

Informe Año 2012

CÁMARA
COLOMBIANA
DEL LIBRO

INFORME VALIDACIONES REALIZADAS



INFORME SOBRE VALIDACIONES EFECTUADAS

PRELIMINAR

Con el fin de obtener datos de calidad, durante la operación “Estadísticas del Libro en Colombia”, se realizan cuatro tipos de validaciones durante el proceso:

1. En la recolección de datos
2. En el análisis y verificación de la información
3. En la tabulación de los datos
4. En la base de datos consolidada

Para cada actividad se realizó validaciones diferentes, estas dependen de la labor desarrollada.

- A. **Recolección de datos:** Actividad realizada mediante el diligenciamiento de la encuesta, la cual está diseñada en formato Excel y que cuenta con validaciones en algunas de sus celdas.
- B. **Análisis y verificación de la información:** Una vez obtenida la información de las encuestas, estos datos son cotejados con bases de datos internas y externas
- C. **Tabulación de los datos:** Es el paso en donde se ingresa la información de cada una de las empresas y de cada una de las variables, en un formato en Excel tipo base de datos. Este archivo cuenta con las mismas validaciones empleadas en el formato de la encuesta.
- D. **Base de datos consolidada:** Tiene relación con las reglas de validación establecidas para la operación estadística y que define los parámetros del contenido de cada una de las celdas de la base de datos final.

DESARROLLO

Las validaciones que se estipularon en el documento metodológico Reglas de Validación de Integridad de Datos, se desarrollaron de la siguiente manera:

Validación A – Recolección de datos: Al momento de recibir y revisar las encuestas, 10 de estas encuestas presentaron error en el diligenciamiento. El total de errores presentados por este tipo de validación es de 65, siendo la variable 79 (ventas por soporte formato papel, número de ejemplares) la que registra el mayor número de errores. El total de variables con inconsistencias es de 12 y 10 empresas presentaron esta situación.

Se realizó la verificación de la información de cada una de las encuestas, confirmando los datos con el responsable de diligenciar la encuesta, de esta manera se validan la consistencia de la información suministrada.

Validación B – Análisis y verificación de la información: En el cruce y análisis de la información entre las encuestas recibidas y las bases de datos (internas y externas), se encontró inconsistencia de datos en 17 encuestas recibidas, con un total de 28 errores en esta validación y 8 variables presentaron inconsistencias de este tipo.

La variable 78 (ventas netas totales), presenta el mayor número de empresas con error al cotejar el dato suministrado en la encuesta con las bases de datos de Supersociedades y Colombia Enterprise.

La acción seguida para corregir esta inconsistencia, es verificar directamente con la fuente de primaria de la información (la editorial) la veracidad del dato.

Validación C – Tabulación de los datos: Luego de realizar las dos validaciones anteriores, se procede con la tabulación de la información en lo que será la base de datos consolidada. Una vez termina esta actividad se revisa la base de datos para verificar que no se presenten errores de digitación.

Como resultado de este proceso de encontraron 5 variables con errores de digitación, en algunos casos se debe a las cifras en decimales que provienen de la encuesta y los otros casos errores humanos. Para solucionar el impase se verifica la información de cada una de las encuestas.

El total de errores de esta categoría es de 5, uno por cada variable.

Validación D – Base de datos consolidada: Esta categoría requiere aplicar las reglas de validación establecidas para la operación estadísticas, a cada una de las variables de la base de datos.

En este proceso se detectaron 17 variables con inconsistencias, en 20 empresas para un total de 44 errores. Mediante la verificación de la regla de validación y la información suministrada en la encuesta se realiza el ajuste en la base de datos. En algunos casos fue necesario consultar la fuente de información primaria para corroborar las cifras.

CONCLUSIONES

RESUMEN

| | |
|-----------------------------|-----|
| Total Errores | 142 |
| Total empresas con errores | 21 |
| Total variables con errores | 42 |

| Validación | Número de Variables | Número de Errores | Número de Empresas |
|------------|---------------------|-------------------|--------------------|
| A | 12 | 65 | 10 |
| B | 8 | 28 | 17 |
| C | 5 | 5 | 3 |
| D | 17 | 44 | 20 |
| | 42 | 142 | |

La validación A es la que presentó el mayor número de errores con el 46% del total de errores generados en la operación. En la mayoría de los casos por un mal diligenciamiento de la encuesta, debido a una mala interpretación de la pregunta consecuencia de no leer el manual de diligenciamiento.

La validación D presenta el mayor número de variables con error y que corresponde al 40% del total. Esta falla se puede atribuir igual a la falta de lectura del manual de la encuesta previo a su diligenciamiento, pues la mayoría de campos que presentaron errores, las aclaraciones están en el manual.

La validación B, registró el 20% de errores, siendo la principal inconsistencia la cifra total de ventas que al ser cotejada con las bases de datos externas no coincidía.

La validación C es la que presenta el menor número de inconsistencias en relación al total de errores, número de empresas y número de variables con error. Cada variable presentó una inconsistencia, relacionada con la validación de celdas realizada a la base de datos y unos errores de digitación.

Al finalizar la actividad se logró obtener que la calidad y consistencia de los datos almacenados en la base de datos consolidada, cumplan con los parámetros que requiere la operación estadística.