constant), OTB, DFN, LLD, FPO, , LLS, Commercial_costs, VL0006, VL0612, VL1218, VL1824,

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	56353,557	3751,938		15,020	,000
	Commercial_costs	,680	,021	,656	31,948	,000
	VL0006	-45863,390	4290,879	-,518	-10,689	,000
	VL0612	-45192,375	4201,071	-,642	-10,757	,000
	VL1218	-38880,479	3824,352	-,367	-10,167	,000
	VL1824	-13778,127	2758,732	-,101	-4,994	,000
	DFN	-10595,346	3196,804	-,152	-3,314	,001
	FPO	-10272,541	3512,559	-,080	-2,925	,004
	LLD	10732,045	3518,427	,058	3,050	,002
	LLS	-8387,966	3328,816	-,098	-2,520	,012
	OTB	-30231,076	2820,707	-,206	-10,718	,000

a. Dependent Variable: Other_variable_cost

Table 5.4: Linear Regression between Other variable costs and Commercial costs, segment, length.

The equation is:

$$\widehat{Other\_variable\_cost} = 56353.56 + 0.68\,Commercial\_Costs - 45863.37VL0006 - 45192.38VL0612 - 38880.48VL1218 - 13778.13VL1824 - 10595.35\,DFN - 10272.54\,FPO + 10732.05\,LLD - 8387.97\,LLS - 30231.08\,OTB$$

The interpretation of the Commercial Costs' coefficient (0.68) is that for every unit increase in Commercial Costs, Other Variable Cost increases on average by 0.68 given that the remaining variables remain constant. Using the same way of thinking we are able to interpret the remaining coefficients of the model. This model is statistically significant since p-value is approaching zero (smaller than 0.05=α, and also smaller than all the usually used significance levels). The model is able to explain R^2=90.1% of the total variability of other variable cost.

6. Data Quality Evaluation

6.1 Methodology Relevance

The methodology adopted and described above is controlled for its proper implementation at all stages. In particular, during the sampling period, AGR.E.R.I research team communicates with the correspondents at regular time intervals in order to ascertain the proper process of collecting the questionnaires. When the questionnaires are collected, the material is evaluated; for example, the number of questionnaires collected per correspondent and the completeness of the data is checked.

The coverage rate of the proposed sample (Table 4.6) is compared with the response rate of sample units per fleet segment. Small discrepancies are allowed in the assessment of rates for reasons such as misclassification of the fishing fleet (see section 6.2) or non-updating of the NFR data during the sampling design.

Moreover, there is an evaluation of the sources of information that contributed to the categorization of the Greek fleet, such as NFR, in which each vessel is assigned to a main and a secondary fishing gear. The evaluation is carried out by calculating the percentage of fishing gears declared in the sampling process and differentiated from the initial classification of the fleet.

6.2 Results/Output Completeness

The annual report meets the requirements of Council Regulation (EC) No 1004/2017 and Commission Implementing Decision (EU) 2019/909. More specifically, the annual report includes all socio-economic variables and activity variables for all sections of the Greek fleet. As already mentioned, the following financial variables are not collected since the Greek management system does not include commercial quotas or other fishing rights:

- *Income from leasing out quota or other fishing rights*
- *Value of quotas or other fishing rights*
- *Lease/rental payments for quota or other fishing rights*

Data collection is compiled in accordance with the principles of Impartiality, Reliability and Objectivity using only official sources of information. The adopted methodology for collecting fisheries data follows international standards and best practices.

Before the statistical inference, the data collected either by census or by probability sampling survey are evaluated using a process of error detection as it is described in the next sections. The presentation of the results for all socio-economic variables and activity variables is detailed for all segments of the fishing fleet.

6.3 Accuracy sampling Errors

6.3.1 Sampling Errors

The desired accuracy for each segment is determined according to the importance of the segments (Table 4.3). As already described in Section 4, this process requires, in general, prior knowledge of the population variance for each variable. Therefore, an auxiliary variable whose variance better reflects the variability of all variables has to be chosen. This variable is "*days at sea*".

