

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE INGENIERIA QUIMICA Y TEXTIL



“Importancia de la Implementación de Sistemas de Reconciliación de Datos y  
Detecciones de Errores Gruesos de medición de instrumentos, en Plantas Industriales  
Químicas y Petroquímicas”

**INFORME DE SUFICIENCIA**

PARA OPTAR EL TITULO PROFESIONAL DE:

**INGENIERO QUÍMICO**

POR LA MODALIDAD DE ACTUALIZACIÓN DE CONOCIMIENTOS

PRESENTADO POR:

JUAN ANTONIO BARBARAN VILLANUEVA

LIMA – PERÚ

2013

## DEDICATORIA

A mi amada esposa María,  
a mis Padres Juan y María,  
a mis suegros Jorge y Amparo,  
a mi hermana Claudia, a mi  
cuñado Jorge, a mi padrino  
Pedro y mi madrina Edith.  
Gracias a ustedes por el apoyo  
y cariño que me brindan  
a cada instante.

## RESUMEN

Las técnicas de reconciliación de datos y detección de errores gruesos de medición, son metodologías que son capaces de mejorar la precisión de los datos de planta por encima de la precisión de la medida de instrumentos y que también son capaces de detectar errores gruesos y pérdidas, tanto en estado estacionario como dinámico de ser este el escenario.

En el capítulo II, se muestra la importancia de la reconciliación de datos y nos indica alguna de sus particularidades, por ejemplo establece que la solución de una reconciliación de datos obtenida de la medición de instrumentos con errores gruesos no es confiable, dado que estos grandes errores se dispersan en todas las variables, causando ajustes infundados.

La reconciliación de datos y la detección de errores gruesos están estrechamente ligadas, por lo que es necesario que sean implementados a la vez para obtener una reconciliación confiable, indica además que solo las leyes de conservación de materia y energía son aceptadas en la reconciliación de datos. Las correlaciones o relaciones aproximadas entre variables del proceso no son recomendables dado que son fuentes adicionales de error. Los datos reconciliados son usados para realizar optimización en línea, planeamiento de la producción y mantenimiento correctivo o preventivo de instrumentos o tanques de almacenamiento, etc., debido a su alta confiabilidad.

En el capítulo III se muestra más detalle sobre la clasificación de los errores de medición y las técnicas de reducción de los mismos, indicando que los errores de medición que ocurren a menudo en la instrumentación del proceso pueden ser pequeños (aleatorios), mientras que otros son mayores y sistemáticos (errores gruesos).

En el capítulo V se muestra aplicaciones de reconciliación de datos y detección de errores gruesos en la industria.

## ÍNDICE

<b>I. INTRODUCCIÓN</b> .....	5
<b>II. IMPORTANCIA DE LA RECONCILIACIÓN DE DATOS Y LA DETECCIÓN DE ERRORES GRUESOS</b> .....	7
2.1 Métodos de acondicionamiento de los Datos de Proceso.....	7
2.1.1 Procedimiento del Sistema de Reconciliación de Datos.....	12
2.2 Formulación del problema de Reconciliación de Datos .....	14
2.3 Beneficios De La Reconciliación De Datos .....	19
<b>III. ERRORES DE MEDICIÓN</b> .....	22
3.1 Clasificación de los errores de medición.....	22
3.1.1 Errores aleatorios .....	23
3.1.2 Errores gruesos .....	23
<b>IV. RECONCILIACIÓN DE DATOS Y LA ESTADÍSTICA</b> .....	26
4.1 Tolerancias e intervalo de confianza .....	26
4.2 Influencia de datos.....	30
<b>V. INSTRUMENTOS DE MEDICION DE FLUJO Y NIVEL</b> .....	31
5.1 Medidor de flujo, Placa de orificio.....	31
5.2 Medidor de nivel por radar en tanques de almacenamiento .....	33
<b>VI. APLICACIONES DE RECONCILIACIÓN DE DATOS Y DETECCIÓN DE ERRORES GRUESOS</b> .....	35
6.1 Balance de materia por reconciliación de datos y detección de Errores Gruesos en Unidades de Procesamiento .....	36
6.1.1 Reconciliación de datos en una Unidad de Destilación Atmosférica .....	38
6.1.2 Reconciliación de datos en una Unidad de Destilación a Vacío.....	41
6.1.3 Reconciliación de datos en una Unidad de Craqueo Catalítico .....	44

<b>VII. ANALISIS COSTO/BENEFICIO</b> .....	47
<b>VIII. CONCLUSIONES</b> .....	50
<b>IX. BIBLIOGRAFÍA</b> .....	51
<b>X. GLOSARIO</b> .....	52
<b>XI. APENDICE</b> .....	53
11.1 Principales compañías licenciadoras de software de reconciliación de datos a nivel mundial.....	53
11.1.1 Pimsoft.....	54
11.1.2 Aspentech.....	55
11.1.3 Soteica.....	56
11.2 Fundamentos de Estadística. ....	57
11.2.1 Conceptos Básicos .....	57
11.2.2 Descripción de Datos .....	58
11.2.3 Datos Agrupados.....	59
11.2.4 Datos no Agrupados.....	73

## I. INTRODUCCIÓN

Partiendo de que todas las mediciones no son más que aproximaciones a la realidad, lo mismo debe de ser cierto para todos los cálculos realizados a partir de ellas, y el mayor reto de todos los programas utilizados hoy en día está en aproximarse tanto como sea posible a la verdad.

La calidad de la data obtenida del proceso afecta significativamente al desempeño y beneficio del uso monitoreo de datos, optimización en línea y control avanzado. Desafortunadamente los datos obtenidos en la planta siempre traen consigo errores que invalidan el modelo utilizado para el proceso de control o de optimización. La reconciliación de data y la detección de errores gruesos son técnicas desarrolladas desde hace 30 años para implementar la exactitud de los datos, de forma que satisfagan el modelo de la planta.

En cualquier planta química, proceso petroquímico, cientos o inclusive miles de variables, tales como caudales, temperaturas, presiones, niveles y composiciones, son medidos rutinariamente y recopilados en los procesos de control, la optimización en línea y la evaluación económica del proceso.

Existen sistemas modernos, los cuales adquieren los datos y procesan un gran volumen de información, frecuentemente tomado con frecuencias del orden de los minutos o inclusive segundos.

Las medidas son inevitablemente dañadas durante la medición, el procesamiento y la transmisión de la señal medida. Siguiendo buenos procedimientos para la instalación y mantenimiento de medidores, es posible asegurar que no habrá errores gruesos en la medida por lo menos por un tiempo. Los errores gruesos causados por descalibración del instrumento pueden ocurrir de repente en un momento en particular y después de ese momento permanecerán con un nivel constante de significancia o magnitud.

Los errores en la data pueden llevar a un deterioro significativo del desempeño de la planta. Los pequeños errores aleatorios y errores gruesos pueden llevar al deterioro del desempeño de los sistemas de control avanzado, mientras que grandes errores gruesos pueden anular las ganancias obtenidas de la optimización del proceso. En algunos casos, data errónea puede llevar al proceso a trabajar en límites antieconómicos, o tal vez peor, a un régimen de operación inseguro.

Se espera que los resultados reconciliados sean más exactos que las mediciones y, más importante que eso, que sean consistentes con las relaciones conocidas entre las variables del proceso, definidas como restricciones. Para que la reconciliación de datos sea efectiva no deben de haber errores gruesos ni en las medidas ni en las restricciones del diagrama de flujo del proceso. Así, la reconciliación de datos y la detección de errores gruesos son aplicadas de manera conjunta para lograr exactitud en la información medida.

La reconciliación de datos y la detección de errores gruesos, ambas logran la reducción de errores tan solo utilizando las propiedades de redundancia entre las medidas. Típicamente, en cualquier proceso las variables están relacionadas unas con otras por restricciones físicas, tales como las leyes de conservación de la materia o energía. Dando un marco a estas restricciones, un número mínimo de grados de libertad es necesario para calcular todos los parámetros del sistema y las variables. Si hay más mediciones que este valor mínimo, entonces existe redundancia en las medidas, que puede ser aprovechada.

## **II. IMPORTANCIA DE LA RECONCILIACIÓN DE DATOS Y LA DETECCIÓN DE ERRORES GRUESOS**

### **2.1 Métodos de acondicionamiento de los datos de proceso**

En cualquier planta química moderna, proceso petroquímico o refinería, cientos o inclusive miles de variables, tales como caudales, temperaturas, presiones, niveles y composiciones, son medidas rutinariamente y recopiladas en los procesos de control, la optimización en línea y la evaluación económica del proceso.

Existen sistemas modernos, los cuales adquieren los datos y procesan un gran volumen de información, frecuentemente tomado con frecuencias del orden de los minutos o inclusive segundos.

La recopilación automática de información no solo permite que la data sea obtenida de manera automática, sino que elimina los errores de la recolección manual. Esto por sí sólo ya aumenta notoriamente la exactitud y validez de los datos obtenidos a partir del proceso. La creciente cantidad de información puede ser tratada luego para aumentar la exactitud y consistencia del proceso mediante una revisión y tratamiento sistemáticos de la información.

Las medidas son inevitablemente dañadas durante la medición, el procesamiento y la transmisión de la señal medida. El error total en una medida es la diferencia entre el valor medido y el valor real, puede ser convenientemente representado como la suma de las contribuciones de dos tipos de errores: Error aleatorio y error grueso.

El error aleatorio es aquel que es inherente a la medida, puede ser por el instrumento o por el procedimiento de medición. Si la medición es repetida bajo las mismas condiciones y con el mismo procedimiento, se puede obtener una medida diferente, dependiendo del error aleatorio. La única forma de caracterizar estos errores es utilizando una distribución de probabilidades.



Estos errores pueden ser causados por diferentes fuentes, tales como: Fluctuaciones en la transmisión de energía, la transmisión de la señal y la conversión de la señal, ruido, filtrado de entradas analógicas, cambios en las condiciones ambientales, etc. Dado que estos errores pueden deberse a diferentes fuentes (sobre las cuales la ingeniería de diseño no tiene control), no pueden ser completamente eliminadas y están siempre presentes en cualquier medida. Usualmente corresponden a componentes de alta frecuencia en la señal medida, y son normalmente de magnitud pequeña excepto en algunos picos ocasionales.

Por otro lado, los errores gruesos son causados por eventos aleatorios, tales como malfuncionamiento de instrumentos (como instalación inadecuada de instrumentos de medición), descalibración, desgaste y corrosión de sensores y depósitos sólidos.

La naturaleza del error aleatorio implica que en todo momento el error tiene una magnitud y signo similar, los que pueden ser desconocidos. Entonces, si la medida es repetida con el mismo instrumento bajo condiciones idénticas, la contribución de un error grueso sistemático al valor medido será siempre la misma.

Siguiendo buenos procedimientos para la instalación y mantenimiento de medidores, es posible asegurar que no habrá errores gruesos en la medida por lo menos por un tiempo. Los errores gruesos causados por descalibración del instrumento pueden ocurrir de repente en un momento en particular y después de ese momento permanecerán con un nivel constante de significancia o magnitud. Otras causas de errores gruesos tales como el desgaste o ensuciamiento de los sensores puede ocurrir gradualmente durante un período de tiempo y por lo tanto la magnitud del error grueso se incrementa lentamente dentro de un período de tiempo relativamente largo. Entonces, los errores gruesos ocurren con menos frecuencia que los errores aleatorios pero sus magnitudes son normalmente mayores que las de los errores aleatorios.

Los errores en la data reconciliada pueden llevar a un deterioro significativo del desempeño de la planta. Los pequeños errores aleatorios y errores gruesos pueden

llevar al deterioro del desempeño de los sistemas de control avanzado, mientras que grandes errores gruesos pueden anular las ganancias obtenidas de la optimización del proceso. En algunos casos, data errónea puede llevar al proceso a trabajar en límites antieconómicos, o tal vez peor, a un régimen de operación inseguro. Es por ello la importancia de reducir los errores gruesos. Muchas técnicas para procesamiento de datos pueden ser utilizadas en conjunto para lograr este objetivo.

Investigación y desarrollo en el área de acondicionamiento de señales ha llevado a la fabricación de filtros analógicos y digitales, los cuales pueden ser usados para atenuar los efectos de las altas frecuencias de ruido en las mediciones. Los errores gruesos pueden ser detectados inicialmente usando pruebas de validación de data. Esto incluye pruebas inclusive cuando la data medida y el promedio en el que cambia están dentro de límites operacionales predefinidos. Hoy en día hay disponibles sensores inteligentes los cuales pueden desempeñar pruebas de diagnóstico para determinar si hay algún problema en los equipos de medición o si el resultado de la medición es aceptable.

Técnicas más sofisticadas incluyen análisis estadísticos de calidad para la información, los que pueden ser usados para detectar errores significativos (afloramientos) en la data del proceso. Estas técnicas son aplicadas a cada variable medida por separado. Así, estos métodos incrementan la precisión de las medidas, dichos métodos no hacen uso del diagrama de flujo del proceso y por lo tanto no aseguran la consistencia de la información con respecto a las relaciones entre las diferentes variables del proceso. Sin embargo, estas técnicas deben de ser usadas como un primer paso para reducir el efecto del error aleatorio en la data y eliminar errores gruesos obvios.

Es posible reducir más adelante los efectos de los errores aleatorios y también eliminar errores gruesos sistemáticos en la información utilizando las relaciones conocidas entre las variables del proceso. Las técnicas de reconciliación de datos y

detección de errores gruesos que han sido desarrolladas en el ámbito de la ingeniería química durante los 30 años pasados es el principal tema a desarrollar.

La reconciliación de datos es una metodología que ha sido desarrollada para incrementar la exactitud de las medidas reduciendo el efecto del error aleatorio en la información. La principal diferencia entre la reconciliación de datos y las técnicas de filtrado de datos es que la reconciliación de datos hace uso explícito de las restricciones del proceso y obtiene estimados de las variables ajustando las mediciones del proceso de forma que las restricciones se cumplan.

Se espera que los resultados reconciliados sean más exactos que las mediciones y, más importante que eso, que sean consistentes con las relaciones conocidas entre las variables del proceso, definidas como restricciones. Para que la reconciliación de datos sea efectiva no deben de haber errores gruesos ni en las medidas ni en las restricciones del diagrama de flujo del proceso. La detección de errores gruesos es una técnica que acompaña a la reconciliación de datos que ha sido desarrollada para identificar y eliminar errores gruesos. Así, la reconciliación de datos y la detección de errores gruesos son aplicadas de manera conjunta para lograr exactitud en la información medida.

La reconciliación de datos y la detección de errores gruesos, ambas logran la reducción de errores tan solo utilizando las propiedades de redundancia entre las medidas. Típicamente, en cualquier proceso las variables están relacionadas unas con otras por restricciones físicas, tales como las leyes de conservación de la materia o energía. Dando un marco a estas restricciones, un número mínimo de grados de libertad es necesario para calcular todos los parámetros del sistema y las variables. Si hay más mediciones que este valor mínimo, entonces existe redundancia en las medidas, que puede ser aprovechada. Este tipo de redundancia frecuentemente es conocido como redundancia espacial y se dice que el sistema de ecuaciones está sobredeterminado.

La reconciliación de datos no puede llevarse a cabo sin redundancia espacial. Sin información extra de las mediciones, el sistema es tan solo medido y no es posible hacer correcciones sobre las medidas erróneas. Peor aún, si menos variables de las necesarias para determinar el sistema son medidas, el sistema es indeterminado y los valores de algunas variables pueden ser estimados solo mediante otras mediciones o si se hacen medidas adicionales.

Un segundo tipo de redundancia que existe en las mediciones es la redundancia temporal. Esta se basa en el hecho de que las mediciones en el proceso son realizadas continuamente en el tiempo con un alto promedio de muestreo, produciendo más medidas de las necesarias para determinar un proceso en estado estacionario. Si se asume que el proceso está en estado estacionario, entonces la redundancia temporal puede ser explotada usando tan solo los promedios de las medidas y aplicando reconciliación en estado estacionario a los valores promedio.

Si el proceso es dinámico, pero sin embargo la solución del estado del proceso es descrita por ecuaciones diferenciales correspondientes a balances de materia y energía, las cuales capturan inherentemente la redundancia en tiempo y en espacio de las variables medidas. Para este tipo de proceso, la reconciliación de datos y las técnicas de detección de errores gruesos han sido desarrolladas para obtener estimados exactos y consistentes con el modelo de ecuaciones diferenciales del proceso.

Procesamiento de señales y técnicas de reconciliación de datos para la reducción de errores pueden ser aplicadas en un proceso industrial como parte de una estrategia integrada referida como acondicionamiento de data o rectificación de data.

