

11. 数据处理

产生有意义的渔业统计要求处理来自不同实地调查的数据。目前，最新的数据处理要求利用计算机系统。本节概述以下内容：

- 需要自动化程序；
- 基础系统功能；
- 数据流；
- 计算机化调查标准；
- 原始数据处理；
- 数据检查和监测；
- 预计过程；
- 基本报告功能；
- 培训和操作准则。

11.1 需要自动化程序

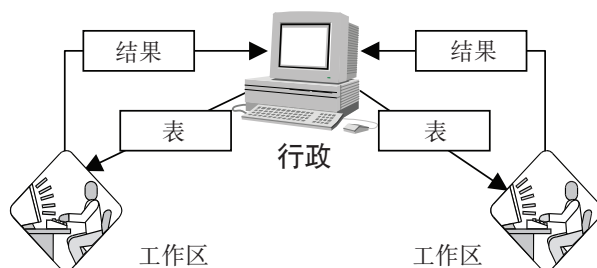
计算机和软件已经成为渔业统计系统中不可分开的部分，应当回应广泛的功能需要。其设计应当是：

- 灵活，以回应变化的调查标准。
- 足以避免软件干预。
- 标准组件，避免处理的瓶颈，允许地方官员处理和分析其自己的数据。
- 持续允许数据制造者定期运行，不需要外部提供协助。

基础渔业数据的计算机化系统的一般功能是：

- 组织调查标准。
- 组织和录入实地收集的样本数据。
- 数据质量报告。
- 预计数的自动计算。
- 预计数的基本报告。
- 与其他用户团体交流预计情况。
- 向其他应用环境输出结果。

11.2 数据流



上图提供了数据处理操作员和渔业统计计划行政管理中心之间数据流的简单系统结构的例子，包括：

- 中央单位（系统管理员或行政管理）负责确立调查标准，并向工作站分散单位分发。
- 分散单位在工作站一级对其自己的原始数据进行操作，产生预计。结果提交给行政管理中心进行整合。

这个结构或类似结构有以下优点:

- 调查标准是行政管理中心单位的职责, 保证了总体上符合确立的标准或有变化时则要求分发给工作站。
- 更靠近原始数据的来源地进行处理, 必要时能够更好和更快采取纠正行动。
- 分散录入原始数据, 减少有时在分散操作中出现的瓶颈。
- 分散用户首先考虑、检查和使用产生的预计。
- 基于分散结果的数据整合一般是行政管理中心单位更为容易的任务。

11.3 调查标准

明确的调查标准帮助减少实地操作, 产生一致性的报告, 将调查结果与来自其他分析和报告的应用情况整合。

11.3.1 调查标准的效力

调查标准通常在一个调查计划整个运行期间有效(通常一年), 此后要进行审议。但是, 调查框架有季节性变化, 因此调查标准的关键是要反映这类变化。

11.3.2 分层和地理区域以及地点

调查标准的计算机化的第一步是建立下表:

- 行政管理或报告分层(主要层)、船籍港和上岸点清单将加入到抽样操作中。
- 在预计情形中利用逻辑分层清单, 例如一个次要层、一个月和船舶/网具类型。

- 在次要层不与主要层重叠时，次要层和主要层相关联。
- 与船籍港和上岸点有关的次要层。这是指导在产量和努力量预计中利用原始数据的关键。

主要层		地点	
代码	描述	代码	描述
0001	沃尔特湖	0001	Dzemeni
0002	其他内陆水域	0002	Kedekope
		0003	Kpatsakope
次要层		0004	Kpeve Tornu
		0005	沃尔特湖分层II未明确的
代码	描述	0006	阿克拉镇
		0007	Gbetekpo
0001	沃尔特湖分层II	0008	Gbevukpo
0002	沃尔特湖分层VII	0009	Logakope
0003	沃尔特湖其他分层	0010	沃尔特湖分层VII未明确的
0004	01WB 将确定	0012	沃尔特湖其他分层（地点）
相关次要层 > 主要层		相关地点 > 次要层	
0001	沃尔特湖	0001	沃尔特湖分层II
0001	沃尔特湖分层II	0001	Dzemeni
0002	沃尔特湖分层VII	0002	Kedekope
0003	沃尔特湖其他分层	0003	Kpatsakope
		0004	Kpeve Tornu