The coefficient of variation is selected as measure of sampling error for each variable and it is given by equation:

$$\mathrm{CV} = \frac{\mathrm{SD}}{\overline{\mathrm{X}}} \tag{6.1}$$

where SD is the standard deviation and $\overline{\mathrm{X}}$ is the average of the variable. Also, the 100(1-α)% confidence interval of mean is estimated for each variable according to the equation (e.g. Särndal et al. 1992):

$$\overline{\mathrm{X}} \pm \mathrm{t}_{\mathrm{n}-1,1-\alpha/2} \frac{\mathrm{S}^{*}}{\sqrt{\mathrm{n}}} \tag{6.2}$$

where n is the sample size of a fleet segment $\mathrm{t}^{2}_{\mathrm{n}-1,1-\alpha/2}$ the percentile of t-student distribution with n degrees of freedom and S^{*} the unbiased estimator of standard deviation of the sample, which is determined by the following equation:

$$\mathrm{S}^{*} = \sqrt{\frac{\sum_{\mathrm{i}}^{\mathrm{n}} (\mathrm{X}_{\mathrm{i}} - \overline{\mathrm{X}})^{2}}{\mathrm{n}-1}} \tag{6.3}$$

As the assumption that the data $X_1, X_2, \ldots X_n$ follow the normal probability distribution does not usually hold, we also use the Wilcoxon-signed rank confidence interval. Let us consider the $N = n(n+1)/2$ possible means of the form $(X_i + X_j)/2$ for all $i \leq j = 1, \ldots, n$. We have to find those means that are larger than a value m (equal to zero unless differently stated). Let $w_{a/2}$ denote the critical values of the statistical function T^{+} of the Wilcoxon signed rank test, where T^{+} is equal to the sum of the positive signed ranks. We find the k=$w_{a/2}$ larger and the k=$w_{a/2}$ smaller such means. The confidence interval with at least 100(1-α)% confidence is given by $[X^{(k)}, X^{(N-k+1)}]$ where $X^{(k)}$ is the *k*-th ordered observation (Hollander et al. (2014)).

6.3.2 Coverage Errors

The selection of sample units (vessels) is random and it is determined before the conduction of the probability sample survey, in order to avoid convenience sample. Also, there are no units of the target population that are excluded from the sample selection procedure. Therefore, the sample is considered to be representative and unbiased to the target population. As a consequence, the coverage error is zero. It is noted that Table 4.6 shows the coverage rate which is the percentage of sample-to-population for each segment of the fishing fleet rather than the percentage of units that cannot be included in the sample.

6.3.3 Measurement Errors

Before the data analysis, data are checked for their completeness, cohesiveness and comparability over time. The data are collected either by census or by the probability survey sampling, and then they are processed using exploratory data analysis, in order to locate measurement and processing errors. Measurement and processing errors are located through unreasonable and extreme values of the data. An unreasonable value is a value that has no natural meaning of interpretation of the variable (e.g. a negative value of a variable that can only take positive values). An extreme value is a value that is considerably remote, compared to the majority of the rest of the variable values.

In order to address the problems related to the unreasonable or extreme values that appear on some vessels in basic technical and economic parameters that are used for the estimation of the data call variables, we attempt to construct benchmark tables with Control Ranges per segment for each parameter. Then, to cope with the unreasonable or extreme value problem, the values that are outside the ranges of the benchmark tables are replaced by the median value of the parameter. For example, Gross value of landings per kilogram of live weight", "Landings per day" and "Days at sea" are auxiliary parameters for the calculation of the variable "Gross value of landings" at vessel level. The parameters "Days at sea" and "Value of landings" are used to "homogenize" the data and subtract that portion of the variance caused by the diversification of fishing effort. The parameters are presented in Table 6.1.

The data used to determine the ranges derive from observations per vessel for the period 2012-2016. The steps to determine the control ranges and to fix the parameter values are summarized as follows:

1. Construct a cumulative frequency distribution for the parameter under consideration per fleet segment.
2. Calculate descriptive statistics.
3. Set the limit equal to the 95% percentage point of the observations.
4. Incorporate an additional threshold up to the level of the two standard deviations (if applicable).
5. Use of physical boundaries (for example values that have to be larger than zero) and experts' knowledge.

6. Check the values per vessel (if many observations lie beyond the specified range then it is expanded by ± 20%).
7. Replace by the parameter median value of the fleet segment if required.
8. Transform each parameter properly, to have annual based values.

