



Food and Agriculture Organization  
of the United Nations



NATIONAL BUREAU OF STATISTICS



---

## **Statistical analysis on the data collected from the PRICE survey 2022 implemented in Nigeria within the 50x2030 Program**

---

### **In depth analysis of fisheries price data, NASS 2022/23**

**By TIOTSOP TSATSOP Armel Blaise**

*Statistical- Economist Engineer,*

*Statistician, ESS*

**Content**

I. Analysis of the quality of data on price..... 4

II. Comparative view of livestock price by geographic area ..... 6

III. Conclusion and recommendations ..... 8

## **FOREWORD**

This document has two objectives. It's intended to the following:

- 1) give an overview of the quality of livestock price data collected through the 2022 price survey in Nigeria, and
- 2) inform on the homogeneity of prices between geographic area

## I. Analysis of the quality of data on price

This section answers two main questions

- i) have we succeeded in improving the completeness and quality of our data, particularly reduce the influence of extreme values?
- ii) Are the rates of missing values reasonable?

**Regarding edition and imputation** of outlier, as a reminder, price processing was carried out in five (5) stages:

- I) check for missing values;
- II) analysis of price consistency within the same combination of unit x size prior to the conversion: that lead to some imputation: for example, (i) by dividing certain excessively large values by 10 or 100; (ii) sometimes by imputing units or size, to make them consistent with given prices. For example, in zones or states where few communities have reported some fisheries, prices were quite high.
- III) converting price into kilogram equivalents, using conversion factor available;
- IV) (micro) edition and imputation of outliers (using p15 lower threshold and Q3/4 + 2\*IQR upper threshold). *That lead to impute about 22% of prices.*
- V) (macro) edition and imputation to obtain estimates (median and mean) that are consistent between states in the same zone. *That lead to impute about 11% of prices.*

Overall, prices varied widely from one area to another. The zone South South is the area of most fish farming, and all fish species are found here. Strictly speaking, only tilapia and catfish are produced in all zones. All other fish species are rather rare in the other zones (see Table1).

**Table 1: Communities that reported fisheries by zones**

Fisheries	North Central	North East	North West	South East	South South	South West	Total
<b>Tilapia</b>	<b>311</b>	<b>353</b>	<b>420</b>	<b>101</b>	<b>545</b>	<b>102</b>	<b>1,832</b>
<b>Catfish</b>	<b>446</b>	<b>345</b>	<b>421</b>	<b>358</b>	<b>796</b>	<b>406</b>	<b>2,772</b>
Carp	43	14	35	2	<b>105</b>	9	208
Shrimps	6	2	13	0	<b>189</b>	8	218
Prawn	7	1	5	2	<b>160</b>	11	186
Oysters	2	0	8	5	<b>69</b>	1	85
Water Snail	24	4	8	29	<b>169</b>	22	256
Common Carp	18	1	4	0	<b>44</b>	3	70
Heterotis	13	6	19	0	<b>102</b>	11	151
Lates	0	0	2	0	<b>48</b>	3	53
Gymnacus	2	1	3	0	<b>63</b>	11	80
Other fish(specify)	6	1	2	0	<b>104</b>	9	122
Total	878	728	940	497	<b>2,394</b>	596	6,033

Due to the small number of communities that reported producing some fisheries, it was not possible to edit and impute extremes values independently in the different zones. Then the strategy to edit and impute outliers had been made as follow:

- For tilapia and catfish, editing and imputation had been made separately by zone;
- For zone South South, editing and imputation had been made for each fish;
- For all other fisheries, editing and imputation had been made at national level (by excluding the zone South South).

