Neiva, 12 de Julio del 2021

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El (Los) suscrito(s):

Ivan Gabriel Molina Perilla, con C.C. No. 1075301891,

Dimitry Andres Riascos Malaver, con C.C. No. 1080187400,

Autor(es) de la tesis y/o trabajo de grado o Titulado DENSIFICACIÓN DE SALMUERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO

presentado y aprobado en el año 2021 como requisito para optar al título de

ESPECIALISTA EN ESTADÍSTICA;

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales “open access” y en las redes de información con las cuales tenga convenio la Institución.

- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.

- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores” , los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE:

Firma: Ivan Molina P.
C.C. 1075 301 891 de Neiva

EL AUTOR/ESTUDIANTE:

Firma:

**TÍTULO COMPLETO DEL TRABAJO:** DENSIFICACIÓN DE SALMUERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO

**AUTOR O AUTORES:**

| **Primero y Segundo Apellido** | **Primero y Segundo Nombre** |
|---|---|
| Molina Perilla | Ivan Gabriel |
| Riascos Malaver | Dimitry Andrés |

**DIRECTOR Y CODIRECTOR TESIS:**

| Primero y Segundo Apellido | Primero y Segundo Nombre |
|---|---|
| Vargas Castellanos | Constanza |

**ASESOR (ES):**

| Primero y Segundo Apellido | Primero y Segundo Nombre |
|---|---|
| | |

**PARA OPTAR AL TÍTULO DE:** Especialista en estadística

**FACULTAD**: Ciencias Exactas y Naturales

**PROGRAMA O POSGRADO:** Especialización en Estadística

**CIUDAD:** Neiva **AÑO DE PRESENTACIÓN:** 2021 **NÚMERO DE PÁGINAS:** 86

**TIPO DE ILUSTRACIONES** (Marcar con una **X**):

Diagramas___ Fotografías___ Grabaciones en discos___ Ilustraciones en general X Grabados___ Láminas___ Litografías___ Mapas___ Música impresa___ Planos___ Retratos___ Sin ilustraciones __ Tablas o Cuadros X

**SOFTWARE** requerido y/o especializado para la lectura del documento: Ninguno

**MATERIAL ANEXO:** Ninguno

**PREMIO O DISTINCIÓN** *(En caso de ser LAUREADAS o Meritoria)*:

**PALABRAS CLAVES EN ESPAÑOL E INGLÉS:**

| Español | Inglés | Español | Inglés |
|---|---|---|---|
| 1. Salmuera | Brine | 6. Pruebas de laboratorio | Lab tests |
| 2. Densidad | Density | 7. Modelos de regresión | Regressions models |
| 3. Cloruro de potasio | Potassium chloride | | |
| 4. Formiato de sodio | Sodium formate | | |
| 5. Correlación | Correlation | | |

**RESUMEN DEL CONTENIDO:** (Máximo 250 palabras)

Durante la vida productiva de un pozo petrolero se requieren realizar intervenciones a este; en las intervenciones es muy importante tener el control del pozo, esto significa que la presión de la formación no supere la presión de fondo de pozo ya que esto puede desencadenar en un descontrol de la presión y que los fluidos dentro del pozo salgan a superficie de forma precipitada, esto puede generar daños ambientales o incluso lesión a personas. Para asegurar que el pozo esté controlado en todo momento se inyecta al pozo un fluido densificado que provea una columna hidrostática que supere la presión de fondo de pozo, a este fluido se le llama comúnmente fluido de control. El fluido usado más comúnmente es la salmuera, la cual consiste en agua y una sal que aumenta la densidad del agua. De acuerdo con las condiciones del pozo, se puede requerir que el fluido de control tenga una determinada densidad, y de acuerdo con la densidad se utiliza la sal que se acomode mejor a los requerimientos. La dificultad que se tiene con la sal como método densificante es el punto de saturación del agua, esto quiere decir que el agua solubilizará la sal hasta cierto punto, de ahí en adelante precipitará la sal que se le agregue no aumentará la densidad de la mezcla. Se utilizarán conceptos estadísticos tales como modelos de regresión y pruebas de hipótesis para obtener modelos estadísticamente significativos que relacionen la cantidad de sal necesaria para preparar un barril de salmuera (Mezcla de cloruro de potasio y formiato de sodio) a la densidad requerida.