Parameters
Days at sea
Crew members
Energy consumption (Liters/day)
Landings or Production (kg/day)
Price (euros/kg)
Repair and maintenance costs (euros/day)
Other Variable costs (euros/day)
Energy cost (euros/day)
Personnel cost (euros/day)
Energy cost / Value of landings (%)
Other Variable costs / Value of landings (%)
Repair and maintenance costs / Value of landings (%)
Fixed costs/ Value of landings (%)
Personnel cost / Value of landings (%)
Total costs / Value of landings (%)

Table 6.1: Parameters utilized to build control ranges and fix prices

The Segments for which these parameters were examined and the corresponding constructed ranges were set to nine and are presented in Table 6.2.

SEGMENTATION	FISHING GEARS	LENGTH CLASS
Small -Scale fisheries	All passive gears	VL0006
	All passive gears	VL0612
Large-Scale fisheries	DFN	VL1218
	HOK	VL1218
	PS	VL1218
	PS	VL1824
	DTS	VL1824
	PS	VL2440
	DTS	VL2440

Table 6.2: Considered parts of the fleet for the construction of the ranges

Alternatively, another process can be used to locate extreme values. Specifically, a value may be characterized as extreme if it is located outside the following interval for each variable (Tukey, 1977):

$$(Q_1 - 1.5*IQR ,Q_3 + 1.5*IQR) \quad (6.3)$$

where Q1 and Q3 are the first and third quartile of the variable, respectively and IQR is the interquartile range.

When unreasonable or extreme values refer to data that have been collected by census, the source of information is checked and data are corrected in case of mismatch. In case of an accord, the unreasonable values or the extreme values are replaced with missing values. The same applies for data that have come from the probability survey sampling, after communicating with the correspondents in order to clarify the validity of these values. It should be noted that very extreme values are defined as the values that are outside the following limit (Tukey, 1977):

$$(Q_1 - 3*IQR ,Q_3 + 3* IQR) \quad (6.4)$$

Since equation 6.4 may yield minimum or maximum values that do not have a physical interpretation for the variable (e.g. the variable days at sea cannot be higher than the value of 365 days), the boundaries are adjusted using values from historical data, literature review (e.g. Pinello et al., 2017) and experts' opinion. In order to find possible omissions in the questionnaire's design or in the correspondent's training, the unreasonable values and the extreme values are segmented into:

- incorrect values due to the questionnaire's design,
- incorrect answers caused by the correspondent/interviewer and
- incorrect answers of the interviewee

The incorrect values due to questionnaire's design include the values that are repeated in different segments of the fishing fleet and are located in data coming from different correspondents. The incorrect answers caused by the correspondent consist of the values that are repeated in the data coming from the same correspondent and are related either to guided responses during the interview, incorrect coding or data entry. The incorrect answers of the interviewee contain the values that cannot be classified into the other two categories. It should be noted that the questionnaire does not include questions in free text format.

6.3.4 Non-response Errors

In order to address the non-response of the sampling unit (vessel), a complementary sample is selected. The complementary sample, as well as the main sample, is drawn randomly. The selection order of the complementary sample is determined in advance to avoid the collection of convenience sample. This action results in the complete non-response rate (CNR) being zero. Also, the partial non-response (PNR) rate of the collected variables is small due to the appropriate training of correspondents and the ability to resolve any questions in close cooperation with the AGR.E.R.I research team. During the interview, the correspondents pay special attention to the variables which had high non-response rates in previous years. Variables with probable high non-response rates have been taken into account before the designing of the questionnaire. Finally, the percentage of PNR is calculated for each variable.

The PNR cases and the missing values which replace the unreasonable/extreme values resulting from the exploratory data analysis, are classified into the following categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). A missing value is classified in MCAR if the probability of missing is independent from the variable or from other variables. A missing value is classified in MAR if the missing is random in the variable but it correlates with other variables. For example, the education level is often missing in vessels with high number of employers. Finally, as MNAR are classified the missing values that are correlated with the variable. For example, the values of annual revenues are missing for vessels with high revenues. The percentage of the three missing values categories is calculated for each variable per fleet segment.

Only the missing values classified in MCAR and MAR are replaced either by unconditional or conditional imputation. In the first case, the missing values of continuous variables are replaced directly by the median of the non-missing values whereas the missing values of nominal variables are replaced by the respective mode value of the variable in stratum. Apparently, if two or more values are missing for a specific variable, all the missing values will be substituted by the same number, a fact that leads to underestimation of the true variance due to the imputation of missing values at the center of the distribution.