上图提供了计算机确立分层、地点和其相关情况的例子。

11.3.3 船舶/网具类型

第二步是确立所有可能的船舶/网具类别表，在预先印制在数据收集表格时，这些类别应当容易被记录员辨认。

捕捞单位	
编码	描述
0001	木排
0002	竹排
0003	机动竹排
0004	大拉网
0005	撒网
0006	刺网
0007	机动刺网
0008	钩线
0009	机动钩线
0010	NIFANIFA
0011	定置网
0012	机动定置网
0013	卷网
0014	机动卷网
0015	万加拉网
0016	机动万加拉网

上图提供了计算机确立船舶/网具类型的例子。

11.3.4 框架调查

下一个任务中确立包含框架调查数据表，要求船籍港、上岸点和船舶/网具类型的相关表。

通常，计算机系统将按照地点 - 船舶/网具表运行，并编制包含所有“地点 - 船舶/网具”组合的空白记录。用户可用每个组合中潜在作业的捕捞单位记录完成这些记录。

地点和船舶/网具类型		单位
Dzemeni	撒网	5
Dzemeni	刺网	12
Dzemeni	机动刺网	6
Dzemeni	钩线	23
Dzemeni	机动钩线	11
Dzemeni	NIFANIFA	0
Dzemeni	定置网	3
Dzemeni	机动定置网	31
Dzemeni	卷网	7
Dzemeni	机动卷网	9
Dzemeni	万加拉网	12
Dzemeni	机动万加拉网	4
Kedekope	木排	67
Kedekope	竹排	12
Kedekope	机动竹排	19
Kedekope	大拉网	3
Kedekope	撒网	0
Kedekope	刺网	0

上图显示了计算机确立框架调查的例子。

11.3.5 物种清单

下一步是确立包含所有可能物种的物种表，在预先印制在数据收集表格时，物种名称应当容易被记录员辨认。

物种	
编码	描述
0001	贝尔鲑脂鲤
0002	西方项鲿
0003	鲿
0004	非洲脂鲤
0005	埃及尼罗鲿鱼
0006	结耙非鲫
0007	黑鳍金鲿;
0008	月琴脂鲤
0009	鳊胡鲶
0010	尖吻复齿脂鲤
0011	裸臀鱼
0012	红宝石鱼
0013	慈鲷
0014	倒立鲶
0015	尼罗河龙鱼
0016	福氏狗脂鲤
0017	狗脂鲤

上图显示了按物种的计算机设置的例子。

11.3.6 标准单位

重要的是以抽样为基础的调查涉及的计量单位在整个统计计划中要一致。本手册中考虑了以下的单位：

重量：应当所有调查实施中采用一致的单位。例如，如果记录上岸量商议的单位为千克，这一单位应当所有数据收集地点采用。（同样的概念适用于货币）。

努力量: 按定义, 努力量单位在不同船舶/网具和捕捞方式之间有差异。但在涉及基础渔业数据的调查中, 需要容易整合来自不同船舶和网具的产量和努力量预计。为统计目的, 一般接受的船-天是一致表示捕捞努力量的合理的好方法。