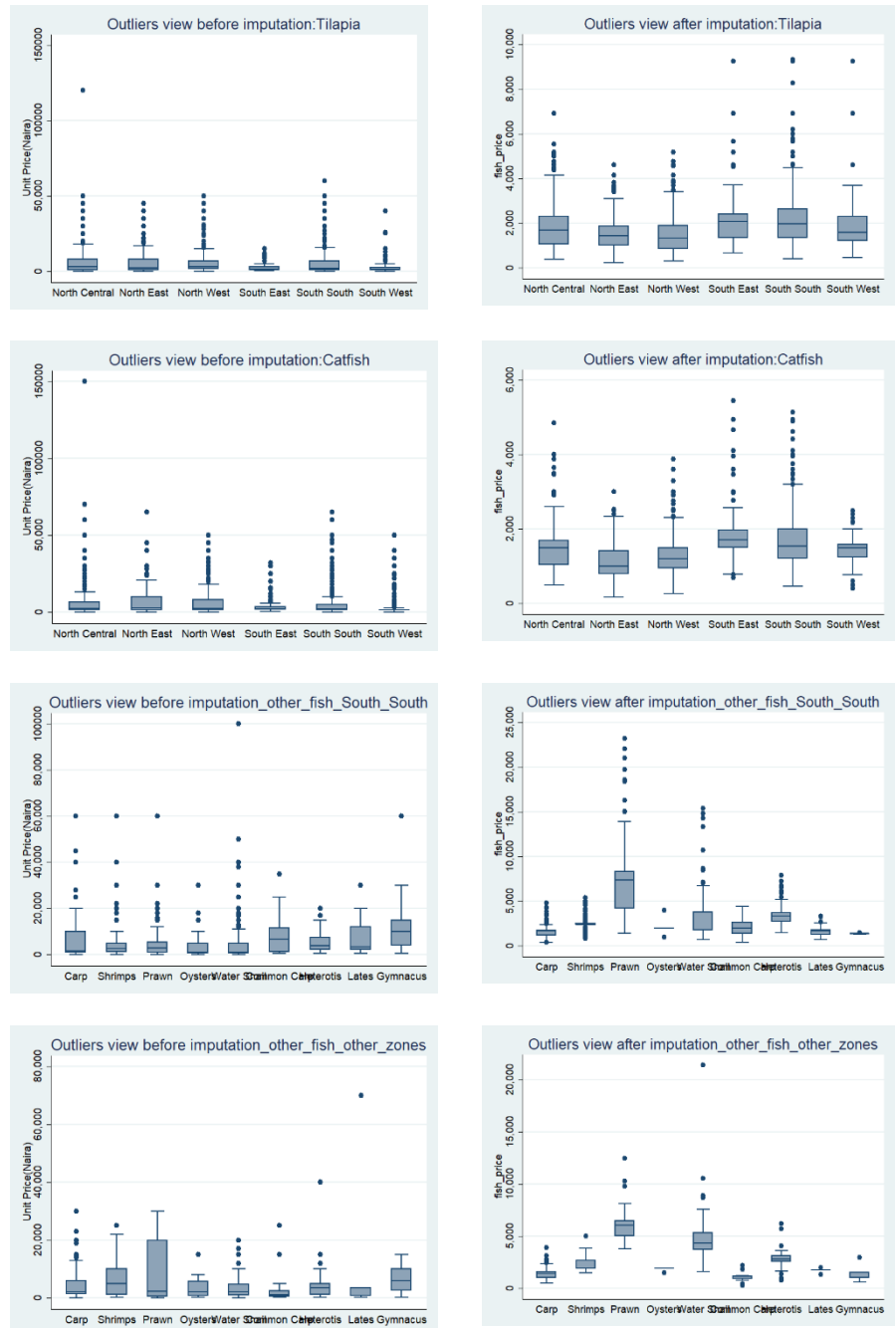
**Regarding outliers:** Box plots of prices before and after the data processing are presented side in Fig 1. It can be seen that the ranges of values are greatly reduced throughout the process. Outliers were imputed using measures of central tendency (median and averages prices for states and zones).

Fig. 1: By Livestock, box plots of price comparison before versus after data processing

If we consider for example the case of Tilapia. Before data cleaning process (*the left box plot*) there is a very wide range of data on price, with values over 100,000 and many values over 25,000. After cleaning process (*the right box plot*) all price are under 10,000.

In the box plot on the right, we can still observe some outliers. Obviously, after one stage of imputation of outliers, news one appears. In addition, analysis of the first price estimates revealed significant differences between the median prices of states within the same zone.