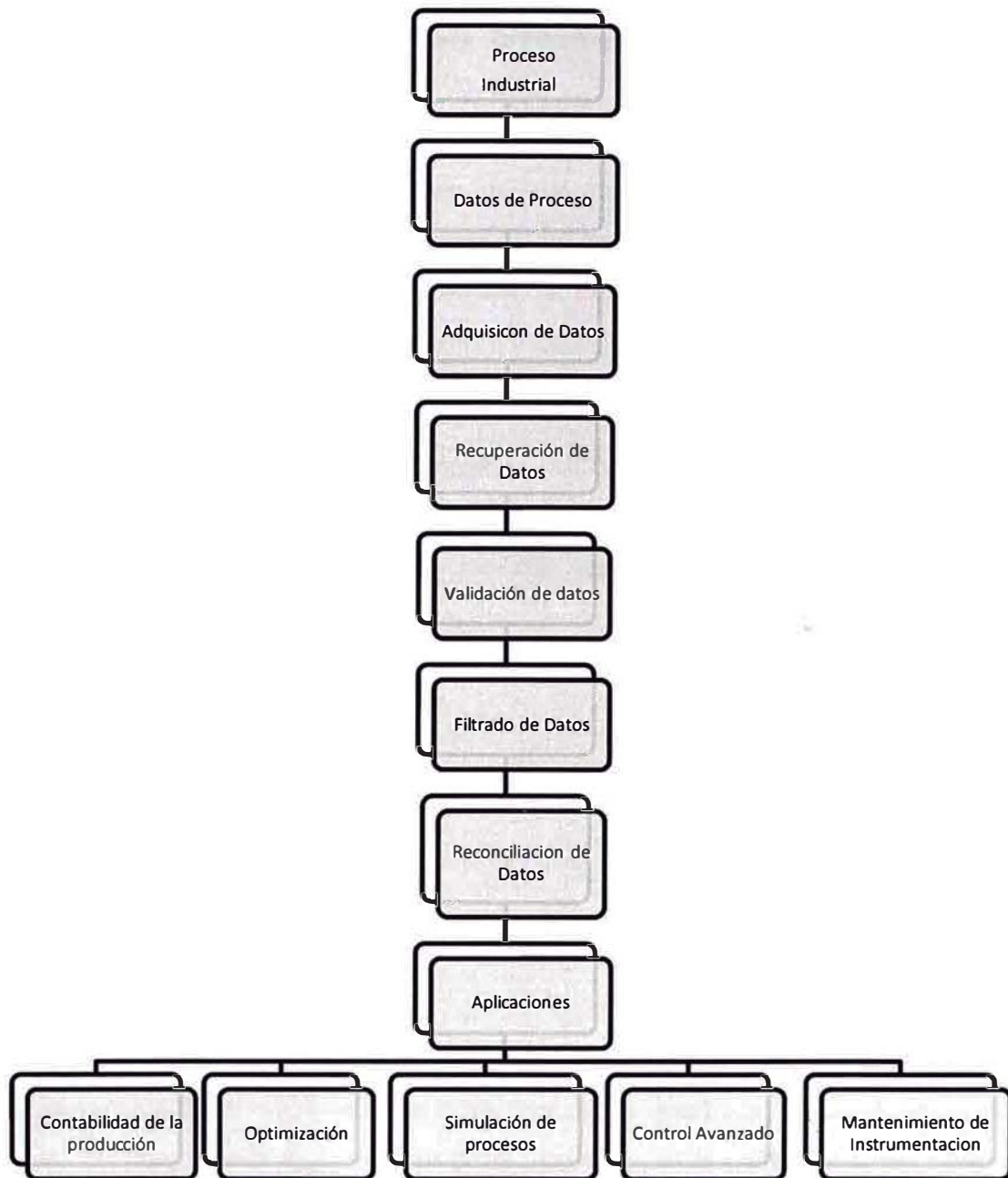
### **2.1.1 Procedimiento del sistema de reconciliación de datos**

En un Proceso Industrial se generan diferentes datos provenientes de las mediciones realizadas por instrumentos de flujo (generalmente placas de orificio) y mediciones de nivel de tanques de almacenamiento, las cuales son adquiridas en campo, en tiempo real por un sistema de transmisión de información de planta, almacenándolas en servidores de alta capacidad, estos datos servirán para realizar la reconciliación de datos y detección de errores gruesos de medición, para lo cual se realiza un balance de materia completo de todo el proceso industrial, con la ayuda de un sofisticado software, que contiene dentro de su lenguaje de programación los algoritmos de reconciliación de datos.

Para realizar el balance de materia, se requiere definir el tiempo o el periodo de tiempo para el cual se realizará el balance, los datos de medición de instrumentos en masa, los que deben ser previamente validados y filtrados, además se debe conocer el porcentaje de tolerancia de error y confiabilidad de todos los instrumentos de medición que generaron los datos de proceso.

El resultado de realizar el balance de materia mediante reconciliación de datos, te generará datos de planta reconciliados mucho más precisos y cercanos al valor real, es decir con un grado de confianza muy elevado, superior al 95 %.

Estos datos reconciliados ya pueden ser utilizados para diferentes fines, por ejemplo, control avanzado, simulación, optimización, contabilidad de la producción, mantenimiento de instrumentación, etc. Ver figura 1.



**Figura 1 - Esquema del sistema de reconciliación de datos**

Fuente: Propia

## 2.2 Formulación del problema de reconciliación de datos.

Como ha sido establecido en las secciones anteriores, la reconciliación de datos incrementa la exactitud de la data del proceso, ajustando los valores medidos de manera que cumplan las restricciones del proceso. La magnitud del ajuste realizado a las medidas es mínimo a partir de que se espera que los errores aleatorios en las medidas sean pequeños. En general, no todas las variables del proceso son medidas por restricciones técnicas y económicas.

Los estimados de variables no medidas así como los parámetros del diagrama de flujo son obtenidos como parte del problema de reconciliación. La estimación de valores no medidos basada en medidas reconciliadas es conocida como *data coaptation*. En general, la reconciliación de datos puede ser formulada por la minimización del siguiente problema de mínimos cuadrados ponderados.

$$\text{Min}_{x_i, u_j} \sum_{i=1} W_i (y_i - x_i)^2 \quad (1)$$

$$\text{Sujeto a } g_k(x_i, u_j) = 0 \quad k = 1; \dots, m \quad (2)$$

La función objetivo (1) define la suma total de los cuadrados de los ajustes hechos a las medidas, donde  $w_i$  es el peso de la medida,  $y_i$  es la medida y  $x_i$  es el estimado reconciliado para la variable  $i$ , y  $u_j$  son las estimaciones de variables no medidas. La ecuación (2) define las restricciones del modelo. Los pesos son escogidos dependiendo de la exactitud de las diferentes medidas.

Las restricciones del modelo son generalmente los balances de materia y energía, pero pueden incluir relaciones de desigualdad impuestas por posibles condiciones de operación del proceso. Las leyes naturales de conservación de materia y energía son usadas típicamente como restricciones en la reconciliación, dado que estas leyes son siempre conocidas. Ecuaciones empíricas u otras, que envuelven muchos parámetros desconocidos no son recomendables, dado que son en el mejor de los casos tan solo aproximaciones. Forzar a las variables medidas a obedecer relaciones inexactas puede causar una solución de la reconciliación de datos inexacta y una detección de errores gruesos incorrecta.

La ley de conservación de la materia o energía puede ser expresada como sigue:

$$\text{Entradas} - \text{Salidas} + \text{Generación} - \text{Consumo} - \text{Acumulación} = 0 \quad (3)$$

EL ámbito de alcance de la ecuación anterior puede ser el flujo total de materia, el flujo de componentes específicos, o el flujo de energía. Si no hay acumulación de ninguno de estas cantidades, entonces estas restricciones son ecuaciones algebraicas lineales y definen una operación en estado estacionario.

Para procesos dinámicos, sin embargo, los términos de acumulación no pueden ser descuidados y las restricciones son ecuaciones diferenciales. Para la mayoría de unidades de proceso, no hay generación o consumo de material. En el caso de reactores, mediante la generación o agotamiento de los componentes debe de ser considerado.



Para algunas unidades simples, tales como divisores de corriente, no hay cambio ni en la temperatura ni en la composición de las corrientes. Para este tipo de unidades, el balance de componentes y energía se reduce a la forma simple siguiente:

$$x_i = y_i \quad (4)$$

Donde la variable  $x_i$  representa temperatura o composición de la corriente  $i$ . La ecuación citada es útil cuando dos o más sensores son utilizados para medir la misma variable, ya sea flujo o temperatura de la corriente.

El tipo de restricciones que son utilizadas en la reconciliación de datos depende del alcance del modelo de reconciliación, la complejidad de las técnicas de resolución utilizadas depende fuertemente de las restricciones utilizadas. Por ejemplo. Si se está interesado en reconciliar sólo los flujos de las corrientes, entonces las restricciones de balance de materia son lineales para las variables de flujo y para la reconciliación de datos. Por otro lado, si queremos reconciliar medidas de presión, temperatura o composición junto con los flujos, entonces estamos frente a un problema de reconciliación no lineal.

Un punto para ser recordado es el tipo de restricciones que podemos imponer legítimamente en la aplicación de reconciliación de datos. Dado que la reconciliación de datos fuerza a los estimados de las variables a satisfacer las restricciones impuestas, este punto gana importancia. Siempre, las restricciones de conservación de materia y energía son incluidas porque son leyes físicas válidas. Debe notarse que sin embargo esas ecuaciones son generalmente escritas asumiendo que no hay pérdida de materia o energía desde el proceso hacia el entorno. Mientras esto puede ser válido para flujos de materia, pueden ocurrir pérdidas significativas de energía por ejemplo por un aislamiento inapropiado en intercambiadores de calor. En estos casos, es mejor no imponer un balance de energía o incluir alternativamente una pérdida indeterminada en la ecuación de balance que pueda ser estimada como parte del proceso de reconciliación.

Otras restricciones aparte de las de conservación de materia y energía, pueden ser ecuaciones que relacionen parámetros de la unidad de proceso. Por ejemplo, el modelo de un intercambiador de calor puede incluir una ecuación que relaciona el poder calorífico con el coeficiente de transmisión de calor, el área de transferencia de calor y las temperaturas de los flujos. La ecuación 5 describe dicha ecuación.

$$Q - UA(LMTD) = 0 \quad (5)$$

Donde  $Q$  es el flujo calorífico,  $U$  es coeficiente global de transferencia de calor,  $A$  es el área de transferencia de calor, y  $LMTD$  es la diferencia media logarítmica de temperatura.

Generalmente, dado que el coeficiente global de transferencia de calor es desconocido y no tiene forma de ser determinado entre la data medida, esta ecuación puede ser incluida y  $U$  puede ser estimado como parte del problema de reconciliación. Si no hay información primaria de  $U$ , sin embargo, y no hay restricciones viables, entonces la inclusión de esta restricción no provee ninguna información adicional y se estimarían todas las otras variables de la misma forma si se incluye o no la restricción. Así, el problema de reconciliación de datos puede ser resuelto igual de bien sin incluir la esta restricción, la constante  $U$  puede ser estimada posteriormente por la ecuación 5 usando los valores reconciliados de flujos y temperaturas.

Por otro lado si  $U$  tiene que ser determinado dentro de límites específicos, o si hay un estimado de  $U$  proveniente de un proceso previo de reconciliación, entonces la restricción puede ser incluida con la información de  $U$  como parte de del problema de reconciliación. El coeficiente global de transferencia de calor puede ser relacionado con las propiedades de las corrientes, sus flujos, temperaturas, y las características del

intercambiador de calor utilizando correlaciones. No es recomendable usar una ecuación en el modelo de reconciliación si la correlación por sí misma puede tener errores y forzar a los flujos y las temperaturas a cumplir la ecuación incrementando la inexactitud de los estimados.

Otra pregunta importante es donde aplicar reconciliación de datos en estado estacionario y donde en estado dinámico. De modo práctico, un proceso nunca está en estado totalmente estacionario. Sin embargo una planta es operada por varias horas o días en relación a un régimen estacionario nominal. Para aplicaciones tales como optimización en línea donde la reconciliación es realizada una vez dentro de un periodo de pocas horas, es apropiado emplear reconciliación en estado estacionario sobre valores promedio de las mediciones dentro del periodo de tiempo de interés.

Durante condiciones transitorias (por ejemplo el cambio de tipo de crudo procesado en una refinería) cuando el alejamiento del estado estacionario es significativo, la reconciliación en estado estacionario no debe ser aplicada porque esta resultaría en ajustes significativos a los valores medidos. Las medidas tomadas durante periodos de transición pueden ser reconciliadas, si es necesario, utilizando un modelo dinámico del proceso. De la misma manera que en aplicaciones de control de procesos donde la reconciliación tiene que ser realizada cada pocos minutos debe realizarse la reconciliación dinámica.

La reconciliación de datos está basada en la asunción de que en las medidas solo está presente el error aleatorio, el cual sigue una curva normal (campana de Gauss). Si está presente un error grueso dentro de una medida, se presentan polarizaciones en algunas mediciones, o si una pérdida significativa que no ha sido contabilizada está presente, entonces la reconciliación de datos no sería precisa. Es por eso que es necesario identificar y eliminar los errores gruesos. Esto es conocido como el problema de detección de errores gruesos.

Aplicaciones tales como simulación y optimización en equipos de proceso existentes permiten crear un modelo del equipo. Esos modelos siempre contienen parámetros que deben de ser estimados con datos de la planta. Esto es conocido como afinamiento del modelo, para lo cual es esencial contar con datos precisos.

El uso de mediciones erróneas en el afinamiento del modelo puede llevar a un aumento en los parámetros incorrectos en el modelamiento los cuales pueden anular todos los beneficios de la optimización.

Hay dos posibles formas en que la reconciliación de datos puede ser aplicada de esta manera las cuales se ilustran usando un ejemplo simple.

Consideremos el problema de optimizar el desempeño de una columna de destilación existente. De los datos de operación, las mediciones de flujos, temperaturas y composiciones de las corrientes de entrada y salida de la columna pueden ser obtenidas. La data reconciliada puede ser usada junto con un modelo plato por plato de la columna con el objetivo de estimar eficiencias de plato. El modelo afinado puede ser usado luego para optimizar el funcionamiento de la columna.

Alternativamente, puede realizarse una parametrización y reconciliación de datos en simultáneo usando el modelo plato por plato de la columna. En ese caso, si las medidas de temperatura en los platos y/o las composiciones están disponibles, también pueden ser usadas y reconciliadas como parte del problema. Obviamente, la segunda opción conlleva a un aumento significativo en el tiempo de cálculo. Esta opción está referida como un modelamiento riguroso en línea y ha sido incorporada en muchos simuladores comerciales de estado estacionario.

La reconciliación de datos puede ser muy útil en la programación de mantenimiento a los equipos de proceso. La data reconciliada puede ser usada para estimar exactamente parámetros clave en el desempeño de los equipos de proceso. Por ejemplo, el coeficiente de transferencia de calor para intercambiadores de calor o el nivel de actividad del catalizador en un reactor pueden ser estimados y usados para

determinar si es que el intercambiador debe de ser limpiado o si el catalizador debe de ser remplazado o regenerado, respectivamente.

Muchas estrategias avanzadas en el control de procesos tales como controladores basados en el modelo o control deductivo requieren estimados exactos de las variables controladas. Las técnicas de reconciliación dinámica de datos pueden ser usadas para derivar estimados exactos para mejores procesos de control.

La detección de errores gruesos no solo aumenta la exactitud de los procesos de reconciliación sino que también identifica problemas en instrumentos los cuales requieren mantenimiento específico y correcciones. La detección de errores grueso incipiente puede reducir costos de mantenimiento y hacer que la operación de la planta sea más fácil. Estos métodos pueden ser ampliados para detectar fallas en los equipos.

### III. ERRORES DE MEDICIÓN

#### 3.1 Clasificación de los errores de medición

Como se mencionó en el capítulo II, hay varias fuentes de error en los instrumentos de medición de campo, debido a la toma de datos en forma manual, el tener múltiples valores para medidas únicas, tener tags incorrectamente asociados a medidas, suposiciones incorrectas de ingeniería y falta de sincronización entre medidas (medidas de los instrumentos en tiempos diferenciados), los cuales determinan errores de medición en virtualmente todos los datos medidos de proceso, algunos de los errores de medición son pequeños y aleatorios (errores aleatorios) y otros son grandes y sistemáticos (errores gruesos), algunos autores prefieren llamarlos errores sistemáticos.

El error total (Et) de una medida es la diferencia entre el valor real y el valor medido

$$Et = \text{Valor real} - \text{Valor medido} \quad (6)$$

La cual puede ser representado por la suma de los dos tipos de errores ya mencionados:

$$Et = \text{Errores aleatorios} + \text{Errores gruesos} \quad (7)$$

### 3.1.1 Errores aleatorios

Si repetimos exactamente las mismas condiciones de operación para un proceso determinado, se observará que para un mismo flujo, el instrumento de medición tendrá diferentes valores de error. Esto es debido a la presencia de errores aleatorios en las mediciones. Los errores aleatorios no pueden ser predichos ni explicados con precisión. No es posible predecir el signo y la magnitud de los mismos y son causados por distintas razones, por ejemplo los cambios en las condiciones ambientales, fluctuaciones en la energía, transmisión de la señal, conversión de la señal, ruido, etc.

Los errores aleatorios no pueden ser eliminados y siempre están en las medidas, estos errores pueden ser caracterizados por una función de probabilidad.

La relación entre el valor medido, el valor real y el error aleatorio en la medición de una variable  $i$  esta expresado por la ecuación (8).

$$y = x + \varepsilon \quad (8)$$

Donde  $y$  es el valor de la medida,  $x$  es el valor real y  $\varepsilon$  es el error aleatorio, el error aleatorio usualmente oscila alrededor del valor 0.

### 3.1.2 Errores gruesos

Generalmente los errores gruesos o también llamados sistemáticos, son usualmente asociados a las fallas de los sensores de medición (Ver figura 2), donde se ilustra gráficamente los tipos más comunes de fallas del instrumento: Bias (sesgo), Complete failure (falla completa), drifting (deriva), y la degradación de la precisión.