Conditional mean imputation is an improvement on unconditional mean imputation. It is also known as regression imputation (RG). To illustrate the idea of the RG method, let us assume that we have (n × k) data set, where the columns are $x_1, \dots, x_k$. Consider a univariate missing pattern with $x_1, \dots, x_{k-1}$ fully observed and x_k observed for the first r values and missing for the last $n - r$ values. The RG method computes the regression of x_k on $x_1, \dots, x_{k-1}$ based on the r complete cases, and then fills in the missing values by the $n - r$ predicted values using the regression model. The models given in section 5.2 can be used to predict the missing values for each observation by substituting the values of the present variables in the regression equations.

Another method that can be used for handling missing values is the complete case analysis (CC). The CC is the most widely used method of handling missing values. This method considers only samples where all the variables are present, and discards those whose information is incomplete. It is also known as case deletion and listwise deletion in the literature (Little and Rubin, 2002). The main advantages of this approach are simplicity and comparability; simplicity since standard statistical analysis (for data without missing values) can be applied without modifications, and comparability of univariate statistics since they are calculated using the same pool of cases. Moreover, if the mechanism of missing values is MCAR, the remaining cases (reduced sample) will be a random subsample of the original sample. This implies that, for any parameter of interest, the estimates are unbiased for the full data set and that they are also unbiased for the listwise deleted data set. Disadvantages of the CC analysis are loss of information, which leads to loss of precision, and bias when the missing data mechanism is not MCAR. A straightforward result of the bias is that complete cases are not a random subsample of the original sample. It should be emphasized that complete case analysis is only used in extreme cases of questionnaires with large missing or problematic sections of the questionnaire.

7 Accessibility and Clarity

Data and metadata are presented and archived in a form that facilitates correct interpretation and comparisons. The end-users are informed about the methodology of statistical processes through the present methodological framework and through a short version presented in the Work Plan. They are also informed about the quality of the statistical results, following the standards of European Statistical System. The methodological framework and the Annual Report are available on the European Union website. The data are stored in databases and are not available for editing online.

The survey files including case information are created using a specific layout. The survey data can be unequivocally mapped to the internationally required fields, codes and formats if asked. Deviations from the prescribed record format must be reconciled before the data can be integrated for further processing.

All data collected are imported into a database, following specifications in the corresponding operational manuals and national record layouts. All data are verified for structural consistency within and across sources and for agreement with the defined formats and record layouts. It has to be assured that sample design and disposition data are recorded for every case. All adaptations are thoroughly tested prior to the production use of the data integration software.

Data are imported on a regular and incremental basis as the survey progresses. Adaptations to the context are reflected in the record layout before data are imported, based on the corresponding documentation. The verification of the database includes the following: 1) a unique ID check, 2) a valid value check for nominal/ordinal variables, 3) a valid range check for continuous variables and 4) cross-table consistency checks.

The delivery of data follows the adapted record layout, prior to any necessary recoding or mapping to re-establish the international record layout, if such a layout exists. Necessary structural adjustments following data submission will be done unless agreed differently. Data delivery is made through secure channels (i.e. the project's SharePoint site or a secured FTP connection), in a folder to which only specific users have access. All data available on web services are accessible only through encrypted connections (HTTPS/SSL) and access control mechanisms. Each user has only a limited set of user rights.

8 Coherence and comparability

The data collection follows the European Regulations and therefore are comparable to those of the other European countries, as the same concepts, definitions and classifications are used. All values of the variables - either collected by census or by the probability sampling survey - have the same target population and the same reference year. Long-term comparability is possible for most segments of the fishing fleet according to section 4.5.

The interviewers could affect to a great extend the data collection process which in turn influences the comparability of the results. The scientific coordinator reviews the planning reports to determine whether the survey requirements have been satisfied. To ensure comparability of the results across countries, the survey design plan will be consistent with the objectives and standards, methodologically sound and operationally practical. The review of the scientific coordinator also gives an advance opportunity to detect potential problems with the survey design plan and to provide advice on dealing with these issues.