Then to bring the median price spreads between states in the same zone down to a more reasonable scale, rather than systematically imputed a large number of all new extreme values, the solution was to proceed via macro editing, by identifying and progressively imputing outliers, while monitoring the effect on reducing price differentials between states. This macro edition step has led to the imputation of about 11% of values of prices.



**Finally, through the various processes of data cleaning, about 33% of data on price had been imputed using central tendency values of prices**

If the same central tendency value was used for a large number of imputations, this would lead to almost identical median prices almost everywhere. To prevent this from happening, in a controlled way, the values used for imputations were varied between national, zone or state averages and medians.

**Regarding missing data**, we still have few missing on conversion factor, concerning 197 of row (3.3% of the sample). Conversion factor had been used for main fisheries. Table 2 gives us the unit with missing conversion factors, by fisheries.

**Table 2: By Fish, units with missing conversion factors**

Unit	Oysters	Gymnacus	Other fish	Total
PIECE SMALL	0	2	6	8
PIECE MEDIUM	1	17	25	43
PIECE LARGE	1	31	16	48
HEAP SMALL	2	5	13	20
HEAP MEDIUM	2	1	7	10
HEAP LARGE	0	0	0	0
BASKET SMALL	2	0	6	8
BASKET MEDIUM	7	0	7	14
BASKET LARGE	0	0	2	2
BASIN SMALL	5	0	3	8
BASIN MEDIUM	6	10	12	28
BASIN LARGE	1	0	7	8
Total	27	66	104	197

Based on the above two points, we can conclude that data was comprehensive, and the quality had been improved by the clearing process.

## II. Comparative view of livestock price by geographic area

The final stage of price data processing was about macro edition, consisted in a review of the estimates guided by the objective of having comparable (median) results between the states of the same zone. At the end of the process, therefore, it makes sense to verify if there are some inequalities between states of the same zone, and between different zones. To do this, the following tests are made:

- i) **To assess the similarity of prices between states within zone, we compare medians and means:** The “K-sample equality-of-medians test”: the null hypothesis that the K samples were drawn from populations with the same median. To say that the (median) prices of states are the same or comparable within zones implicitly refers to the null hypothesis. If not, the medians are statistically distinct between states of a zone (Table3).

The strategy for this test uses the procedure chi-squares test as follow: Firstly, for each livestock, the sample of each zone is split into two groups: communities with price values greater than (or equal to) the median (the group “yes”), and the groups with price less than the median (the group “no”).

Then, the chi-squares test is performed by cross-tabulating the group variable with the states (within the zone). Since the chi-square test is an asymptotic test, there must be at least no (or few) expected values below five. Then, in respect of the strategies used for imputing outliers, we compare values of prices between states and within each zone only for tilapia and catfish. For all others fisheries, we firstly compare prices between states of South South (Table 3). Secondly, we compare price of this zone with those of all other zones together, using a t-test (table 4).

Fisheries for which the medians are statistically equal between geographic area are highlighted in table 3. We can see that in most cases, medians are statistically

different. For confirmation purposes, we also perform a comparison of means test. To the Fisher statistic, we add the (robust) Welch's test which corrects the Fisher statistic in the event of unequal variances between groups (see Table3 and Table4).

**Tests of equality of medians / means between states within zones:** Table3 contains results of the two tests (equality of medians and equality of means). The main finding is that, medians and averages of prices are different between states and between zones, at a 5% level.