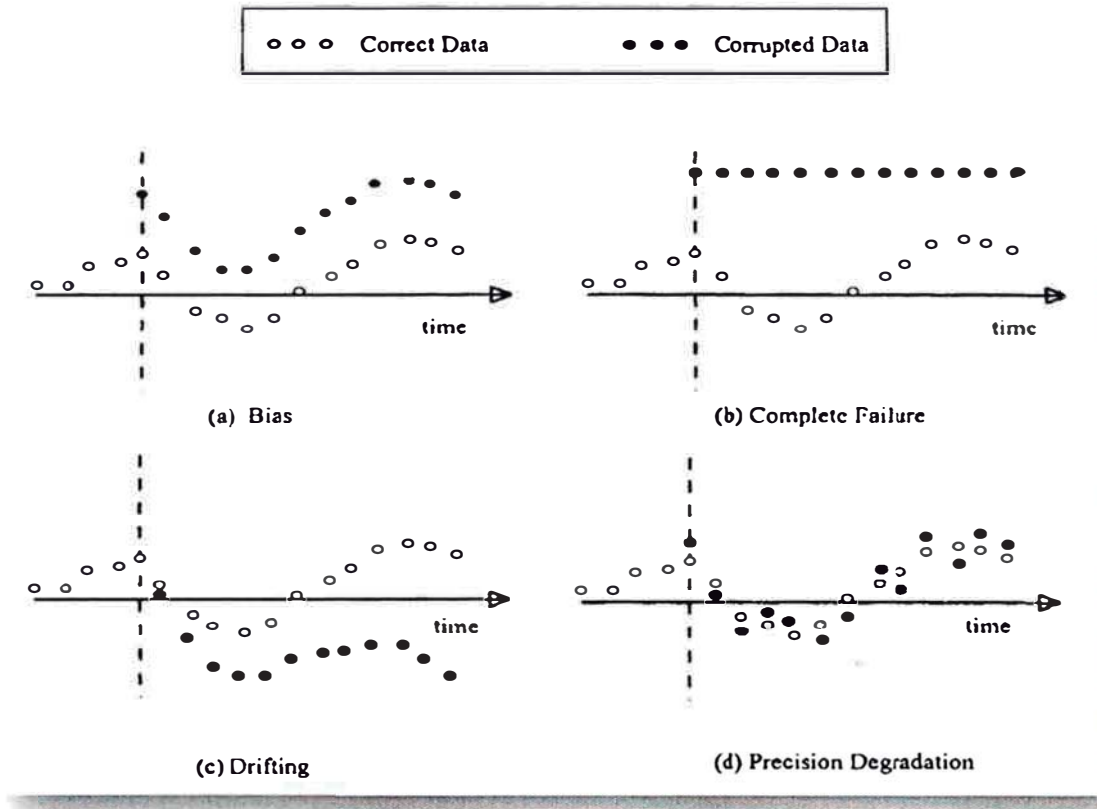
Si un error grueso existe en un valor medido, la ecuación de medición (8) se cambia a la (9) de la siguiente manera:

$$y = x + \varepsilon + \delta \quad (9)$$

Donde  $\delta$  es la magnitud del error grueso, se debe tener en cuenta que las fugas del proceso, que también se clasifican como errores gruesos no se pueden modelar con la ecuación (9).

Los errores gruesos afectan significativamente la precisión datos de proceso en la industria. Tienen que ser detectados y eliminados. Algunos de ellos, tales como valores atípicos ocasionales (picos), se pueden detectar mediante el uso de técnicas de filtrado o control de calidad estadístico (también conocido como estadística de control de procesos). Otros tipos pueden ser más difíciles de detectar sin un modelo físico por lo cual la reconciliación de datos es la herramienta adecuada en la mayoría de los casos.





**Figura 2: Fallas de sensores de medición**

Fuente: Narasimhan S., Jordache C. "Data Reconciliation & Gross Error Detection"

Editorial Gulf Publishing Company, (2000)

## IV. RECONCILIACIÓN DE DATOS Y LA ESTADÍSTICA

### 4.1 Tolerancias e intervalo de confianza

Como ya se ha mencionado no existen mediciones perfectas. Las mediciones siguen una distribución gaussiana alrededor de una media (campana de Gauss).

Estadísticamente sabemos que el 95% de las lecturas de las mediciones de instrumentos van a estar dentro de 2 desvíos estándar de la media (desviaciones estándar).

Por ejemplo, si el valor medido es 100 toneladas y la tolerancia es de 2% en el instrumento, entonces dos desvíos estándar son iguales a 2% de la lectura. El valor verdadero debería estar entre 98 y 102 toneladas el 95% de las veces. El intervalo de confianza del medidor es de 95%. Ver Figura 3

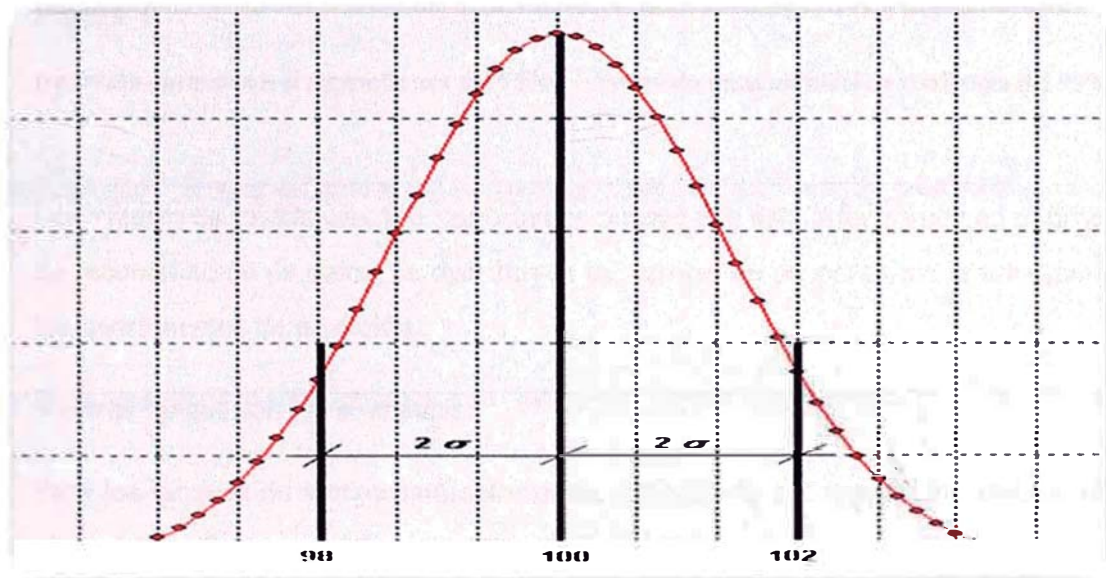
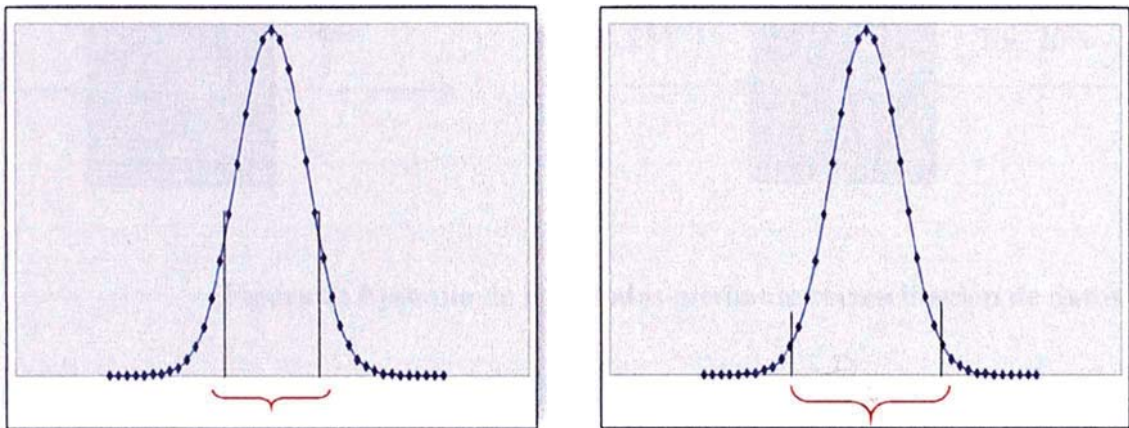


Figura 3: Campana de gauss

Fuente: Propia

Si el número de  $\sigma$  se fija en 3 toneladas, esto definiría un intervalo de confianza del 99 %, lo cual significa que van a ser reportados menos errores, pues ahora las lecturas “buenas” pueden caer en un rango más amplio. Ver figura 4.



**Figura 4:**

**Intervalo para un nivel de confianza del 95%**      **Intervalo para un nivel de confianza del 99%**

Los valores de tolerancias son importantes porque con esta información en el proceso de reconciliación de datos, se distribuyen los errores en proporción a la tolerancia de los instrumentos de medición.

Valores típicos son los siguientes:

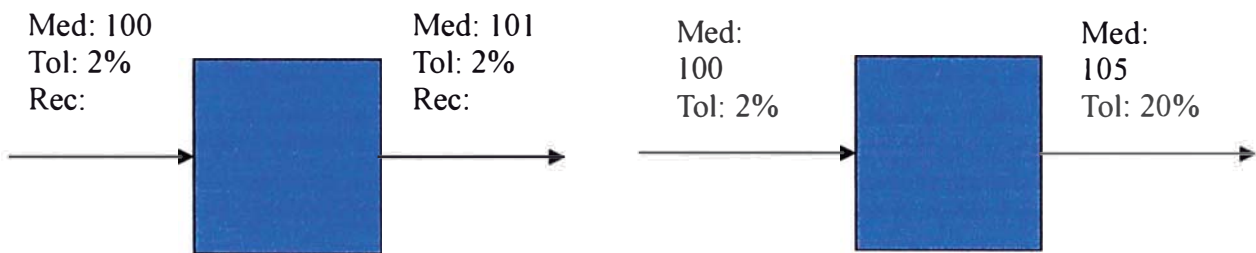
Para los tanques de almacenamiento: masa equivalente a 2 mm de incertidumbre en altura (valor fijo).

Instrumentos de medición de transferencia de custodia: 0.5%

Instrumentos de medición de datos de proceso comunes: 2%

Valores estimados (flujos que no tienen instrumento de medición real en planta): 10%

Tras la reconciliación de los datos, las correcciones de la medición son mayores en los instrumentos de medición con mayores tolerancias.



**Figura 5: Ejemplo de resultados mediante reconciliación de datos**

Donde:

Med: Medida de instrumento

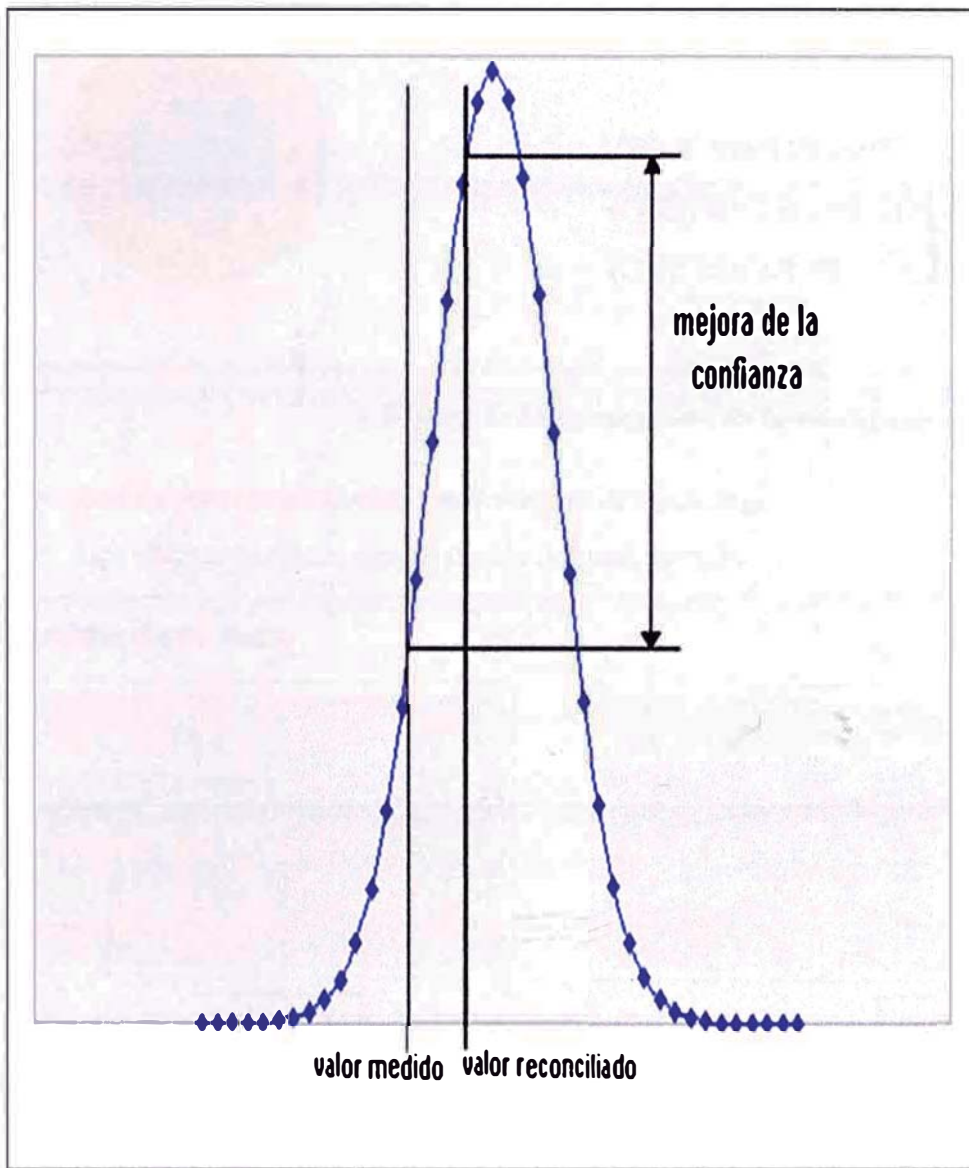
Tol: tolerancia del instrumento de medición de planta

Rec: valor ajustado, luego del proceso de Reconciliación de Datos

Se puede inferir que los valores reconciliados son mejores, es decir más cercanos al valor real que los valores medidos, pero esto se cumple si y solo si las medidas tomadas de los instrumentos no tienen errores gruesos.

Además se debe considerar que todas las tolerancias deben ser fijadas correctamente antes del proceso de reconciliación de datos, entonces los valores reconciliados son demostrablemente mejores que los medidos.

Pero cabe resaltar que esto cumple para los medidores con las mayores tolerancias en planta de proceso. Para los medidores más precisos no hay mejoras sustanciales en la medición. Ver Figura 6 y 7.



**Figura 6: Mejoramiento de la confianza de los datos**

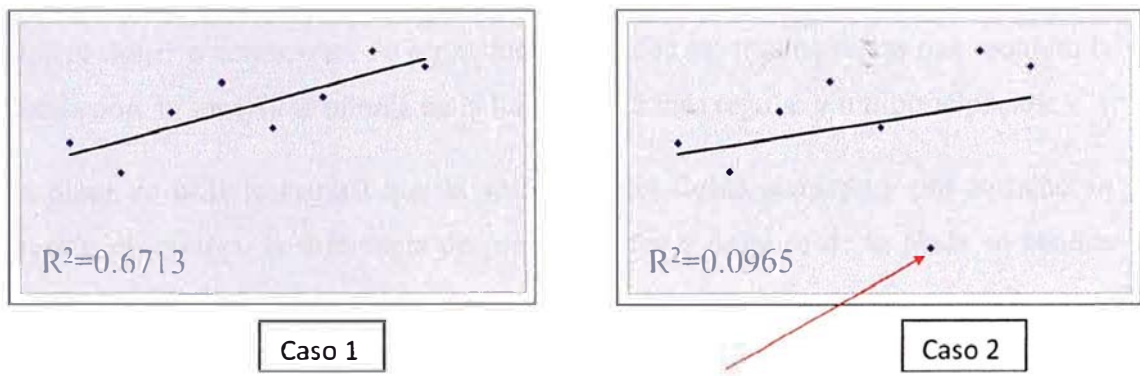
Fuente: Propia



**Figura 7: Mejoramiento de la confianza de los datos**

- Los valores reconciliados caerán dentro del área azul.
- Los valores medidos caerán dentro del área naranja.

#### 4.2 Influencia de datos



**Figura 8: Mejoramiento de la confianza de los datos**

La influencia mide el efecto de una medición sobre el ajuste realizado al reconciliar los datos.

La reconciliación puede verse afectada por un problema de regresión. Lo que se aprecia en el primer caso, es que los puntos están más o menos distribuidos a lo largo de una recta y la distancia entre el punto y la recta es aproximadamente igual para los puntos y van a tener influencias más o menos similares.

En el caso del segundo caso, el punto más alejado es el que va a tener mayor efecto sobre el ajuste y mayor influencia.

## V. INSTRUMENTOS DE MEDICION DE FLUJO Y NIVEL

### 5.1 Medidor de flujo, placa de orificio

Las Placas de orificio son elementos de medición de flujo aplicables para gases, líquidos y vapores limpios. Actualmente se estima que al menos un 75% de los medidores de flujo industriales son dispositivos de presión diferencial, siendo el más popular la placa de orificio. La placa de orificio consiste en una placa perforada que se instala en la tubería, el caudal se puede determinar por medio de las lecturas de la caída de presión diferenciales.

La línea de instalación debe ser de sección circular y de diámetro constante, sin obstrucciones o conexiones de otras líneas en todos los tramos rectos que requiera la instalación, la superficie interna de la línea será lo más regular y uniforme posible.