**ABSTRACT:** (Máximo 250 palabras)

During the productive life of a petroleum well, it is required to make interventions to this. In those interventions it is particularly important to make sure the pressure of the formation does not surpass the well's bottom pressure since it may cause the fluids inside the well to rise up to the surface hastily. This could cause environmental damages or even some injuries to people. In order to guarantee that the well is under control in every moment, it gets injected with a densificated fluid that provides a hydrostatic column which exceeds the well's bottom pressure. This fluid is called control fluid. The most used fluid is brine, which consists in water

and a salt which increases the density of it. According to the pression in the well's bottom, top, and a safety factor, it may be required that the control fluid has a particular density. Depending on such density, the required amount of salt according to the requirements must be added. The challenge with the salt as a densifier method is the water's saturation point. This means that the water will solubilize the salt until certain limit, thenceforth it will precipitate the salt added without increasing the density of the brine. Statistics concepts such as regression models and hypothesis testing will be used in order to obtain the statically significant models to associate the brine required to prepare a barrel of brine (Sodium formate and potassium chloride mix) with the required density.

**APROBACION DE LA TESIS**

Nombre Jurado: JAIME POLANIA PERDOMO

Firma:

Nombre Jurado: EDGAR ANDRES BERNAL CASTRO

Firma:

# DENSIFICACIÓN DE SALMUERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO

UNIVERSIDAD SURCOLOMBIANA

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ESPECIALIZACIÓN EN ESTADISTICA

NEIVA – HUILA

2021

# DENSIFICACIÓN DE SALMERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO

**IVAN GABRIEL MOLINA PERILLA**

**DIMITRY ANDRES RIASCOS MALAVER**

**UNIVERSIDAD SURCOLOMBIANA**

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES**

**ESPECIALIZACIÓN EN ESTADISTICA**

**NEIVA – HUILA**

**2021**

# DENSIFICACIÓN DE SALMUERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO

**IVAN GABRIEL MOLINA PERILLA**

**DIMITRY ANDRES RIASCOS MALAVER**

**Proyecto de grado para optar al título de: Especialista en estadística**

**DIRECTOR**

**Ing. Constanza Vargas Castellanos**

**UNIVERSIDAD SURCOLOMBIANA**

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES**

**ESPECIALIZACIÓN EN ESTADISTICA**

**NEIVA – HUILA**

**2021**

**DENSIFICACIÓN DE SALMUERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO**

**ÁREA DE INVESTIGACIÓN:**

**FLUIDOS DE CONTROL**

**PRESENTADO AL COMITÉ DE PROYECTOS DE GRADO DEL PROGRAMA DE ESPECIALIZACIÓN EN ESTADISTICA**

UNIVERSIDAD SURCOLOMBIANA
NIT: 891180084-2

ACREDITADA DE ALTA CALIDAD
Resolución 11233 / 2018 - MEN

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES**

**ESPECIALIZACIÓN EN ESTADÍSTICA**

*CARTA DE ACEPTACIÓN*

En calidad de Coordinador del Posgrado Especialización en Estadística, programa reconocido por el Ministerio de Educación Nacional mediante Resolución de Registro Calificado No. 3683 del 2 de marzo de 2018 y adscrito a la Facultad de Ciencias Exactas y Naturales de la Universidad Surcolombiana, me permito informar que el trabajo de investigación titulado: **"DENSIFICACIÓN DE SALMERA TIPO KCL Y NACOOH PARA FLUIDOS DE CONTROL DE POZO"** presentado por los estudiantes Ivan Gabriel Molina Perilla y Dimitry Andrés Riascos Malaver; es ACEPTADO como trabajo de grado para optar el título de Especialista en Estadística.