11.4 原始数据处理

需要处理的原始数据是实地收集的关于船舶活动性和上岸量的单个数据。设计和实施针对这些数据的计算机系统是复杂的任务, 要求相当的努力, 只能在这里简要回顾。

11.4.1 录入船舶活动性数据

计算机程序必须足够灵活, 以处理不同抽样计划收集的数据。数据录入直接从按月、按船籍港或按船舶/网具类型排列的文件进行。

船舶活动性调查-2000年7月-船籍港表-船舶/网具类型: 大拉网											
活跃 抽样 框架			活跃 抽样 框架			活跃 抽样 框架			活跃 抽样 框架		
1			9			17	4	6	25		
2			10			18	3	7	26		
3	4	8	11			19			27	5	8
4	3	5	12			20	3	6	28		
5			13			21	3	5	29		
6			14			22			30		
7			15	3	6	23			31		
8			16	4	4	24			记录员 H.Y. TSIKPO		

上图为用于录入关于船舶活动性数据的通用计算机屏幕的例子。记录活跃船舶数以及在特定一天在一个船籍港抽样的总数, 尽管在有要求时, 也提供框架调查数据。

11.4.2 录入活动天数数据

活动天数数据提供了在次要层、一个月以及具体船舶/网具类型的预计情形中预计捕捞努力量的时间提高因素。因此，计算机表将包含次要层和船舶/网具类型的所有组合。这些可以由计算机系统自动创建。对特定月份，这些记录需要按活动天数对应的每一组合更新。最初，该表包含的是零。

次要层和船舶/网具类型		天数
KETU	APW独木舟	27
KETU	大拉网	27
KETA	APW独木舟	27
KETA	大拉网	27
KETA	定置网	27
KETA	流刺网	27
东DANGBE	APW独木舟	27
东DANGBE	大拉网	0
东DANGBE	定置网	27
东DANGBE	流刺网	27
西DANGBE	APW独木舟	0
西DANGBE	大拉网	27
西DANGBE	钩线	27
西DANGBE	定置网	0
西DANGBE	流刺网	27
TEMA市	APW独木舟	0
TEMA市	大拉网	0
TEMA市	钩线	0
TEMA市	定置网	0

11.4.3 录入上岸量数据

上岸量数据录入直接从按月、分层、船籍港或船舶/网具类型排列的文件进行。下图为用于录入取样的上岸量的通用计算机屏幕例子。

上岸量调查-2001年7月 表格: 0034	
分层: SW沿岸	船舶/网具类型:
上岸点: DENU	
日期: 2001年7月23日	单位: 1
记录员:	时期: 2天
Yelowomi Paul	抽样: 147千克

物种	重量	数量	价格	价值
鲷	5	16	6.5	32.50
石首鱼	0	0	0	0.00
圆小沙丁鱼	124	0	1.29	160.00
龙虾	0	0	0	0.00
未列名小沙丁鱼	0	0	0	0.00
蓝圆海参	0	0	0	0.00
鲷	18	0	0	108.00
未列名的鲷鱼	0	0	0	0.00
海螺	0	0	0	0.00
西鲱	0	0	0	0.00
鲨鱼	0	0	0	0.00
对虾	0	0	0	0.00

11.5 数据检查和监测

在产生捕捞努力和产量预计之前，为确定完成情况以及原始数据的质量，必须对特定数量的数据进行检查和监测。这类控制功能涉及：

- **监测**: 总结清单和报告将提供在每个预计情形中的船舶活动性以及上岸量样本可获得性的快速显示。
- **数据范围检查**: 提供的清单显示产量、样本努力量和价格的“末端”值（数值范围）。这些必须与录入数据的格式自动链接。在怀疑数值高或低时，可检查清单进行核实。
- **样本规模检查**: 提供的清单显示预期的船舶活动性以及上岸量的样本规模和准确性。这些按照推理决定，在第9和10节做了论述。

11.6 预计程序

以计算机为基础的预计过程涉及以下计算步骤:

11.6.1 预计捕捞努力量

- (a) 采用船舶活动性样本、活动天数和框架调查数据对一个次要层、一个月和一个船舶/网具类型情形进行适当预计。
- (b) 在每个情形中计算BAC。
- (c) 计算预计的BAC准确性。
- (d) 计算总BAC变化和其置信限。
- (e) 按空间和时间解释BAC变化。
- (f) BAC将活动天数和框架调查数据结合，产生捕捞努力量预计。
- (g) 计算努力量变化和置信限。