La placa de orificio origina que la velocidad del fluido aumente y por lo tanto su presión disminuya, la diferencia de presión, antes y después de la placa se conoce como presión diferencial.

Conociendo la presión diferencial, las características geométricas y la densidad del fluido se puede calcular el flujo volumétrico siguiendo la siguiente ecuación:

$$Q = \frac{C_d}{\sqrt{1-\beta^4}} \frac{\pi}{4} d^2 \sqrt{\frac{2h}{\rho}} \quad (10)$$

Donde:

$C_d$  es el coeficiente de descarga, el cual es un valor de ajuste que compensa la distribución de velocidad y las pérdidas de carga menores que no fueron tomadas en cuenta al obtener la ecuación de flujo volumétrico. Ver Figura 9.

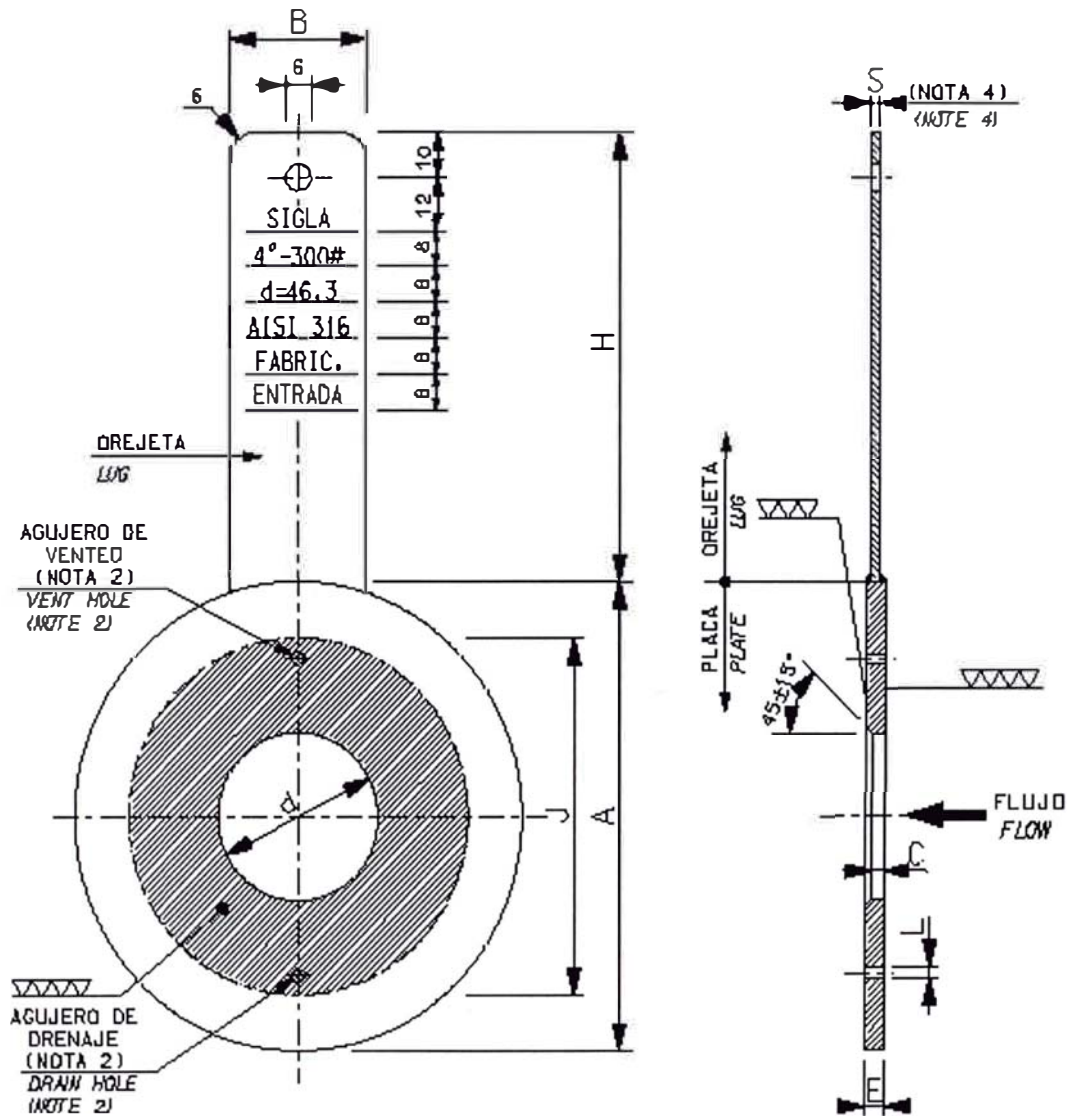


Figura 9: Medidor de placa de orificio

Fuente: Propia



## **5.2 Medidor de nivel por radar en tanques de almacenamiento**

Los sensores de radar miden distancias de forma continua y sin contacto y requieren de poco mantenimiento. Asimismo son prácticamente insensibles a la atmósfera del proceso (vapor, presión, polvo o temperaturas extremas).

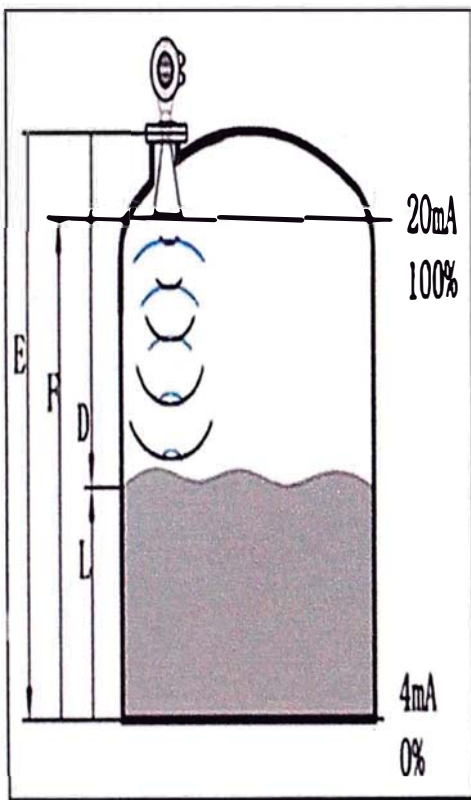
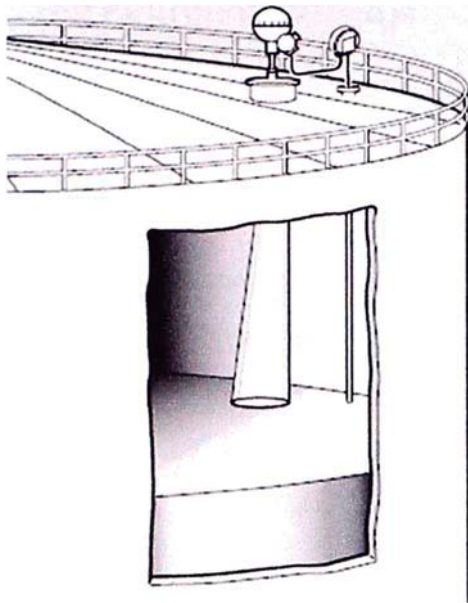
Los instrumentos de radar miden el tiempo de recorrido de un impulso de radar desde su emisión hasta su recepción. Este tiempo de recorrido es proporcional a la distancia y por tanto al nivel del producto.

Las señales de radar no están ligadas por un medio que las transmita, y se propagan a la velocidad de la luz (300 000 000 m/s). Los sensores de radar emiten generalmente señales de 6 a 26 GHz.

El transmisor emite señales de radar en forma de impulsos de microondas reflejados por dos materiales con constantes dieléctricas diferentes (el entorno del sensor y el producto).

Un receptor interno recoge los impulsos de radar reflejados (eco) y utiliza el tiempo de recorrido del impulso (emisión / recepción) para calcular la distancia y por tanto el nivel.

La exactitud mínima de la medida de nivel será menor o igual a +/- 1 mm y repetibilidad de +/- 0,5 mm.



Donde  $D = \frac{v}{2dt}$  y  $v = \frac{C}{\sqrt{e}}$

**D = Distancia del emisor al líquido**

**v = Señal de velocidad**

**dt = Tiempo de recorrido**

**C = Velocidad de la luz**

**e = Constante dieléctrica**

**Figura 10: Medidor de nivel tipo radar.**

Fuente: Propia

## **VI. APLICACIONES DE RECONCILIACIÓN DE DATOS Y DETECCIÓN DE ERRORES GRUESOS**

La tecnología de reconciliación de datos es ampliamente aplicada en la actualidad en química, petroquímica y otras industrias de procesamiento de materiales. Es aplicada fuera de línea, tales como la contabilidad de la producción, control de mermas, programas de mantenimiento o en conexión con aplicaciones en línea, tales como el proceso de optimización y control de procesos avanzados.

En este capítulo se presenta la principal aplicación industrial de la reconciliación de datos y la detección de errores gruesos. Sólo se describen las características generales, sin entrar en detalles acerca de la solución particular técnica o el software utilizado.

Otras implementaciones industriales y de software para datos que necesitan de reconciliación no han sido divulgadas públicamente o la fuente de la información publicada no es fácilmente accesible.

El análisis de este capítulo se centra en los balances de materia por reconciliación de datos y detección de errores gruesos en unidades de procesamiento y su uso para la contabilidad de la producción, control de mermas (ímbalances o variaciones) y mantenimiento de instrumentos de medición.

### **6.1 Balance de materia por reconciliación de datos y detección de errores gruesos en unidades de procesamiento**

La reconciliación de unidades de proceso (Balance de materia y energía para unidades de procesamiento) fue el primer tipo de aplicación de reconciliación de datos y detección de errores gruesos. El interés en la aplicación de métodos de conciliación de datos industriales comenzó en la década de 1980, cuando la gestión de la planta se percató de los beneficios de usar un sistema de reconciliación de datos, por lo cual un software para la reconciliación de datos y la detección de errores gruesos llegó a estar disponible.

Anteriormente, las operaciones manuales para la reconciliación de datos de producción solía ser realizado por ingenieros de la planta para fines contables.

Se utilizaban datos no acondicionados para el modelado de procesos, simulación y optimización.

El filtrado de datos se utilizó principalmente para aplicaciones de control de procesos. El rendimiento de todos estos sistemas que utilizan datos de proceso era pobre, ya que los datos de entrada venían acompañados de múltiples errores.

En la actualidad, los sistemas integrados que se usan para simulación de procesos, optimización y fiscalización incluyen un sistema de reconciliación de datos, que precede a todas las aplicaciones que hacen uso de los datos de proceso.

El uso de la reconciliación de datos, hace posible que se las medidas estén libres de errores gruesos y satisfagan los balances de materia y de energía, mejorando notablemente la precisión de los modelos de procesos utilizados para la optimización y aumenta la calidad de las operaciones de control.

Existen muchas implementaciones de software de reconciliación de datos en las refinerías y plantas petroquímicas. Las más simples son para aplicaciones industriales en torno a unidades individuales, especialmente para la destilación o columnas de separación, tales como unidades de recuperación de hidrocarburos aromáticos, craqueo de naftas, y el crudo de petróleo (Destilación Atmosférica, Destilación al Vacío, Craqueo Catalítico, etc.).

**Datos necesarios para la reconciliación de datos en una planta industrial: En la tabla 1.**

<b>Data</b>	<b>Descripción</b>
<b>Configuración de la Planta</b>	Unidades de Proceso, tanques y sus líneas de conexión
<b>Inventario de Tanques</b>	Inventario de cada tanque en la Planta
<b>Movimientos (De unidad a unidad/tanque) y Transferencias (De tanque a tanque)</b>	Data de Movimientos: tiempo del inicio y fin, fuentes y destinos y la cantidad transferida
<b>Recibos</b>	Data de recibos: tiempo del inicio y fin, fuentes y destino y la cantidad recibida.
<b>Despachos</b>	Data de Despacho: tiempo del inicio y fin, fuentes y destino y la cantidad recibida.
<b>Medidores/Factores de precisión</b>	Factores de precisión de instrumentos de medición de flujo (confiabilidad)
<b>Laboratorio</b>	Resultados de pruebas de laboratorio (Densidad y composición)

**Tabla 1**

Fuente: Propia

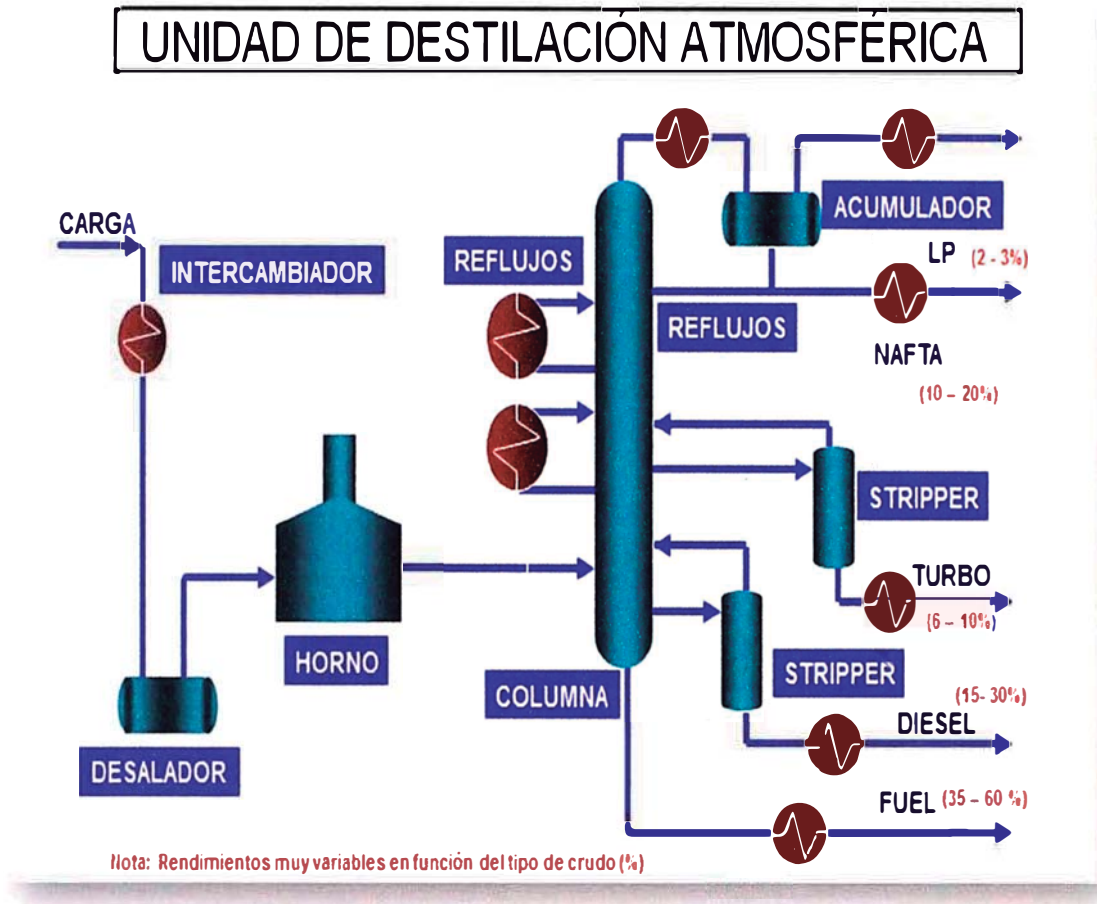
### **6.1.1 Reconciliación de datos en una unidad de destilación atmosférica**

La destilación atmosférica es un proceso que permite el fraccionamiento natural del crudo sin afectar sus estructuras moleculares; basándose en las diferentes temperaturas de ebullición de sus componentes, con la finalidad de obtener productos finales e intermedios.

El proceso inicia con el precalentamiento de crudo en una red de intercambio de calor con productos terminados y reflujos fríos hasta una temperatura de 120°C aproximadamente, temperatura óptima para el inicio del proceso de desalado, cuyo propósito es eliminar sales e impurezas que podrían ocasionar corrosión en equipos y tuberías dentro de la unidad.

Luego del desalado, el crudo es derivado hacia una segunda red de intercambio de calor para continuar con el precalentamiento hasta una temperatura de 240°C aproximadamente (dicha temperatura varía según la configuración de la red de intercambio de calor) y de esta manera hacer su ingreso al horno de la unidad, cuya función primordial será elevar la temperatura del crudo hasta 360°C-370°C, antes del umbral de descomposición térmica.

El horno transfiere la energía térmica necesaria para producir el cambio de fase del crudo y en la zona de carga de la columna fraccionadora se disminuye la presión del sistema para producir el flash de la carga. Los vapores ascienden a través de la columna fraccionadora donde ocurre el equilibrio de fases (líquido-vapor), la fase líquida se logra con los hidrocarburos que regresan a la columna en forma de reflujos por el tope de la columna fraccionadora se obtiene gas combustible, GLP y gasolinas, los cortes laterales están conformados por las naftas, turbo, diesel, gasóleos y por los fondos se obtiene el crudo reducido que es enviado como carga a la columna de destilación a Vacío para continuar con su fraccionamiento. Ver figura 11.



**Figura 11: Diagrama de la unidad destilación atmosférica**

Fuente: Propia

A continuación se muestra un balance másico y volumétrico en la unidad de destilación primaria para una refinería de petróleo, donde se podrá apreciar la comparación entre los valores obtenidos a partir de datos de producción medidos por los instrumentos de flujo y los datos obtenidos después de realizar la reconciliación de datos.