Para constancia se firma en la Ciudad de Neiva, a los doce (12) días del mes de julio del año 2021.

**JAIME POLANIA PERDOMO**
Coordinador

Vigilada Mineducación

Sede Central / Av. Pastrana Borrero - Cra. 1
Sede Administrativa / Cra. 5 No. 23 - 40
www.usco.edu.co / Neiva - Huila
PBX: 875 4753
PBX: 875 3686
Línea Gratuita Nacional: 018000 968722

## Dedicatoria

*Yo Ivan Gabriel Molina Perilla dedico este proyecto especialmente a mi madre Maria Isabel Perilla Suarez (Q.E.P.D), ya que sin ella nada de esto no hubiera sido posible; por todos los esfuerzos que realizo para educarme y hacerme la persona que hoy soy. Quisiera que pudiera estar presente conmigo en este momento, pero donde quiera que esté, sé que estará orgullosa, así como yo lo estoy de haber tenido una madre como ella. A mis abuelos Maria Ana Cecilia Suarez Castañeda y Jose Gerardo Perilla, porque fueron mi apoyo y mis segundos padres luego de la partida de mi madre, ellos estuvieron presentes en los mejores y peores momentos de mi vida, sin ellos, no podría haber llegado hasta este punto. A mis tíos Carlos Perilla Suarez y Fabio Nelson Perilla Suarez (Q.E.P.D), y a mi tía Disney Perilla Suarez por su apoyo incondicional durante toda mi vida, y su constante trabajo por hacerme mejorar día a día.*

*Yo Dimitry Andrés Riascos Malaver dedico este proyecto a toda mi familia, especialmente a mis padres Hector Riascos Vásquez y Blanca Cecilia Malaver Romero, a mis hermanos Hector Javier Riascos Malaver y Margarita Riascos Malaver quienes fueron mi apoyo fundamental y, además, la guía para ser la persona que soy hoy en día. A todos mis abuelos (QEPD) que siempre creyeron y confiaron que yo sería un buen profesional.*

## Agradecimientos

Agradecemos a Dios y a nuestras familias por guiarnos y darnos fortaleza para realizar este proyecto.

Agradecemos a PARKO SERVICES S.A. por brindarnos su apoyo e instalaciones para realizar este proyecto.

A la universidad Surcolombiana por permitirnos formarnos con los conocimientos en estadística necesarios para realizar el proyecto.

Al profesor Carlos Eduardo Alonso Malaver y a la ing. Constanza Vargas Castellanos, por guiarnos durante el desarrollo de este proyecto en los ámbitos estadísticos y de conocimientos técnicos en fluidos de control.

# Contenido

## Lista De Figuras

## Lista De Tablas

## Resumen

**Título**: Estandarización en las concentraciones para la densificación de salmueras tipo KCL y NACOOH para fluidos de control de pozo en las empresas prestadoras de servicio de fluidos.