11.6.2 预计产量和价值

采用抽样的上岸量对一个次要层、一个月和一个船舶/网具类型情形进行适当预计。

- (a) 在每个情形计算总CPUE。
- (b) 计算预计的CPUE准确性。
- (c) 计算总CPUE变化以及其置信限。
- (d) 按空间和时间解释CPUE变化。
- (e) 计算样本物种比例。
- (f) 计算样本价格。
- (g) 产生平均鱼体规格预计（重量）。
- (h) 预计的CPUE与努力量预计组合，产生总产量预计。
- (i) 计算产量预计和相关置信限的变化。
- (j) 样本物种比例与按总产量的预计组合，产生按物种的产量预计。
- (k) 样本价格与按物种的产量组合，产生按物种的价值预计。
- (l) 合计按物种的价值，产生上岸量总价值。

上述计算步骤对每一个次要层、月份和船舶/网具类型情形重复进行。在该过程末尾，产生以下的数据分程序：

11.6.3 数据分组

- (a) 在主要层和总计一级将产量、努力量和价值分组。
- (b) 在主要层和总计一级计算平均CPUE和价格。

11.7 基本报告

在编制预计数据的基本报告方面有多种办法。一般在按月报告产量/努力量预计的报告功能中，其构成“第一代”统计数。应当考虑以下几点：

- (a) 第一级报告应当是预计情形（分层），在此产生所有的计算和相关统计指标以及诊断。
- (b) 在分析结果前，用户应当检查系统信息，确定每个预计情形的完成水平。
- (c) 必须报告预计过程涉及的所有数据，需要时允许对结果的手工核实。
- (d) 报告顺序一般应当按照11.7节论述的计算步骤进行。

11.7.1 系统诊断

以下的例子显示在（不完全）预计过程中的系统信息。对每个预计情形，要显示预计结果描述。

不同预计情形显示的信息通知用户：

- (a) CPUE的准确性低于90%。继续预计。
- (b) 没有活动天数以及无框架数据（因此无提高因数）。不能预计。

(c) 无上岸量或数据。不能预计。

(d) 有限的地理覆盖范围。BAC和CPUE的准确性低于90%。

KETA	大拉网	预计
CPUE准确性低于90%		
.....		
KETA	钩线	不能预计
无活动天数		
无框架数据		
.....		
KETA	定置网	不能预计
无上岸量		
无努力量数据		
.....		
KETA	流刺网	预计
只有1个上岸点		
努力量只有1个点		
BAC准确性低于90%		
CPUE准确性低于 90%		
计算的CPUE无差异		

11.7.2 预计努力量

在以下的例子中，在3部分描述了预计的努力量。

- (a) 预计的BAC和产生的准确性可按照展示的抽样信息进行核实。
- (b) BAC变化大（29%），按空间和时间进行解释。注意时间变化（20.5%）是主因。
- (c) 预计的捕捞努力量可采用预计的BAC和活动天数以及框架调查提高因子核实。还要展示预计努力量的置信限。

KETA: 大拉网

努力量预计

BAC-船舶活动性系数.....	25.000%
准确性.....	91.173%
抽样单位.....	120
活跃.....	30
地点数	2
天数.....	10
BAC变化.....	28.912%
BAC空间变化.....	8.393%
BAC时间变化.....	20.920%
95%时BAC低限.....	10.833%
95%时BAC高限.....	39.167%
框架调查中的单位.....	168
活动天数.....	27.000
预计努力量（天）.....	1134
95%时努力量低限.....	491
95%时努力量高限.....	1777