Después de la reconciliación de datos se reducen los imbalances entre la entradas y salidas en proceso, en unidades de masa el imbalance siempre debe ser cero, por

principio de conservación de la materia, la cual maneja la reconciliación de datos, mientras que en volumen depende de la calidad de los flujos reconciliados.

Se puede apreciar que los medidores de flujos de producción de diesel (Prod-Diesel liviano) y de residual (Prod-Residual), tienen un error de medición alto, siendo necesaria sus intervenciones para calibración.

Se aprecia la disminución del imbalance de 1168 a 61 barriles, es decir de 1.93% a 0.1 % en volumen. Ver Figura 12.

Unidad de Destilación Atmosférica	Masa, toneladas		Volumen, barriles	
	Medido	Reconciliado	Medido	Reconciliado
Intakes				
Carga-UDP	8421	8421	60437	60437
Total In	8421	8421	60437	60437
Outturns				
Prod-Gas UDP	24	24	160	160
Prod-Nafta Liviana	944	1023	8195	8878
Prod-Nafta Pesada	222	224	1816	1834
Prod-Turbo	912	740	6800	6245
Prod-Diesel liviano	1910	2233	14017	17034
Prod-Residual	4283	4177	28281	26225
Total Out	8295	8421	59269	60376
Imbalance (In-Out)	126	0	1168	61
	1.49%	0.00%	1.93%	0.10%

**Figura 12: Balance de una unidad de destilación atmosférica, mediante reconciliación de datos.**

Fuente: Propia



### **6.1.2. Reconciliación de datos en una unidad de destilación a vacío**

El crudo reducido proveniente de la unidad de destilación primaria se procesa en la unidad de destilación a vacío con la finalidad de destilar componentes que a presión atmosférica se craquearían, ya que se tendría que elevar tanto la temperatura a la salida del horno que se superaría el umbral de descomposición térmica.

La energía térmica necesaria para vaporizar el crudo reducido es suministrada en los hornos de la unidad que trabajan en un rango de temperaturas de salida de 395 a 405°C, dicha temperatura será función de la cantidad y calidad de los productos que se deseen obtener.

La columna trabaja a vacío con una presión absoluta de unos 20 mmHg en la zona de carga y está formada por empaques que permiten el contacto líquido-vapor, favoreciendo así la transferencia de masa.

Los productos que se obtienen son: gasóleo ligero de vacío que es derivado al pool de diesel, gasóleo pesado de vacío que es derivado como carga a las unidades de craqueo catalítico y residual de vacío que es derivado como carga a la unidad de visbreaking o hacia tanques de asfaltos (según su calidad). Ver figura 13.

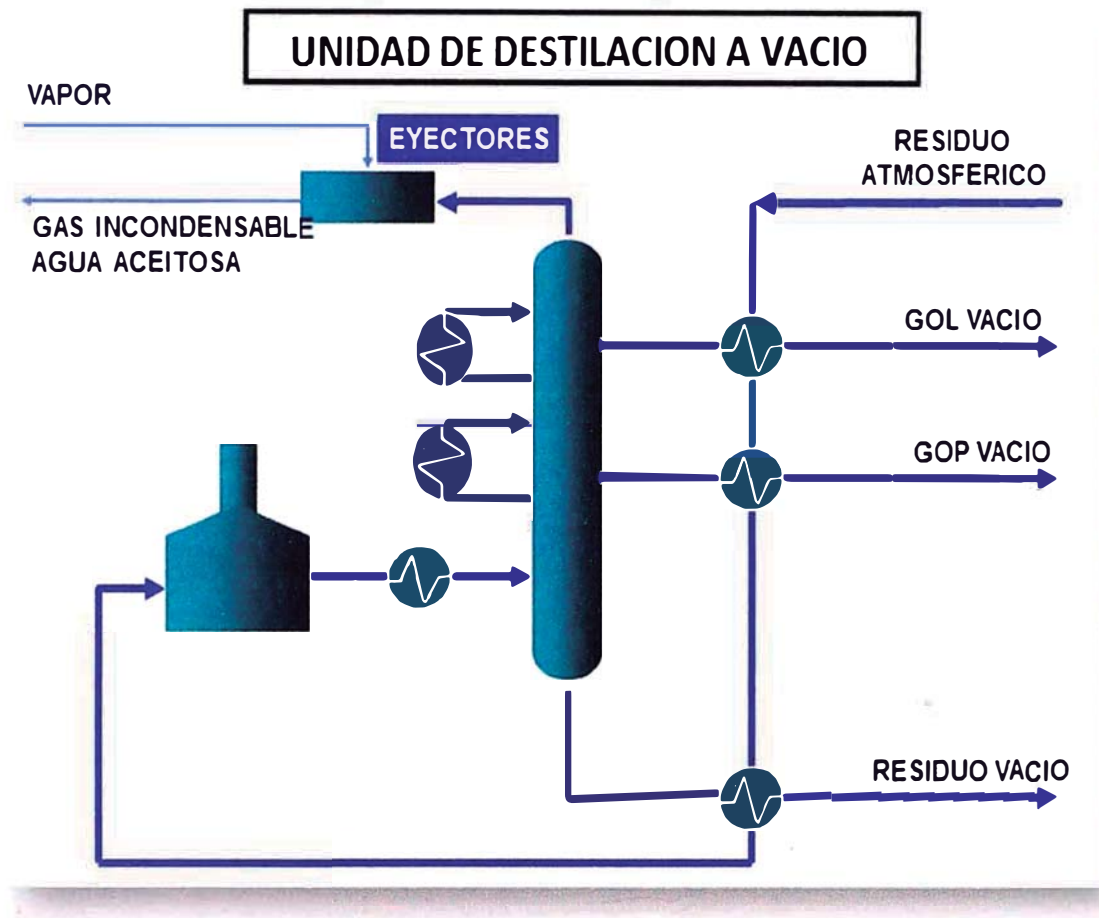


Figura 13: Diagrama de la unidad destilación al vacío

A continuación se muestra un balance másico y volumétrico en la unidad de destilación al vacío para una refinería de petróleo, donde se podrá apreciar la comparación entre los valores obtenidos a partir de datos medidos por los instrumentos de flujo y los datos obtenidos después de realizar la reconciliación de datos.

Después de la reconciliación de datos se reducen los imbalances entre la entradas y salidas en proceso, en unidades de masa el imbalance siempre debe ser cero, por

principio de conservación de la materia, la cual maneja la reconciliación de datos, mientras que en volumen depende de la calidad de los flujos reconciliados.

Se puede apreciar que el medidor de flujo de producción de residuos de vacío (Prod-FC, tiene un error de medición alto, siendo necesaria su intervención para calibración.

Se aprecia la disminución del imbalance de 235 a 229 barriles, es decir de 0.80 % a 0.75 % en volumen. Ver Figura 14.

Unidad de Destilación al Vacío	Masa, toneladas		Volumen, barriles	
	Medido	Reconciliado	Medido	Reconciliado
Intakes				
Carga2-FL	4448	4651	29371	30708
Total In	4448	4651	29371	30708
Outturns				
GAS-UDV1	10	10	59	59
Prod-GL	573	587	4115	4218
Prod-GP	1789	1865	12626	12851
Prod-SW	127	148	960	961
Prod-FC	1876	2041	11376	12391
Total Out	4375	4651	29136	30479
Imbalance (In-Out)	73	0	235	229
	1.64%	0.00%	0.80%	0.75%

**Figura 14: Balance de una unidad de destilación vacío, mediante reconciliación de datos.**

Fuente: Propia

### 6.1.3 Reconciliación de datos en una unidad de craqueo catalítico

Para la conversión del gasóleo pesado de vacío en destilados ligeros se utiliza el proceso denominado craqueo catalítico en lecho fluido (FCC). En este proceso la rotura de las moléculas de hidrocarburos componentes del gasóleo pesado de vacío da lugar a moléculas de menor tamaño en el rango de destilación de las gasolinas y del gasóleo ligero, se produce además GLP y un residuo pesado que se utiliza en la formulación del fuel oil y por último coque que se utiliza como fuente de energía para el propio proceso.

La unidad puede dividirse en tres secciones bien diferenciadas.

- 1.-Reactor/Regenerador, donde tienen lugar las reacciones de cracking y combustión del coque formado.
- 2.-Fraccionador Principal, donde se separan los productos obtenidos de acuerdo con su posterior utilización.
- 3.-Concentración de Gases, donde se separan y purifican los gases.

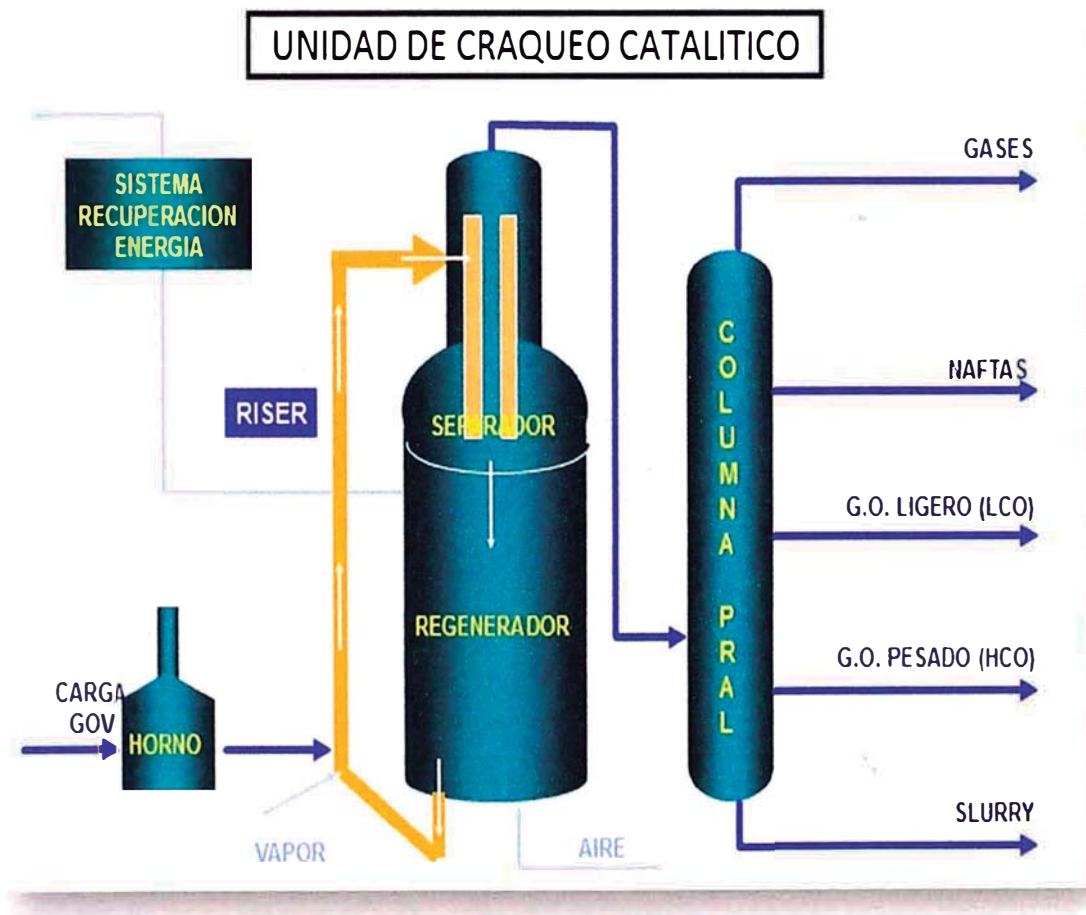
En la sección reactor/regenerador el catalizador fluidificado circula de forma continua entre el reactor y el regenerador. La alimentación se introduce en la línea por la que circula el catalizador regenerado al reactor (Riser). La reacción de cracking tiene lugar en dicha línea. Al llegar al reactor se separan los productos de cracking del catalizador, los productos se hacen pasar a través de ciclones, para separar las partículas de catalizador arrastradas, y pasan a la sección de fraccionamiento. El catalizador separado pasa por una zona de despojamiento en donde se separan los hidrocarburos retenidos en el catalizador, y circula hasta el regenerador. El control de la circulación se realiza mediante válvulas de diseño especial.

En el regenerador se inyecta el aire necesario para lograr la combustión del coque producido durante la reacción y que es causante de la pérdida temporal de actividad del catalizador, una vez regenerado se circula al reactor reiniciando el ciclo. Los gases de combustión pasan por unos ciclones para separar las partículas de

catalizador arrastradas y se envían a la chimenea pasando previamente por un sistema de recuperación energética, caldera e vapor o turbina de expansión.

Los productos de reacción a la salida del reactor se envían al fraccionador principal en donde se condensan y separan los productos principales. En esta sección se recupera calor precalentando la alimentación y generando vapor.

La nafta y gases salen por la cabeza de la columna y se envían a la sección de concentración de gases. Ver figura 15.



**Figura 15: Diagrama de la unidad de craqueo catalítico**

Fuente: Propia

A continuación se muestra un balance másico y volumétrico en la unidad de craqueo catalítico para una refinería de petróleo, donde se podrá apreciar la comparación entre los valores obtenidos a partir de datos medidos por los instrumentos de flujo y los datos obtenidos después de realizar la reconciliación de datos.

Se aprecia un imbalance de -2536 (ganancia volumétrica medida) vs -1403 barriles, (ganancia volumétrica reconciliada) es decir de -13.19 % a -6.86 % en volumen, se puede apreciar que el medidor de carga de gasóleo pesado a la unidad de craqueo catalítico (Carga GP), tiene un error de medición alto (1110 barriles de diferencia con el valor reconciliado), siendo necesaria su intervención para su calibración. Ver Figura 16.

Unidad de Craqueo Catalítico	Masa, toneladas		Volumen, barriles	
	Medido	Reconciliado	Medido	Reconciliado
<b>Intakes</b>				
Carga-GasUDP	24	19	160	154
Carga1-FL	597	617	4051	4181
Carga-GP	2189	2351	15019	16129
<b>Total In</b>	<b>2810</b>	<b>2986</b>	<b>19230</b>	<b>20464</b>
<b>Outturns</b>				
Prod1-CO	366	375	58	62
Prod-FG	149	151	949	954
Prod-C3	223	230	2529	2546
Prod-C4	248	234	2807	2798
Prod-NC	1384	1380	11649	11700
Prod-LCO	307	316	2035	2042
Prod-DCO	290	300	1738	1765
<b>Total Out</b>	<b>2967</b>	<b>2986</b>	<b>21766</b>	<b>21867</b>
<b>Imbalance (In-Out)</b>	<b>-157</b>	<b>0</b>	<b>-2536</b>	<b>-1403</b>
	<b>-5.60%</b>	<b>0.00%</b>	<b>-13.19%</b>	<b>-6.86%</b>

**Figura 16: Balance de una unidad craqueo catalítico, mediante reconciliación de datos.**

Fuente: Propia

## VII. ANALISIS COSTO/BENEFICIO

El costo promedio de implementar un sistema de reconciliación de datos que elimine los errores gruesos en una refinería de petróleo con una capacidad de procesamiento de 80 KBPD (Miles de barriles por día), incluyendo el servidor y aplicativos necesarios, donde se almacenarán los datos medidos y reconciliados de los instrumentos en tiempo real, oscila alrededor de los 410,000 US\$, siendo el periodo necesario para la implementación de 1 año.

El resumen de los costos se presenta a continuación en la Tabla 2.

**Tabla 2**

COSTOS	MONTO (US\$)
Costos Internos	50,000
Costos de Consultoría	40,000
Costos de Desarrollo	150,000
Costos de Software	135,000
Costos de Hardware	35,000
<b>TOTAL (US\$)</b>	<b>410,000</b>

Fuente: Propia

La utilización de este tipo de tecnología típicamente resulta en la disminución pérdidas / mermas, reducidas a menos del 0.5 %.

Los beneficios típicos para las refinerías de petróleo que están entre 80 - 200 KBPD de capacidad de procesamiento, oscilan en el orden de 1,000,000 - 4,000,000 US\$ por año.

La implementación de esta tecnología satisface una necesidad básica, que es la de conocer exactamente lo que sucedió en el negocio.

Por su complejidad, en refinerías de petróleo y petroquímicas es donde existe el mayor número de aplicaciones de este tipo.

La experiencia indica que la implementación de un sistema de balance de masa por reconciliación de datos resultará en por lo menos una reducción de la incertidumbre de aproximadamente el 1% de la carga de la planta (este número es conservador y puede ser sustancialmente mayor en refinerías de petróleo / petroquímicas que realizaban balances volumétricos en vez de másicos con anterioridad). Es razonable y conservador suponer que sólo la cuarta parte de esta cantidad se recuperará. Es por eso que se estima que la recuperación será del 0,25% de la carga.

Para una refinería de 100,000 barriles por día de capacidad de procesamiento, que opere durante 350 días al año, y asumiendo un margen de refinación de 2 US\$ por barril, se pueden esperar beneficios por 175,000 US\$ por año provenientes únicamente de la reducción de pérdidas de inventario.

El planeamiento determina las cantidades de componentes y productos terminados a producir para optimizar la ganancia de la planta. Cuanto más lejos del plan, más lejos del óptimo económico.

La reconciliación de datos ayuda además para satisfacer los compromisos de ventas que pueden traer aparejados penalidades o una mala relación comercial con los clientes, es importante su cumplimiento para asegurar que todos los componentes son fabricados en las proporciones adecuadas para ser mezclados, a fin de producir productos vendibles. Las desviaciones llevan a tener que comprar productos o utilizar productos de alto valor económico para satisfacer los compromisos de ventas de un producto de menor valor económico, esto por una mala programación de la producción (downgrade).

Un buen balance de materia es la única manera confiable de determinar la desviación de lo producido con respecto a lo planificado.