**Table 3: By livestock and by zone, K-sample equality-of-medians test and tests of equality of means**

Livestock	K-sample equality-of-medians test ( <i>H<sub>0</sub>: Prices of states came from a population with a same median</i> )			Homoscedasticity test ( <i>H<sub>0</sub>: equal variance</i> )				Tests of Equality of Means ( <i>H<sub>0</sub>: Equal means between state</i> )			
	Zone	Pearson chi2	P-value	Statistic of Levene	ddl1	ddl2	P-value	ANOVA (Fisher test)		Welch Correction	
								F	P-value	Statistic of Welch	P-value
Tilapia	North Central	39.96	<0.001	24.9	6	304	<0.001	14.8	<0.001	5.4	<0.001
Tilapia	North East	40.79	<0.001	4.4	5	342	0.001	7.3	<0.001	9.6	<0.001
Tilapia	North West	119.559	<0.001	8.9	5	405	<0.001	16.5	<0.001	72.4	<0.001
Tilapia	South East	10.48	0.015	5.5	3	96	0.002	1.9	0.140	4.8	0.007
Tilapia	South South	168.59	<0.001	15.8	5	535	<0.001	34.5	<0.001	46.0	<0.001
Tilapia	South West	33.80s	<0.001	9.4	4	96	<0.001	3.3	0.015	22.7	<0.001
Catfish	North Central	36.07	<0.001	6.6	6	439	<0.001	8.4	<0.001	8.7	<0.001
Catfish	North East	132.09	<0.001	7.8	5	333	<0.001	33.2	<0.001	54.8	<0.001
Catfish	North West	133.51	<0.001	13.2	5	407	<0.001	30.3	<0.001	48.9	<0.001
Catfish	South East	58.18	<0.001	64.5	4	353	<0.001	88.5	<0.001	15.6	<0.001
Catfish	South South	77.73	<0.001	21.4	5	786	<0.001	34.6	<0.001	30.5	<0.001
Catfish	South West	114.97	<0.001	4.6	5	400	<0.001	21.7	<0.001	24.0	<0.001
<b>Comparison between states of South South, by fisheries</b>											
Carp	South South	11.25	0.024	2.3	4	98	0.069	1.1	0.379	1.0	0.424
Shrimps	South South	12.08	0.017	1.4	4	161	0.243	2.1	0.082	2.5	0.047
Prawn	South South	23.31	<0.001	6.7	4	139	<0.001	7.5	<0.001	5.2	0.001
Water Snail	South South	22.08	0.001	0.7	4	136	0.606	1.5	0.202	2.4	0.061
<b>Common Carp</b>	<b>South South</b>	<b>5.02</b>	<b>0.170</b>	18.8	1	38	<0.001	3.7	0.061	3.5	0.074
Heterotis	South South	21.87	<0.001	1.0	3	98	0.403	1.9	0.134	4.2	0.015
Lates	South South	5.94	0.114	0.6	2	36	0.553	3.8	0.032	5.3	0.014

**Table 4: By Fisheries, K-sample equality-of-medians test and tests of equality of means: comparing South South with others zones<sup>(a)</sup>**

Fisheries	Test of median		Test for Equality of Variances			t-test for Equality of Means			
	Pearson chi2	P-value	Equal variances modality	Levene	P-value	t	df	P-value	Mean Difference
Tilapia	66.34	<0.001	Assumed	19.422	<0.001	-8.581	1810	<0.001	-567.08
			Not assumed			-7.934	864.4	<0.001	-567.08
Catfish	32.47	<0.001	Assumed	1.009	0.315	-6.070	2752	<0.001	-206.92
			Not assumed			-6.395	1642.5	<0.001	-206.92
Carp	<b>0.55</b>	<b>0.457</b>	Assumed	1.425	0.234	-1.632	156	0.105	-220.45
			Not assumed			-1.826	146.9	0.070	-220.45
Water Snail	5.86	0.015	Assumed	10.934	0.001	4.836	187	<0.001	2541.16
			Not assumed			4.060	63.1	<0.001	2541.16
Heterotis	24.79	<0.001	Assumed	0.922	0.339	-3.295	138	0.001	-778.02
			Not assumed			-3.491	74.6	0.001	-778.02
Shrimps	10.44	0.001							
Prawn	17.38	<0.001							
Oysters	23.38	<0.001							
Common Carp	24.47	<0.001							
<b>Lates</b>	<b>0.74</b>	<b>0.390</b>							
Gymnacus	0.31	0.577							

(a) Whenever there is no statistics, it could not be computed

### **III. Conclusion and recommendations**

---

The data-processing process has improved the pertinence of the data on price, mainly by imputing extreme values and many missing values, and also by imputing some units and size to be consistent with the price recorded by the surveyor.

Then, estimated median prices are significantly different between zones and between states of the same zone.

Due to imputation strategies and low numbers in some areas and states, median estimates by state should be taken with caution.