To facilitate comparability in data analysis, the dataset is mapped into a highly structured, standardized record layout. In addition to specifying the position, format and length of each field, the record layout includes a description of each variable and indicates the categories and codes to be provided for that variable. Upon receiving the file, the data manager performs a series of range checks to ensure compliance with the prescribed format and run flow and consistency edits on the file. When anomalies are detected, proper personnel is notified of the problem and asked to submit cleaned files.

9 Confidentiality, Transparency and Security

9.1 Introduction

The dissemination of statistics deriving from the socio-economic variables is carried out by AGR.E.R.I following the statistical principles of the European Statistics Code of Practice, in particular the principle of statistical confidentiality. AGR.E.R.I takes all appropriate precautions to ensure that individual statistical units (fishing vessels) cannot be identified by technical or other means reasonably practicable by third parties.

The confidential information transmitted by the departments of the Ministry of Rural Development and Food and the Ministry of Maritime Affairs and Insular Policy to AGR.E.R.I are used exclusively for statistical purposes and only the authorized members of the AGR.E.R.I research team has the exclusive right of access to that information. In addition, the members of the AGR.E.R.I research team, in any employment relationship, are bounded by confidentiality and have the obligation to use the data accessed exclusively for statistical purposes. Any other use of such data is prohibited beyond the end of their duties.

9.2 Respondents

Participants in the survey are provided with the following information on the data collection delivered in person by the interviewer:

- funding, purpose and duration of research
- explanation of how the respondent was selected for the study
- interview procedures to be followed by the respondent
- voluntary nature of participation
- expected risks and benefits
- maintenance of confidentiality
- right to withdraw from the study at any time without penalty

The institute will maintain confidentiality of respondent data. All the involved staff sign a pledge of confidentiality and a non-disclosure form.

9.3 Researchers-AGR.E.R.I

Researchers maintain a copy of the following documents:

- scripts, letters, fact sheets and any other materials provided to give respondents information they need to make an informed decision about participation;
- consent protocols;
- confidentiality procedures and protocols;
- confidentiality agreements completed by staff;

AGR.E.R.I possesses substantial practical experience in the following areas:

- Survey management;
- Probability sample design and sample selection;
- In-person data collection;
- Computer-assisted personal interviewing (CAPI);
- Instrument and materials translation/adaptation;
- Interviewer training;
- Achievement of high response rates on surveys;
- Maintenance of data confidentiality;
- Data processing, including data capture, coding and editing;
- Sample weighting and estimation.

The staff is encouraged to report any confidentiality concerns that limits data sharing. A document describes in detail the encryption software on computers, the group policies which limit user access on computers, who has administrative control on the computers and how is security on folder level realized in the host operating system of computers.

A confidentiality agreement is signed by the involved staff including: Field supervisors; Interviewers; Coders; and Data processing personnel. These agreements are kept in a physical file for future use if appropriate. There is special education of training interviewers on the importance of data security. The main goal is to maintain the continued confidentiality and security of the survey materials and respondent data during data capture, coding, scoring, and processing. All of the work for coding, scoring, and processing will be carried out within the premises and reach of the survey organization.

Training of the interviewers include the following components:

- An introduction to survey research, providing examples of types of survey questions and interviewing terminology;
- The conventions for asking survey questions and recording answers;
- Written and oral exercises on asking questions, recording responses and applying probing techniques to obtain accurate data;
- Gaining respondent co-operation, in which the following concepts are discussed:
    - Interviewer behavior and style when making contact with the respondent;
    - The importance of making effective and fast connections with the respondent;
    - Methods to overcome resistance and address respondent concerns;
    - Written and oral exercises on refusal aversion techniques and how to answer respondent questions.
- Standards and ethics in survey research, including information on informed consent, data confidentiality, and data security and written exercises that include consent and confidentiality scenarios that interviewers may confront during data collection;

9.4 Data Collection Process

Initially, the respondent is located in the port and a letter printed on official stationery is provided. An attractive brochure is developed and it is given in conjunction with the official letter to the respondent during the contact. This brochure further serves to legitimize the study, stress the study’s importance and motivate respondent co-operation. The following information is included in the brochure:

- Why the study matters and why the respondent should participate
- Topics included in the survey
- How the respondent’s name and port was obtained
- Why another person cannot substitute the respondent

- Confidentiality of the data
- Users of the data
- Length of the interview
- URL of the survey website

Upon agreement of the respondent to participate in the study and the signing of the Confidentiality declaration form (see Section 9.5) a face-to-face interview is contacted and the collected data are transferred to the database by the interviewer.