**Descripción**: Durante la vida productiva de un pozo se requieren realizar intervenciones a este para reactivar o incrementar la producción de petróleo y gas; en las intervenciones es muy importante tener el control del pozo, esto significa que la presión de la formación no supere la presión de fondo de pozo ya que esto puede desencadenar en un descontrol de la presión y que los fluidos dentro del pozo salgan a superficie de forma precipitada, esto puede generar daños ambientales o incluso, algún tipo de lesión a personas. Para asegurar que el pozo esté controlado en todo momento se inyecta al pozo un fluido densificado que provea una columna hidrostática que supere la presión de fondo de pozo, a este fluido se le llama comúnmente fluido de control. El fluido usado más comúnmente es la salmuera, la cual consiste en agua y una sal que aumenta la densidad del agua (si se desea se pueden agregar anticorrosivos, surfactantes, entre otros aditivos). De acuerdo con la presión existente en el fondo del pozo, la presión en cabeza y un factor de seguridad, se puede requerir que el fluido de control tenga una determinada densidad, y de acuerdo con la densidad se utiliza la sal que se acomode mejor a los requerimientos. La dificultad que se tiene con la sal como método densificante es el punto de saturación del agua, esto quiere decir que el agua solubilizará la sal hasta cierto punto, de ahí en adelante precipitará la sal que se le agregue sin aumentar la densidad de la mezcla. En Colombia los productos más comúnmente usados por sus beneficios operacionales son el cloruro de sodio, cloruro de potasio, el formiato de sodio y el formiato de potasio, ya sea como producto puro o como mezcla entre ellos. Como punto de partida se abordarán los conceptos referentes a las sales de cloruro de potasio y sales de formiato, ya que normalmente el formiato de potasio se vende como producto

líquido debido a la complejidad de su preparación, también se partirá de la información que traen algunas tablas creadas por empresas de las cuales actualmente se tienen algunas diferencias representativas a nivel práctico a la hora del cálculo de las densidades requeridas. Se utilizarán conceptos estadísticos tales como modelos de regresión y pruebas de hipótesis para obtener modelos estadísticamente significativos.

Palabras claves: Salmuera, densidad, cloruro de potasio, formiato de sodio, correlación, pruebas de laboratorio, modelos de regresión

## Abstract

**Title:** Standardization in the concentration of the densification of KCL and NACOOH brines for the well's control fluids in the companies that provide the service.

**Description:** During the productive life of a well, it is required to make interventions to reactivate or increase the petroleum and gas production. In those interventions it is particularly important to make sure the pressure of the formation does not surpass the well's bottom pressure since it may cause the fluids inside the well to rise up to the surface hastily. This could cause environmental damages or even some injuries to people. In order to guarantee that the well is under control in every moment, it gets injected with a densificated fluid that provides a hydrostatic column which exceeds the well's bottom pressure. This fluid is called control fluid. The most commonly used fluid is brine, which consists in water and a salt which increases the density of it (anti-corrosives, surfactants, and other additives may be added). According to the pression in the well's bottom, top, and a safety factor, it may be required that the control fluid has a particular density. Depending on such density, the required amount of salt according to the requirements must be added. The challenge with the salt as a densifier method is the water's saturation point. This means that the water will solubilize the salt until certain limit, thenceforth it will precipitate the salt added without increasing the density of the brine. In Colombia, the most commonly used products due to the operational benefits are sodium chloride, potassium chloride, sodium formate and potassium formate, either as a pure product or as mixture of them. As starting point, concepts regarding the potassium chloride salts and formate salts will be addressed, since normally potassium formate is sold in a liquid state because of the complexity of its preparation. Furthermore, the information contained in tables created by some companies (which have some practical representative differences when making the calculus of the required

densities) will be taken into account. Statistics concepts such as regression models and hypothesis testing will be used in order to obtain the statically significant models.

**Key Words:** Brine, density, potassium chloride, sodium formate, correlation, lab tests, regression models

## Objetivos

- Realizar pruebas de laboratorio que permitan determinar a través de tablas o correlaciones la relación existente entre las concentraciones requeridas de sal, o volumen de fluido, y densidad final de la salmuera.
- Analizar la viabilidad económica de aumentar la densidad base de KCl en las mezclas de KCl+NaCOOH.