Se puede considerar que la imposibilidad de cerrar un buen balance, fuerza a la planta a degradar un cargamento de producto terminado por mes. Para una refinería de 100 KBPD de capacidad de procesamiento, que se ve forzada a degradar mensualmente una cantidad de 40,000 barriles, asumiéndolo a un precio diferencial igual a sólo 2 US\$ por barril, las pérdidas mensuales serían de 80,000 US\$ por mes. Si asumimos, en forma conservadora, que con la reconciliación de datos se evita el 50% de estas degradaciones, los beneficios anuales que se obtendrían serían de 480,000 US\$ por año.

Sólo teniendo en cuenta los beneficios por reducción de la pérdida de inventarios y por mejora en el seguimiento y contabilidad de la producción se estiman en 655,000 US\$ anuales de beneficios para una refinería de petróleo de 100 KBPD de capacidad de procesamiento.

Por lo anteriormente mencionado, se calcula que el periodo de recupero de la inversión está en el orden de no más de 6 meses.

## VIII. CONCLUSIONES

1. Todas las mediciones de instrumentos industriales siempre traen consigo porcentajes de aproximación y los cálculos realizados a partir de los datos recogidos, no son más que aproximaciones y en muchos casos difieren en gran magnitud con la realidad, por lo tanto el reto de la Reconciliación de Datos es acercarse tanto como sea posible a la realidad.
2. La reconciliación de datos y detección de errores gruesos son técnicas para incrementar la exactitud de los datos.
3. La reconciliación de datos hace uso explícito de las relaciones entre variables del proceso. Ajusta las mediciones de modo que las restricciones del proceso se cumplan, el cual es el caso de los balances de materia.
4. No deben de existir errores gruesos ni en la información ni en el modelo del proceso, la reconciliación de datos no puede desarrollarse en forma precisa, si no existe redundancia en las medidas de los instrumentos.
5. Es muy útil en simulación de procesos y optimización de los procesos productivos, ya que a partir de la reconciliación de datos se pueden obtener valores confiables de rendimientos de producción de las unidades de proceso, muy importante en la toma de decisiones en la industria.
6. La reconciliación de datos es útil para una óptima programación del mantenimiento de equipos y calibración de instrumentos de medición de flujo y nivel de tanques de almacenamiento.

## **IX. BIBLIOGRAFÍA**

1. Narasimhan S., Jordache C. "Data Reconciliation & Gross Error Detection"  
Editorial Gulf Publishing Company, (2000)
2. Bodington E. "Planning, Scheduling and Control Integration in the Process Industries", Editorial Mc Graw-Hill, (1995)
3. Romagnoli J., Sanchez M. "Data processing and Reconciliation for Chemical Process Operations", Editorial Academic Press, (2000)
4. Rodriguez L. "Probabilidad y Estadística para Ingenieros"  
Editorial ICM Espol, (2007)
5. Canavos G. "Probabilidad y Estadística", Editorial McGraw-Hill, (1992)
6. Pimsoft, Sigmafine  
[http://www.pimsoftinc.com/files/BR\\_Sigmafine\\_Overview.pdf](http://www.pimsoftinc.com/files/BR_Sigmafine_Overview.pdf)
7. Aspentech, Aspen Advisor  
[http://www.aspentech.com/brochures/Aspen\\_Advisor.pdf](http://www.aspentech.com/brochures/Aspen_Advisor.pdf)
8. Sotetica, S-TMS Results Analyzer  
[http://www.soteticavisualmesa.com/descargas/STMS\\_brochure\\_June08.pdf](http://www.soteticavisualmesa.com/descargas/STMS_brochure_June08.pdf)

## X. GLOSARIO

1. **Data coaptation:** Obtención de valores de flujos no medidos en un proceso, como resultado del balance de materia mediante reconciliación de datos.
2. **Tags:** Una etiqueta o tag es una clave asignada a un dato almacenado en un repositorio (servidor). Las etiquetas son en consecuencia un tipo de metadato, pues proporcionan información que describe al dato y que facilita su recuperación.
3. **Redundancia de medición:** Más de un punto medida para un mismo flujo, en diferentes posiciones del sistema.
4. **Imbalances o variaciones:** Desviaciones entre las entradas y salidas, calculadas mediante un balance de materia, luego de la reconciliación de datos.
5. **Crudo reducido:** Último corte de producción de la unidad de destilación primaria, también llamado residual de primaria en las refinerías, se obtiene de los fondos de la unidad, los cuales se enviarán como carga a la unidad de vacío.
6. **Visbreaking:** Unidad de procesamiento en las refinerías de petróleo, en la cual se realiza un craqueo térmico moderado del residuo de vacío y así obtener productos de mayor valor económico.
7. **Riser:** Parte del reactor donde se inicia la reacción por el contacto entre la alimentación y el catalizador en unidad de craqueo catalítico.
8. **Downgrade:** Proceso de degradación de un producto, al mezclarse con otro de menor valor económico, el caso contrario es conocido como upgrade.

## XI. APENDICE

### 11.1 Principales compañías licenciadoras de software de reconciliación de datos a nivel mundial.

#### 11.1.1 Pimsoft

La Cía. Pimsoft tiene en el mercado a su producto SIGMAFINE 4, el cual utiliza un motor de reconciliación de datos robusto, para conciliar y validar las mediciones iniciales, dando así la confianza a los datos validados, lo que le permite a las empresas actuar con rapidez y seguridad, debido a la consistencia de la información.

La Cía. Pimsoft es una de la licenciadoras más grandes a nivel Mundial, el cual tiene entre sus clientes de Oil & Gas a las siguientes empresas: Ver Tabla 3.

**Tabla 3**

Adriatic LNG	ORLEN	Shell
Bapco	Pemex Refinacion	Slovnaft
Bayernoil	Petrobras	Suncor Energy
Česká Rafinérská	<b>Petroperu</b>	Iplom
Chevron	PTT Aromatics	MOL Group
Chinese Petroleum	Qatar Liquefied Gas	Murphy Oil
ConocoPhillips	Raffineria di Gela	Essar Oil UK
Countrymark Cooperative	Raffineria di Milazzo	Galp Energia
Edison Gas	Raffineria di Roma (TotalErg)	GASCO
Engen Petroleum	Refineria de Cartagena	Hovensa
ERG	SASOL	OMV Petrom

Fuente: <http://www.pimsoftinc.com>

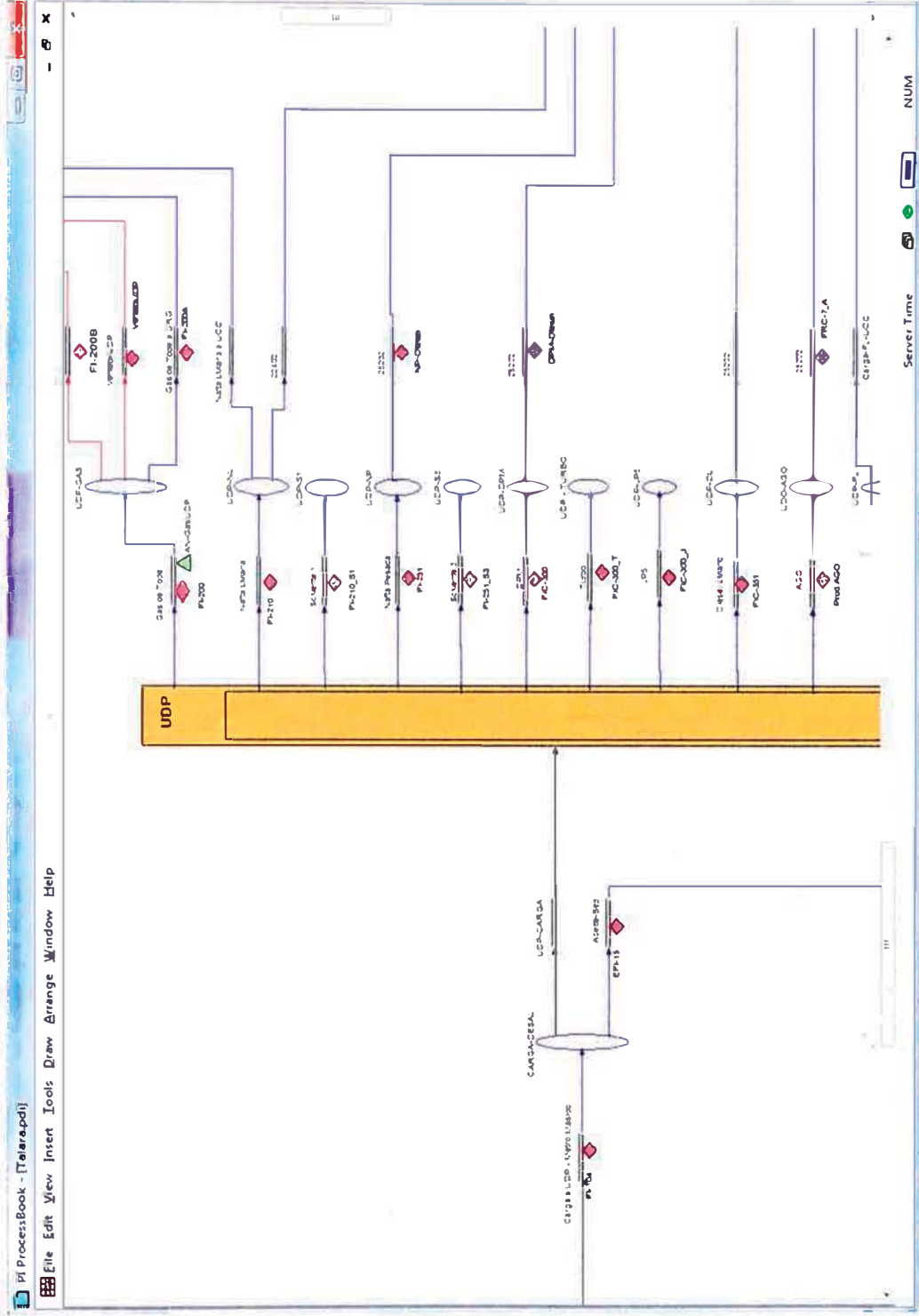


Figura 17: Software de reconciliación de datos, Sigmafine 4

Fuente: <http://www.pimsoftinc.com>

### 11.1.2 Aspentech

La Cía. Aspentech tiene en el mercado a su producto Aspen Advisor 5, el cual combina una interfaz gráfica de fácil manejo, basada en Microsoft Windows.

Cuenta con un algoritmo de reconciliación basado en el seguimiento de inventario y movimientos de productos, por dentro y fuera de una planta de procesamiento.

La Cía. Aspentech entre sus principales clientes a las siguientes empresas: Ver tabla 4

**Tabla 4**

Marathon Oil	BP	SABIC
Petrofac	Shell Global Solutions	ConocoPhillips
Kuwait Oil Company	Sasol	Westlake

Fuente: <http://www.aspentech.com>

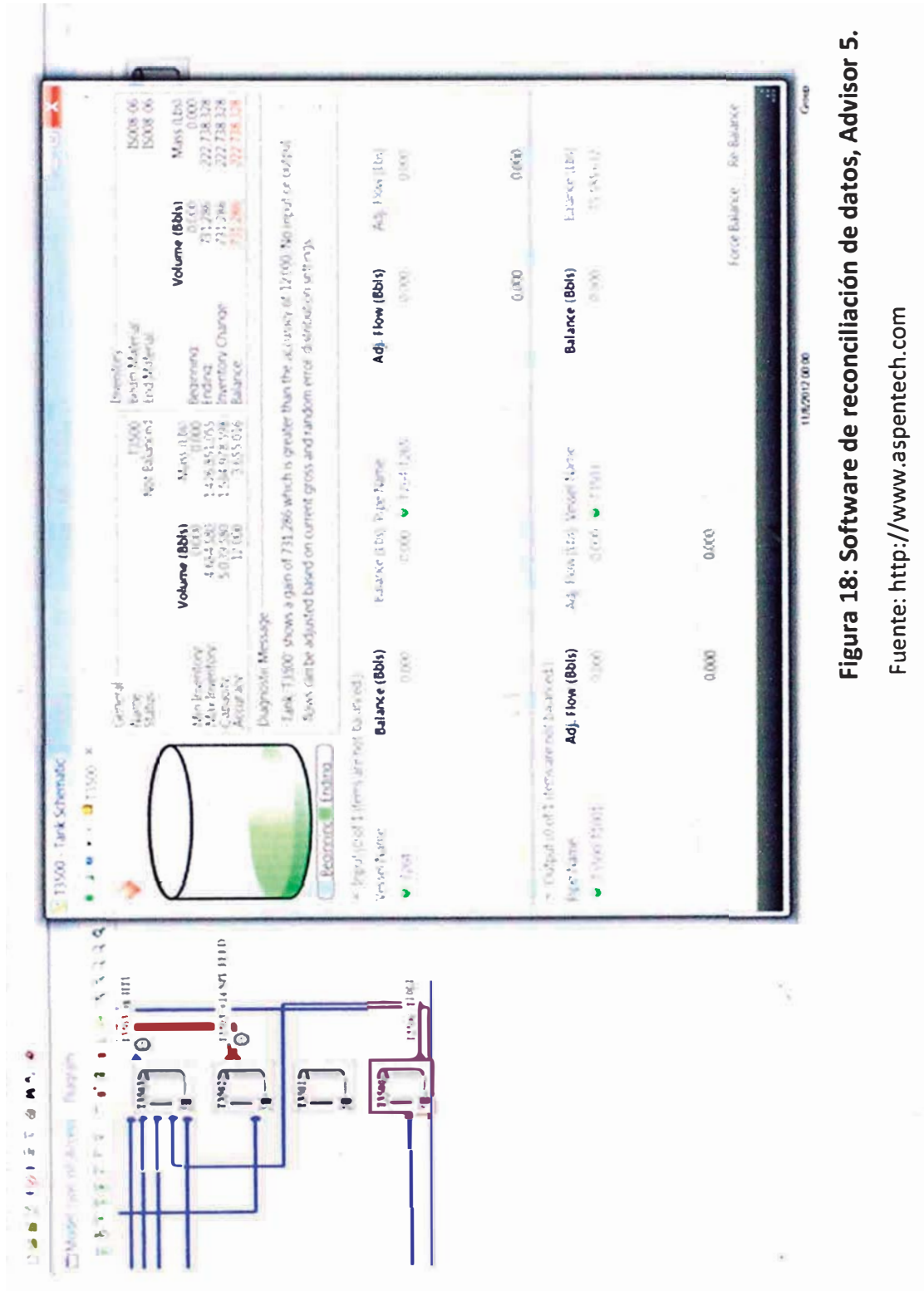


Figura 18: Software de reconciliación de datos, Advisor 5.

Fuente: <http://www.aspentech.com>



### 11.1.3 Soteica

La Cía. Soteica tiene en el mercado a su producto S-TMS Results Analyzer, el cual ofrece una completa solución de la técnica de reconciliación de datos, que se soporta en las iniciativas de detección de pérdida de masa en entornos industriales complejos.

La Cía. Soteica tiene entre sus principales clientes a las siguientes empresas: Ver tabla 5.

**Tabla 5**

Banner Pharmacaps	Ergon Refining	Petroperu
Chevron Phillips	Indian Oil	Degussa
DuPont	Petrobras	Lyondell

Fuente: <http://www.soteicavisualmesa.com>

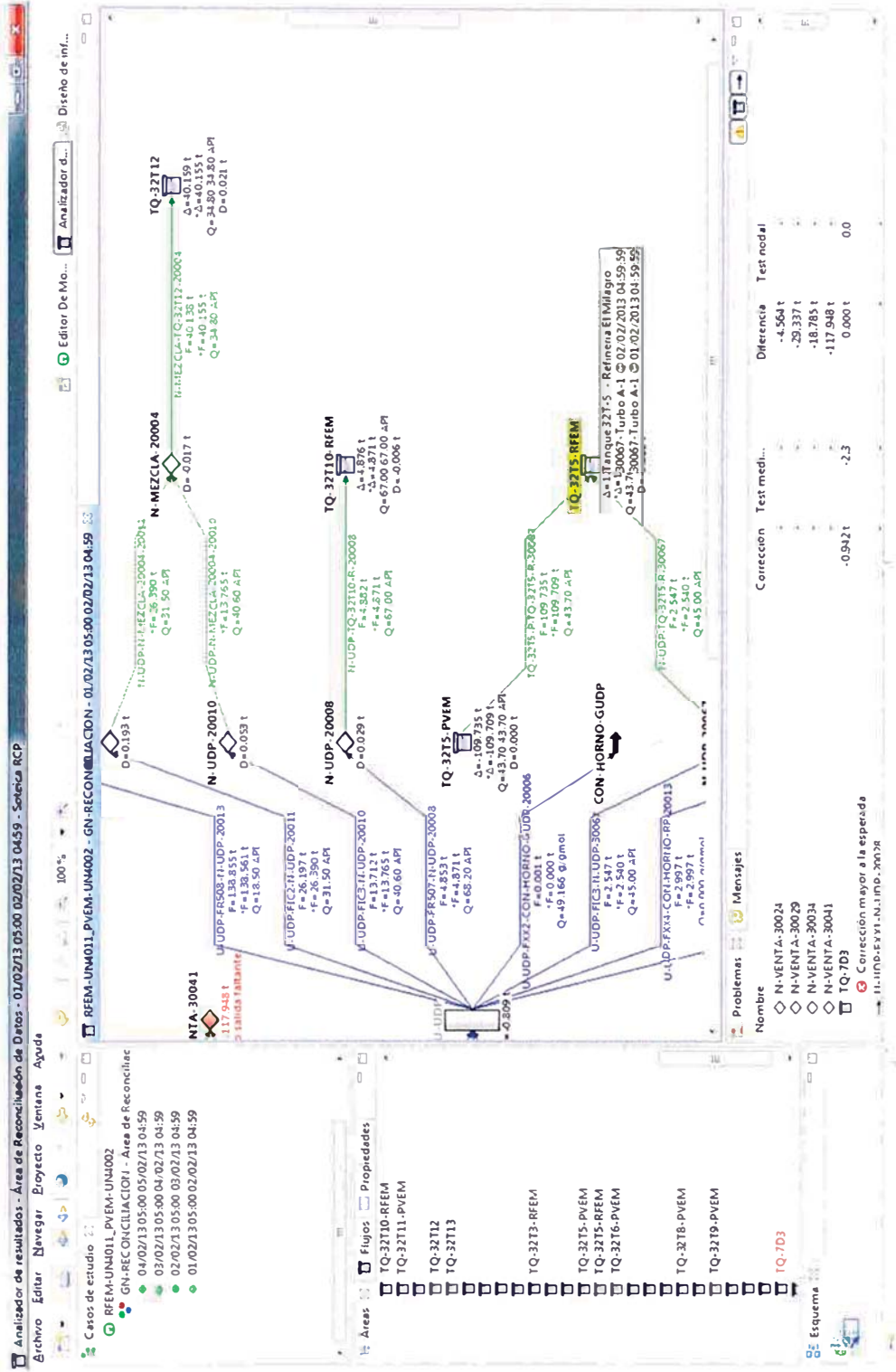


Figura 19: Software de reconciliación de datos, S-TMS Results Analyzer.