9.5 Confidentiality Declaration Form

SAMPLE CONFIDENTIALITY FORM

Confidentiality Agreement

____

Name: ____

Phone number: ____

E-mail: ____

Address: ____

____

This is to certify that I, ____, have agreed to provide ____ services within the framework of the:

National Fisheries data Collection Program

I understand that the signing of this form is an acknowledgment of my professional responsibility to maintain complete integrity of security for this project. I declare that I will not divulge any project information, trial materials, test materials, processes, contents or results, or any other materials, documents, or information pertaining to the project, or its clients or suppliers, to any person or organization, as directed under the terms of the project.

I understand that the above does not apply to information that is in the public domain.

I have read and accept the conditions as outlined above.

Signature ____

Date ____

10 General Data Protection Regulation

10.1 Data Collection

AGR.E.R.I collects personal information from respondents using alternative methodologies. The collection is done with the consent of the respondents. The data collected relate to demographic, social and/or economic data relevant to the purpose of each survey.

10.2 Time of Keeping the Data in Identifiable Form

When conducting quantitative surveys, which are carried out by the researcher using the method of personal interviews, the Institute shall keep the data of the respondents in a personalized form. Such data shall be kept in this format for a limited time, only for the purposes of checking the quality and authenticity of the data and the reliability of the research.

The institute checks at least 15% of the completed questionnaires during or immediately after the survey is completed. Respondents' personal data, at the latest after three months, are anonymized to make it impossible to identify the personal data with the survey data.

10.3 Rights

Respondents, as long as the institute holds their personal data in identifiable form, have the following rights:

- The right to withdraw their consent. The revocation shall take effect upon its submission.
- The right to disclose their data. Respondents will be notified of their data within 15 days of receiving the request at the latest.
- The right to correct their data. The correction will apply to any processing that may occur after the initial disclosure of the data to the Institute.
- The right to delete their data. Deletion will take place no later than 30 days after the request is received.
- The right to report to the supervisory authority in case of breach of data security or illegal processing.

The above rights are exercised for all types of rights upon request to the Institute. The institute will contact interested parties to confirm the request while the email address with the request to the Institute's Data Protection Officer (DPO) is: Vasileios Paliouras (dpo@elgo.gr).

10.4 Transmission of Data to Third Parties

The Institute does not transmit personal data to third parties. Personal data are transmitted only with the consent of the respondent after being informed of the time the data is to be kept in identifiable form and the full contact details of the recipient. Once the data has been transmitted, the respondent may exercise his rights to the recipient for as long as his data is retained. The institute shall ensure the confidentiality of the interviewee's data prior to their dispatch, by agreement with the recipient. When the Institute uses subcontractors for the processing of personal data, it shall enter into a contract with them containing the standard contractual terms drawn up by the EU for this purpose in order to fully guarantee the rights of the respondents.

10.5 Confidentiality Agreements regarding the collection and processing of research data

The staff (Employees, Partners, and Technicians) of the Institute sign Confidentiality Agreement, under which they are bound by the data protection rules when collecting, processing, and analyzing the data of the respondents.

10.6 Data Protection Officer and Contact Details

Data Protection Officer (DPO) is Vasileios Paliouras, dpo@elgo.gr. The institute's offices are in Athens, Terma Alkmanos str., 11528 Ilisia. The contact number is 210 2756596

References

European Commission. 2006. Evaluation of the capital value, investments and capital costs in the fisheries sector. Study No. FISH/2005/03 (available at www.framian.nl/includes/download.asp?file_id=173).

European Commission, 2010. Commission Decision (2010/93/EC): Adopting a multiannual Community programme for the collection, management and use of data in the fisheries sector for the period 2011–2013. *Official Journal of the European Union*, L41/8, 18 December 2009.

European Commission, 2016. Commission Implementing Decision (EU) 2016/1251: Adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019. *Official Journal of the European Union*, L 207/113, 12 July 2016.

Eurostat, 2008. Survey sampling reference guidelines 2008 edition: Introduction to sample design and estimation techniques. Luxembourg: Office for Official Publications of EC.