11.7.3 预计产量

在该例子中，在三个部分描述了总产量预计。

- (a) 预计的总CPUE和产生的准确性可按照展示的抽样信息进行核实。注意产生的准确性稍低于可接受的90%，原因是采用了30个样本，而不是要求的31个。
- (b) CPUE变化大（32%），按空间和时间进行解释。注意时间变化（27.5%）是主因。
- (c) 预计的总产量可采用预计的CPUE和上文描述的预计的捕捞努力量核实。产量的复合变化很高（43%），原因是CPUE和捕捞努力量变化大。还要展示预计总产量的置信限。

产量预计

CPUE.....	402.967
准确性.....	89.798%
准确性90%时的样本规模.....	31
抽样上岸量.....	30
抽样产量.....	12,089
抽样努力量.....	30
地点数.....	2
天数.....	10
CPUE变化.....	31.993%
CPUE空间变化.....	4.421%
CPUE时间变化.....	27.572%
95%时CPUE低限.....	150.284
95%时CPUE高限.....	655.650
预计产量（千克）.....	456,964
产量变化.....	43.121%
95%时产量低限.....	70,747
95%时产量高限.....	843,182

11.7.4 按物种的产量

在以下例子中，按物种的结果展示在3个栏目中，描述了：

- (a) 按物种和有关努力量的预计产量。
- (b) 按物种的CPUE。
- (c) 每类物种的平均重量。
- (d) 按物种的抽样价格和预计产值。

所有上岸量的总价值和其单位价值的概要见该报告顶部。

总价值 (1,000C) 221 571
 均价 (1,000C/千克) 0.485

按物种的产量	重量 努力量	CPUE 平均重量	价值 价格
鳀鱼	362,899 (79.4 %) 1,134	320.017 0.000	152,244 (68.7 %) 0.420
狐鲣	26,366 (5.8 %) 1,134	23.250 0.000	8,490 (3.8 %) 0.322
圆小沙丁鱼	29,030 (6.4 %) 1,134	25.600 0.000	28,341 (12.8 %) 0.976
蓝圆鲹	38,669 (8.5 %) 1,134	34.100 0.000	32,496 (14.7 %) 0.840

11.7.5 总计

以下的例子显示计算具体船舶/网具类型的总合计（流刺网）。数字来自不同次要层该船舶/网具类型所有统计的分组。

总合计: 流刺网			
框架调查单位.....	4		
预计努力量.....	27		
CPUE.....	35.000		
预计产量（千克）.....	945		
总价值（1000C）.....	851		
均价（1000C/千克）.....	0.900		
按物种产量	重量 努力量	CPUE 平均重量	价值 价格
大西洋小型金枪鱼	203 （21.4 %） 27	7.500 0.000	162 （19.0 %） 0.000
鲨鱼	473 （50.0 %） 27	17.500 0.000	473 （55.6 %） 1.000
鲣鱼	270 （28.6 %） 27	10.000 0.000	216 （25.4 %） 0.800

11.8 培训和操作准则

基础渔业数据的计算机系统总体评估不仅涉及设计标准，还涉及渔业工作人员有效运行该系统的能力。培训包括：

- 由数据操作员控制所有系统功能。
- 定期准备数据备份。
- 可用性快速启动的系统操作指南。
- 进一步处理评估产量/努力量预计数的方法。
- 数据录入、预计和提交数据程序的有效监测。

概 要

本节介绍了有关基础渔业数据自动化处理的一般情况，包括：

- (a) 自动程序需要有力、标准组件和持续的计算机系统
- (b) 系统基本功能
- (c) 数据流。分散系统结构的优点
- (d) 计算机化调查标准，包括分层、地点和联系；物种和船舶/网具分类；框架调查、活动天数以及标准计量单位
- (e) 处理关于船舶活动性和上岸量数据
- (f) 数据检查和监测
- (g) 预计过程、涉及数据、统计指标和诊断
- (h) 基本报告职责
- (i) 培训和操作准则的重要性