Fuente: <http://www.sotecoavisualmesa.com>

## **11.2 Fundamentos de estadística**

### **11.2.1. Conceptos básicos**

El tema de la estadística moderna abarca la recolección, presentación y caracterización de información para ayudar tanto en el análisis de datos como en el proceso de la toma de decisiones.

La estadística se divide en dos:

**Estadística descriptiva.** Puede definirse como aquellos métodos que incluyen la recolección, presentación y caracterización de un conjunto de datos con el fin de describir apropiadamente las diversas características de ese conjunto de datos.

Ejemplos: promedio de bateo, índice de accidentes, tasa de mortalidad, calcular desviación estándar, etc. :

**Inferencia estadística o estadística inferencial.** Puede definirse como aquellos métodos que hacen posible la estimación de una característica de una población o la toma de una decisión referente a una población, basándose sólo en los resultados de la muestra.

#### **Teoría de decisión.**

Algunos de los problemas más importantes de la inferencia estadística se refieren a la evaluación de los riesgos y las consecuencias que pueden ocurrir al hacer generalizaciones a partir de una muestra de datos. Esto incluye una estimación de la probabilidad de tomar decisiones erróneas, las posibilidades de hacer predicciones incorrectas. En los últimos años, se han hecho intentos de abordar todos estos problemas dentro del marco de referencia de una teoría unificada llamada teoría de decisión. Si bien esta teoría tiene muchas ventajas conceptuales y teóricas, plantea algunos problemas de aplicación que son difíciles de resolver. Para entenderlos debe

comprenderse que, por muy objetivamente que se planee un experimento o investigación, es imposible eliminar todos los elementos subjetivos. Un elemento de subjetividad interviene aun cuando definimos elementos como “bueno” o “mejor” con respecto a la razón de criterios de decisión (por ejemplo buscaremos una línea recta que “mejor se ajuste a un conjunto dado de pares ordenados de datos).

La subjetividad influye mucho en la elección de los métodos estadísticos o fórmulas empleadas en una situación específica.

**Población (o universo):** Es la totalidad de elementos o cosas bajo consideración.

**Una muestra:** Es la porción de la población que se selecciona para su análisis.

**Un parámetro:** Es una medida de resumen que se calcula para describir una característica de toda una población.

**Una estadística:** Es una medida de resumen que se calcula para describir una característica de una sola muestra de la población.

**Muestra aleatoria:** Es una muestra elegida independientemente de todas las demás, con la misma probabilidad que cualquier otra y cuyos elementos están elegidos independientemente unos de otros y con la misma probabilidad.

### 11.2.2 Descripción de datos

Los datos ayudan a los encargados de la toma de decisiones a hacer conjeturas bien fundamentadas acerca de las causas y, por lo tanto, sobre los efectos probables de ciertas características de algunas situaciones. Por lo demás, el conocimiento de las tendencias adquirido con la experiencia permite conocer los posibles resultados y planear con anticipación.

Algunas veces puede ser satisfactorio presentar los datos tal como se encuentran y obtener información directamente de ellos; otras veces solo habrá que agruparlos y presentarlos en forma gráfica o tabulada, aquí el uso de las tecnologías computacionales es mucha utilidad y rapidez.

### **11.2.3 Datos agrupados**

Cuando la muestra consta de más de 30 datos, lo aconsejable es agrupar los datos en clases y a partir de estas determinar las características de la muestra y por consiguiente las de la población de donde fue tomada.

Los datos agrupados se pueden resumir gráficamente, o en tablas, y mediante el uso de medidas numéricas, como la media, la amplitud, la desviación estándar, y otras más. El nombre que reciben los datos ordenados en grupos o categorías es el de distribución de frecuencia.

#### **Distribución de frecuencia**

Una forma de sintetizar los datos consiste en valerse de una tabla o distribución de frecuencia. Tomemos como ejemplo el inventario promedio en días de 20 tiendas de conveniencia.

En las tablas que se mostraran a continuación se han incluido datos idénticos referentes al inventario promedio y se han dispuesto primero como un arreglo en orden ascendente y luego como una distribución de frecuencia. Para obtener la tabla 7 se tuvo que dividir los datos en grupos de valores semejantes. Después se registraron el número de puntos graficados (observaciones) de datos que caían dentro de cada grupo.

**TABLA 6:** Arreglo de datos del inventario promedio (en días) de 20 tiendas de artículos de conveniencia.

<b>Inventario promedio</b>	<b>Día</b>
2.0	1
3.4	2
3.8	3
4.1	4
4.1	5
4.3	6
4.7	7
4.9	8
5.5	9
5.5	10
3.4	11
3.8	12
4.0	13
4.1	14
4.2	15
4.7	16
4.8	17
4.9	18
5.5	19
5.5	20

Fuente: Propia

**TABLA 7:** Distribución de frecuencia del inventario promedio (en días) de 20 tiendas de artículos de conveniencia (6 clases)

<b>Clase (grupo de observaciones de datos con valores semejantes)</b>	<b>Frecuencia (número de observaciones en cada clase)</b>
<b>2.0 a 2.5</b>	<b>1</b>
<b>2.6 a 3.1</b>	<b>0</b>
<b>3.2 a 3.7</b>	<b>2</b>
<b>3.8 a 4.3</b>	<b>8</b>
<b>4.4 a 4.9</b>	<b>5</b>
<b>5.0 a 5.5</b>	<b>4</b>

Fuente: Propia

La distribución de frecuencia es una tabla que organiza los datos en clases; es decir, en grupos de valores que describen una característica de los datos. El inventario promedio es una característica de las 20 tiendas de conveniencia

Una distribución de frecuencia muestra el número de observaciones provenientes del conjunto de datos que caen dentro de cada una de las clases. Si podemos determinar la frecuencia con que ocurren los valores en cada clase de un conjunto de datos, estaremos en condiciones de construir una distribución de frecuencia.

### **Características de las distribuciones de frecuencia relativa**

Hasta ahora se ha expresado la frecuencia con que ocurren los valores en cada clase como el número total de observaciones que caen en dicha clase. También se puede expresar la frecuencia de cada valor como una fracción o porcentaje del número total de observaciones. La frecuencia de un inventario promedio, digamos de 4.4 a 4.9, es 5 en la tabla 7 y de 0.25 en la tabla 8. Para obtener este último valor, dividimos la

frecuencia de esta clase (5) entre el número total de observaciones en el conjunto de datos (20). La respuesta puede expresarse como una fracción ( $\frac{5}{20}$ ), un decimal (0.25) o un porcentaje (25 %). Una distribución de frecuencia relativa presenta las frecuencias en fracciones o porcentajes.

**TABLA 8:** Distribución de frecuencia relativa del inventario promedio (en días) de 20 tiendas de artículos de conveniencia

<b>CLASE</b>	<b>FRECUENCIA</b>	<b>Frecuencia relativa: Fracciones de observaciones en cada clase</b>
<b>2.0 a 2.5</b>	<b>1</b>	<b>0.05</b>
<b>2.6 a 3.1</b>	<b>0</b>	<b>0.00</b>
<b>3.2 a 3.7</b>	<b>2</b>	<b>0.10</b>
<b>3.8 a 4.3</b>	<b>8</b>	<b>0.40</b>
<b>4.4 a 4.9</b>	<b>5</b>	<b>0.25</b>
<b>5.0 a 5.5</b>	<b>4</b>	<b>0.20</b>
	<b>20</b>	<b>1.00 suma de las frecuencias relativas de todas las clases</b>

Fuente: Propia

La suma de todas las frecuencias relativas es de 1.00 o 100 %. Esto sucede porque una distribución de frecuencia relativa para cada clase con su fracción o porcentaje correspondiente de los datos totales. Por lo anterior, las clases en cualquier distribución de frecuencia simple o relativa son exhaustivas. Todos los datos encajan en una u otra categoría. Observe también que las clases son mutuamente excluyentes; e decir, ninguna observación cae dentro de más de una categoría.



**Clases discretas.**

Los esquemas de clasificación pueden ser cualitativos o cuantitativos y discretos o continuos. Las clases discretas son entidades individuales que no pasan de una clase a la siguiente sin una ruptura. Son discretas las siguientes clases: el número de hijos de las familias, el número de camiones que poseen las compañías transportistas, las ocupaciones de los graduados universitarios.

Los datos continuos pueden pasar de una clase a la siguiente sin ruptura alguna. Contienen una medida numérica como el peso de unas latas de tomates, los kilogramos de presión sobre el concreto, o el promedio de calificaciones de los universitarios el último semestre.

**Construcción de una distribución de frecuencia**

Ahora que hemos aprendido a dividir una muestra en clases, ya se esta en condiciones de tomar datos brutos y construir una distribución de frecuencia. A continuación tomaremos el siguiente ejemplo para el desarrollo de la distribución de frecuencia:

**TABLA 9: Concentraciones de cloro en partes por millón (ppm) en 30 galones de agua tratada.**

<b>Concentración de cloro en ppm</b>	<b>Numero de muestra</b>
16.2	1
15.4	2
16.0	3
16.6	4
15.9	5
15.8	6
16.0	7
16.8	8
16.9	9
16.8	10
15.7	11
16.4	12
15.2	13
15.8	14
15.9	15
16.1	16
15.6	17
15.9	18
15.6	19
16.0	20
16.4	21
15.8	22
15.7	23
16.2	24
15.6	25
15.9	26
16.3	27
16.3	28
16.0	29
16.3	30

Fuente: Propia

Para analizar los datos de esta tabla se sigue los siguientes pasos:

**1.- Escoger el tipo y número de clases para dividir los datos.** En este caso, ya se ha optado por clasificar los datos según la medida cuantitativa del número de ppm del cloro en el agua tratada, en vez de hacerlo a partir de un atributo cualitativo como color o el olor del agua.

Después necesitamos decidir cuántas clases utilizar y el intervalo (la distancia que debe comprender cada clase).

El número de clases por utilizar depende principalmente del número de observaciones en los datos. Esto es, un número mayor de observaciones requiere un mayor número de grupos de clase. En general la distribución de frecuencia debe tener al menos cinco clases pero no más de 15. Si no hay suficientes agrupamientos de clase o si hay demasiados, se obtendrá poca información.

A continuación se muestra la tabla 10, que nos puede ser útil para seleccionar el número de clases, aclarando que esta designación no es obligatoria y puede ser a decisión del analista.

**Tabla 10: Selección de número de clases**

TAMAÑO DE MUESTRA, Nº DE DATOS	NUMERO DE CLASES (K)
<b>Menos de 50</b>	<b>5 a 7</b>
<b>50 a 99</b>	<b>6 a 10</b>
<b>100 a 250</b>	<b>7 a 12</b>
<b>250 en adelante</b>	<b>12 a 15</b>

Fuente: Propia

Para designar el número de clases se usará la letra K, para este ejemplo se usará K= 6 clases.

Es deseable que el ancho de cada agrupamiento de clase (intervalo del inicio de una clase al inicio de la siguiente) sea igual. Para determinar el ancho de cada clase, el alcance de los datos se divide entre el número de agrupamientos de clase deseado:

$$\text{Ancho de Intervalo} : \frac{\text{Rango}}{\text{número total de intervalos de clase}}$$

$$A = \frac{R}{K} \quad R = \text{Val. mayor} - \text{Val. menor}$$

La amplitud o ancho del intervalo se calcula:

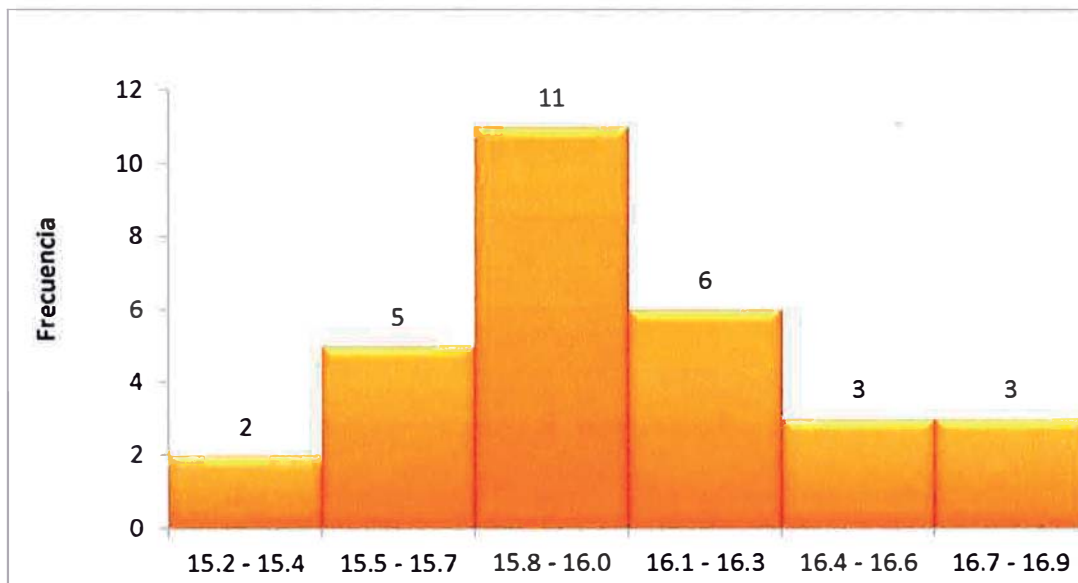
$$A = \frac{16.9 - 15.2}{6} = 0.283 \approx 0.30 \text{ ppm}$$

Se han clasificado los datos según la media cuantitativa de cuantas ppm se encuentran en el agua tratada. Se escogieron seis clases para cubrir el intervalo de 15.2 a 16.9 y en consecuencia se utilizará 0.3 ppm como el ancho de los intervalos de clase.

**2.- Clasificar los puntos de datos en clases y contar el número de puntos en cada clase:** Esta información aparece en la tabla 5. Toda observación de datos encaja por lo menos en una clase y ninguna observación lo hace en más de una clase. Por lo que nuestras clases son exhaustivas y mutuamente excluyentes. Observe que el límite inferior de la primera clase corresponde a la menor observación de datos de la muestra, y que el límite superior de la última clase corresponde a la observación mayor de los datos.

**TABLA 11: Concentraciones de cloro en muestras de agua tratada con intervalos de clase de 0.3 ppm**

<b>CLASE</b>	<b>FRECUENCIA</b>
15.2 - 15.4	2
15.5 - 15.7	5
15.8 - 16.0	11
16.1 - 16.3	6
16.4 - 16.6	3
16.7 - 16.9	3
<b>Total 30</b>	



**Figura 20: Histograma de frecuencia absoluta**

Fuente: Propia

## Histogramas

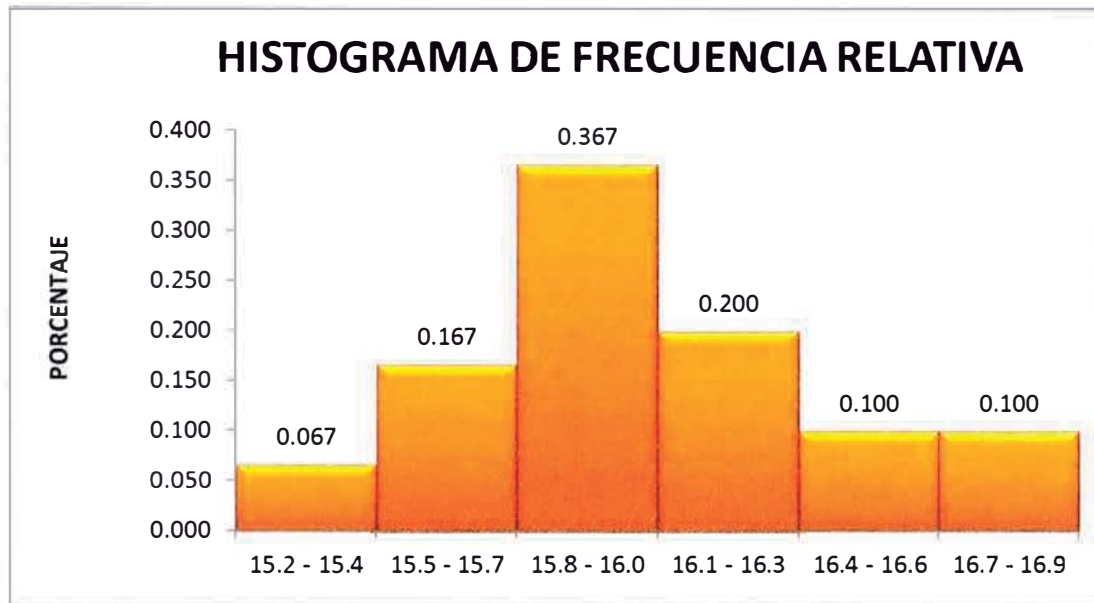
La figura 20 es un ejemplo de un histograma. El histograma es una serie de rectángulos, todos ellos de anchura proporcional a la gama de valores dentro de una clase y también de altura proporcional a los elementos que caen dentro de la clase. Si las clases que empleamos en la distribución de frecuencia tienen el mismo ancho, las barras verticales del histograma lo tendrán también. La altura de la barra de cada clase corresponde al número de elementos de esta última.