Eurostat, 2014**.** ESS handbook for quality reports. General and regional statistics Collection: Manuals and guidelines. Office of the European Union.

Hollander, M., Wolfe, D.A. and Chicken, E. (2014). Nonparametric Statistical Methods, 3rd Edition, Wiley.

Horvitz, D. & Thompson, D., 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* 47: 663–685. EC study No. FISH/2005/03

Little, R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data. (Second Edition.) New York: Wiley.

Moura, C., 2016. *Quality guidelines for the DCF. Guidelines for socio-economic data. Ad hoc contract on quality.*

Pinello, D., Gee, J. & Dimech, M., 2017. Handbook for fisheries socio-economic sample survey: principles and practice. FAO Fisheries and Aquaculture Technical paper No 163, Rome, FAO.Dattalo, P., 2008. Determining sample size. Oxford: Oxford University Press.

Särndal, C. E., Swensson, B. and Wretman, J., 1992. Model assisted survey sampling. New York: Springer-Verlag

STECF (Scientific, Technical and Economic Committee for Fisheries), 2009. Report of the Sub-group on Research Needs (SGECA/SGRN 09-02). Edited by Paul Connolly & Tiit Raid. European Commission, 22-27 JUNE 2009, GALWAY.

Thompson, S., 2002. Sampling. Wiley New York.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison Wesley, Reading, MA

Annex

Fishing activity variables		Data Source	Data Collection
Capacity	GT	Fleet register	A
	kW	Fleet register	A
	Vessel Age	Fleet register	A
Effort	Days at sea	Fleet register / Questionnaire	A/B
	Fishing days	Fleet register / Questionnaire	A/B
	kW * Fishing Days	Fleet register / Questionnaire	A/B
	GT * Fishing days	Fleet register / Questionnaire	A/B
	Number of trips	Fleet register / Questionnaire	A/B
	Number of fishing operations	Fleet register / Questionnaire	A/B
	Number of nets/Length	Fleet register / Questionnaire	A/B
	Number of hooks, Number of lines	Fleet register / Questionnaire	A/B
	Numbers of pots, traps	Fleet register / Questionnaire	A/B
Landings	Value of landings total and per commercial species	Fleet register / Questionnaire	A/B
	Live Weight of landings total and per species	Fleet register / Questionnaire	A/B
	Prices by commercial species	Fleet register / Questionnaire	A/B

A: Census, B: Probability Sample Survey

Table A.1: The activity variables, the data sources and the data collection method.

	Economic variables	Data Source	Data Collection
Income	Gross value of landings	Questionnaire	B
	Income from leasing out quota or other fishing rights	N/A	N/A
	Other income	Questionnaire	B
Labour costs	Personnel costs	Questionnaire	B
	Value of unpaid labour	Questionnaire	B
Energy costs	Energy costs	Questionnaire	B
Repair & maintenance costs	Repair and maintenance costs	Questionnaire	B
Other operating costs	Variable costs	Questionnaire	B
	Non-variable costs	Questionnaire	B
	Lease/rental payments for quota or other fishing rights	N/A	N/A
Subsidies	Operating subsidies	Questionnaire	B
	Subsidies on investments	Questionnaire	B
Capital costs	Consumption of fixed capital	Fleet Register/ Questionnaire	A/B
Capital value	Value of physical capital	Fleet Register/ Questionnaire	A/B
	Value of quota and other fishing rights	N/A	N/A
Investments	Investments in tangible assets, net	Questionnaire	B
Financial position	Long/short Debt	Questionnaire	B
	Total assets	Questionnaire	B
Employment	Engaged crew	Questionnaire	B
	Unpaid labour	Questionnaire	B
	Total hours worked per year	Questionnaire	B

A: Census, B: Probability Sample Survey, N/A: Not Applicable.

Table A.2: The economic variables, the data sources and the data collection method.

Social variables	Data Source	Data Collection
Employment by gender	Questionnaire	B
FTE by gender	Questionnaire	B
Unpaid labour by gender	Questionnaire	B
Employment by age	Questionnaire	B
Employment by education level	Questionnaire	B
Employment by nationality	Questionnaire	B
Employment by employment status	Questionnaire	B
FTE National	Questionnaire	B

B: Probability Sample Survey

Table A.3: The social variables, the data sources and the data collection method