Un histograma que se sirve de la frecuencia relativa (tabla 12) de las observaciones de datos en cada una de las clases y no del número real de observaciones recibe el nombre de histograma de frecuencia relativa (figura 21). Este tiene la misma forma que un histograma de frecuencia absoluta hecho con el mismo conjunto de datos. La única diferencia entre el histograma de frecuencia absoluta y el de frecuencia relativa, es la escala vertical de la izquierda, en el primero es el número absoluto de observaciones en cada clase y en el segundo es el número de observaciones en cada clase como una fracción del número total de ellas.

**TABLA 12:** Concentraciones de cloro en muestras de agua tratada con intervalos de clase de 0.3 ppm usando la frecuencia relativa.

CLASE	FRECUENCIA	FRECUENCIA RELATIVA
15.2 - 15.4	2	0.067
15.5 - 15.7	5	0.167
15.8 - 16.0	11	0.367
16.1 - 16.3	6	0.200
16.4 - 16.6	3	0.100
16.7 - 16.9	3	0.100
<b>Total</b>	<b>30</b>	<b>1.000</b>

Fuente: Propia



**Fig. 21: Histograma de Frecuencia Relativa**

Fuente: Propia

### **Polígono de frecuencias**

Aunque de menor uso, los polígonos de frecuencias son otro medio de representar gráficamente tanto las distribuciones de frecuencia simples como las de frecuencia relativa. Para construir un polígono de frecuencias, marcamos las frecuencias sobre el eje vertical y los valores de la variable que vamos a medir las marcamos sobre el eje horizontal, tal como lo hicimos con los histogramas. El siguiente paso consiste en graficar cada frecuencia de clase dibujando un punto sobre su marca de clase, o punto medio, y conectar los puntos consecutivos con una recta para formar un polígono (figura de muchos lados).

A continuación en la tabla 13 se muestran los datos de una distribución de frecuencias para elaborar un polígono.

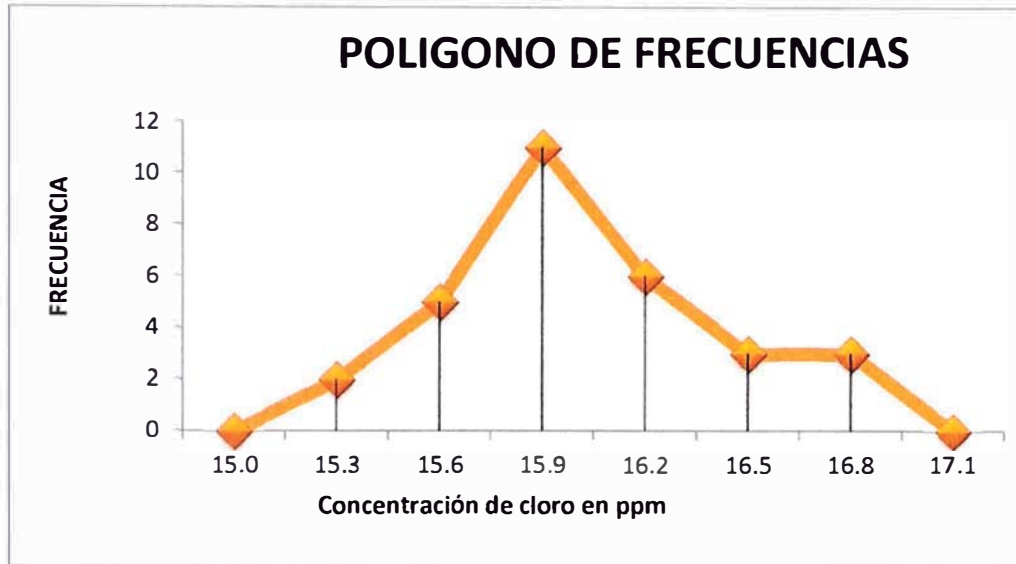
**TABLA 13:** Datos de la concentración de cloro en ppm (distribución de frecuencia) para graficar polígono de frecuencias.

<b>CLASE</b>	<b>FRECUENCIA</b>	<b>CENTRO CLASE</b>
	0	15.0
<b>15.2 - 15.4</b>	2	15.3
<b>15.5 - 15.7</b>	5	15.6
<b>15.8 - 16.0</b>	11	15.9
<b>16.1 - 16.3</b>	6	16.2
<b>16.4 - 16.6</b>	3	16.5
<b>16.7 - 16.9</b>	3	16.8
	0	17.1
	30	

Fuente: Propia

En la figura 21 se muestra un polígono de frecuencias construido con los datos de la tabla 13. Si comparamos esta figura con las figuras anteriores vemos que se han agregado clases en cada extremo de la escala de valores observados. Estas dos nuevas clases contienen cero observaciones, pero permiten al polígono alcanzar el eje horizontal en ambos extremos de la distribución.





**Figura 22: Polígono de frecuencias**

Fuente: Propia

El polígono es simplemente una gráfica lineal que une los puntos medios de todas las barras en un histograma.

Se llama polígono de frecuencias relativas a aquel que usa la frecuencia relativa de los puntos de datos en cada clase y no el número real de puntos.

### Ojivas

Una distribución de frecuencia acumulativa nos permite ver cuántas observaciones se hallan por arriba o por debajo de ciertos valores, en lugar de limitarnos a anotar los números de elementos dentro de los intervalos. Por ejemplo, si queremos saber cuántos galones contienen menos de 17.0 ppm, podemos servirnos de una tabla que

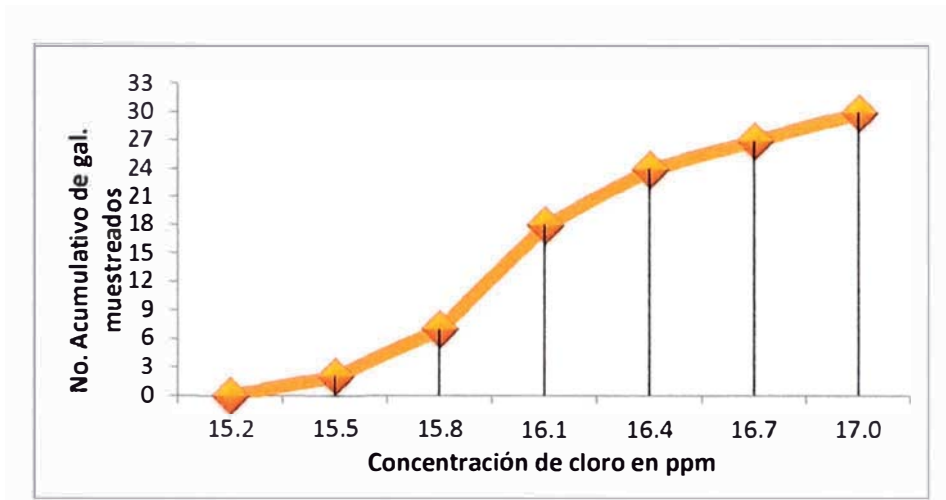
incluya frecuencias acumulativas “menores que” en nuestra muestra como se observa en la tabla 14.

**TABLA 14:** Distribución de frecuencia acumulativa “menor que” de las concentraciones de cloro en ppm.

<b>CLASE</b>	<b>FRECUENCIA ACUMULATIVA</b>
<b>Menor que 15.2</b>	0
<b>Menor que 15.5</b>	2
<b>Menor que 15.8</b>	7
<b>Menor que 16.1</b>	18
<b>Menor que 16.4</b>	24
<b>Menor que 16.7</b>	27
<b>Menor que 17.0</b>	30

Fuente: Propia

Se llama ojiva a la gráfica de una distribución de frecuencia acumulativa. La ojiva de una distribución de este tipo se muestra en la figura 23. Los puntos graficados representan la cantidad de galones que tienen menos cloro que las partes por millón indicadas sobre el eje horizontal.



**Figura 23: Ojiva “menor que” de la distribución de las concentraciones de cloro en ppm para 30 galones de agua tratada.**

Fuente: Propia

En ocasiones la información que se utiliza se presenta a partir de frecuencias “mayores que”. La ojiva apropiada para tal información tendrá una pendiente hacia abajo y hacia la derecha.

También es posible construir una ojiva de una distribución de frecuencia relativa, de la misma manera que una absoluta.

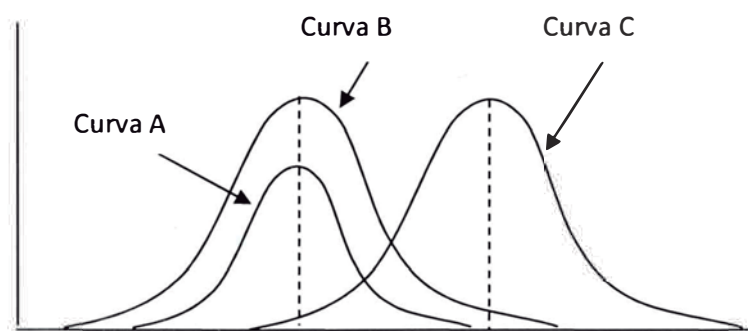
#### **11.2.4 Datos no Agrupados**

Cuando la muestra que se ha tomado de la población o proceso que se desea analizar, es menor de 30 elementos en la muestra, entonces estos datos son analizados sin necesidad de formar clases con ellos y a esto se le llama tratamiento de datos no agrupados.

Con frecuencia un conjunto de números se pueden reducir a una o unas cuantas medidas numéricas sencillas que resumen el conjunto total. Tales medidas son fáciles de comprender que los datos originales, no procesados. Dos importantes características de los datos que las medidas numéricas pueden poner de manifiesto son: 1) el valor central o típico del conjunto y 2) la dispersión de los números.

### Medidas de Tendencia Central para datos no agrupados

Estas medidas se utilizan para indicar un valor que tiende a ser el más representativo de un conjunto de números. Las tres medidas que más comúnmente se emplean son la media, la mediana y la moda. En la figura 24 podemos apreciar el significado gráfico de las medidas de tendencia central.



**Figura 24: Comparación de la localización central de las tres curvas**

Fuente: Propia

### Media Aritmética

Es lo que viene a la mente de la mayoría de las personas cuando se menciona la palabra “promedio” Como este término tiene ciertas propiedades matemáticas deseables, es la más importante de las tres. La media aritmética se calcula al sumar los valores de un conjunto de datos y dividir el resultado de la suma entre el número de valores del mismo.

La media poblacional se calcula:

$$\mu = \frac{\sum_{i=1}^n X_i}{N} \quad (11)$$

Donde  $\mu$  = Media Poblacional y N = N° datos de la población

La media muestral se calcula de la siguiente manera:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (12)$$

Donde:  $\bar{X}$  = media muestral y n= n° datos en la muestra

### Media Geométrica

Algunas veces estamos manejando cantidades que cambian a lo largo de un periodo; entonces necesitamos conocer una tasa promedio de cambio, como el crecimiento promedio a través de un periodo de varios años. En tales casos la simple media aritmética no es apropiada, porque no proporciona las respuestas correctas. Lo que necesitamos encontrar es la media geométrica. Se calcula con la siguiente formula:

$$G = \sqrt[n]{X_1 * X_2 \dots * X_n} \quad (13)$$

### La media Ponderada

Nos permite obtener un promedio que tiene en cuenta la importancia de cada valor para el total global.

$$\bar{X}_w = \frac{\sum_{i=1}^n X_i w_i}{\sum_{i=1}^n w_i} \quad (14)$$

Donde  $w_i$  = Peso asignada a cada observación

### La Mediana

Es una medida de tendencia central diferente a las medias. La mediana es un solo valor del conjunto de datos que mide el elemento central en los datos una vez que estos han sido ordenados según su valor o magnitud. Para calcular la mediana se presentan dos casos:

a).- Cuando el número de datos en la muestra es impar. En este caso después de ordenar los datos de la muestra en cuanto a su magnitud, es decir de mayor a menor o

de menor a mayor valor, se procede a localizar aquel dato que se encuentra justo en el centro de los datos o en la parte central de los mismos.

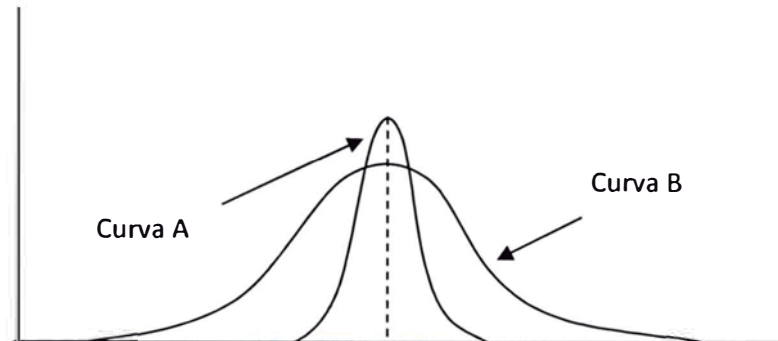
b).- Cuando el número de datos en la muestra es par. En este caso después de ordenar los datos en cuanto a su magnitud, observamos que en la parte central de los datos no se encuentra dato alguno, en este caso la mediana, tomará el valor promedio de dos datos; el que se encuentra antes de la parte central y el que se encuentra después de la parte central.

### **La Moda**

Se define como aquel valor o valores que más se repiten entre los datos que se han obtenido en una muestra. Una vez que los datos se han ordenado es posible observar la moda o modas de la misma, la distribución de los datos puede ser amodal (carece de moda), unimodal (tiene una sola moda), bimodal (tiene dos modas), polimodal (tiene más de dos modas).

### **Medidas de Dispersión para datos no agrupados**

Cuando se tiene una muestra de datos obtenida de una población cualquiera, es importante determinar sus medidas de tendencia central así como también el determinar qué tan dispersos están los datos en la muestra. Por ejemplo las medidas de dispersión indican si los valores están relativamente cercanos uno del otro o si se encuentran dispersos. En la figura 25 se aprecia la dispersión que hay entre las dos curvas.



**Figura 25: Dispersión entre dos curvas**

Fuente: Propia

La dispersión mide cuán próximos están los valores de un grupo entre sí, es conveniente considerar cuatro variables de dispersión: La amplitud de variación (o rango), la desviación media, la varianza y la desviación estándar. Todas estas medidas, excepto la amplitud de variación, toman a la media como punto de referencia. Un valor cero indica que no hay dispersión, la dispersión aumenta a medida que se incrementa el valor de la medida (amplitud, varianza, etc.).

### **Amplitud de variación o rango**

Es la medida más sencilla de calcular y comprender. Es la diferencia entre el valor mayor y el valor menor. Se determina de la siguiente manera:

$$R = \text{Valor Mayor} - \text{Valor menor}$$

(15)



La principal limitación de la amplitud de variación o rango es que considera solamente los valores extremos de un conjunto, y no proporciona mayor información respecto a los demás valores del mismo.

Se ha encontrado determinado número de medidas de dispersión que utilizan la media como punto de referencia. Todas incluyen la obtención de la desviación, o diferencia entre cada valor y la media ( $x_i - \bar{x}$ ). Se consideran tres de estas medidas. La primera trata sobre la desviación absoluta respecto de la media, mientras que las otras dos se concentran en las desviaciones cuadradas a partir de la media.

#### **Desviación absoluta media (DAM).**

Mide la desviación promedio de valores con respecto a la media del grupo, sin tomar en cuenta el signo de la desviación. Se obtiene al restar la media a cada valor del grupo, eliminando el signo de la desviación, hallando después el promedio.

$$DAM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (16)$$

#### **Varianza**

La varianza de una muestra se calcula casi en la misma forma que la desviación media, con dos pequeñas diferencias:

- 1) las desviaciones se elevan al cuadrado antes de ser sumadas y,
- 2) se obtiene el promedio, utilizando  $n - 1$  en lugar de  $n$ .

$$s_x = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad (17)$$

### Desviación estándar

La desviación estándar es simplemente la raíz cuadrada positiva de la varianza. De este modo si la varianza es 81 la desviación estándar es 9. Para obtener la desviación estándar se debe calcular la varianza y hallar su raíz cuadrada.

Las fórmulas para la desviación estándar son:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - [(\sum x_i)^2 / n]}{n-1}} \quad (18)$$

La desviación estándar es una de las medidas de resumen que más se suele utilizar para distribuciones, y desempeña un papel importante en la estadística. Es importante notar que las unidades de la desviación estándar son las mismas que las de la media.