## Justificación

Actualmente no existen unas tablas suministradas por la academia para la preparación de salmueras tipo KCl (Cloruro de potasio), NaCOOH (Formiato de sodio) y mezclas entre ellas; algunas empresas han creado sus propias tablas, pero hay una variación muy grande entre las concentraciones usadas para alcanzar dichas densidades cuando se incluye el formiato de sodio, por tanto, se requiere por medio de pruebas de laboratorio y de campo generar unas tablas y/o correlaciones que al ponerse a prueba cumplan con repetibilidad y reproducibilidad. Generar una metodología clara para la preparación de las salmueras nombradas anteriormente permitirá mejorar los procesos de planeación y logística de procesos que las involucren, además de dar paso a posibles futuros trabajos que generen optimizaciones económicas.

## Introducción

A nivel académico hay poca información sobre las cantidades/concentraciones de la cantidad sal que se requiere para lograr densificar una salmuera hasta cierta densidad; algunas empresas han creado sus propias tablas, pero hay una variación muy grande en las concentraciones usadas para alcanzar dichas densidades cuando se incluye el formiato de sodio. Debido a esta problemática se requiere de tablas y/o correlaciones por medio de los cuales se establezca la concentración de sal y volúmenes de líquido necesarios para las distintas densidades requeridas, y esto se logrará a partir de pruebas de laboratorio repetitivas por medio de ensayo y error, y pruebas de campo para su validación.

Se piensa que estas dos correlaciones son lineales por tanto se calcularán dos puntos en las correlaciones, realizando varias repeticiones para confirmarlos, y con ellos se predecirán puntos intermedios y se observará si la densidad y volumen son los esperados. Además de esto se debe confirmar si hay diferencias representativas en la cantidad de sal necesaria al cambiar de proveedor teniendo en cuenta que los proveedores pasan un filtro inicial para lograr una salmuera que cumpla las condiciones de calidad requerida por las empresas que contratan el servicio de fluidos de control.

Con las correlaciones obtenidas se realizará un análisis económico para observar la viabilidad de en próximos estudios aumentar la cantidad de cloruro de potasio usado, aumentando la densidad base del cloruro de potasio, en las mezclas de cloruro de potasio y formiato de sodio en búsqueda de generar ahorros en el costo total de la química.

# 1. Marco teórico

## 1.1. Generalidades Estadísticas

### 1.1.1. Estadística Descriptiva

Una de las principales funciones de la estadística es la de organizar, resumir y describir una información. Esto se realiza a través de cuadros, gráficas y medidas, estás ultimas divididas principalmente en posición, forma y dispersión.

Las medidas de posición o tendencia central permiten determinar la posición de un dato o valor en un conjunto de datos, el cual es representativo del total de la información.

- Promedio: La mayoría de las veces que se habla del promedio se refiere a la media aritmética, y su función es la de resumir un grupo de datos, pero tiene como desventaja que es muy sensible a datos extremos (outliers).

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

- Mediana: Es un dato tal que divide el grupo de datos en 2, por tanto, solo es superado por el 50% del total de datos. No tiene una fórmula matemática, sino que se debe seguir un procedimiento para hallarlo:

  **1.** Si el número de observaciones es impar, y no está agrupado, se organiza de menor a mayor y se selecciona el dato que está en la posición n/2.

  **2.** Si el número de observaciones es par, y no está agrupado, se organiza de menor a mayor y se promedia los datos en la posición n/2 y (n/2)+1.

  **3.** Para datos agrupados discretos se busca el valor n/2 en la columna de frecuencia acumulada, si se encuentra directamente se promedia su valor de Xi con el siguiente, si no se encuentra en la columna de frecuencia acumulada se toma el dato puntual de Xi inmediatamente mayor a este.

**4.** Para datos agrupados continuos en intervalos se realiza una interpolación entre el límite inferior y superior del intervalo que contiene en su frecuencia acumulada el valor correspondiente o inmediatamente superior a n/2.

- Moda: Es el dato que más se repite, ya sea uno, dos o más veces en el caso de una distribución multimodal. Para una distribución simétrica se tiene que $\bar{x} = M_e = M_o$.
- Percentil: Tiene un concepto muy parecido a la mediana, con la principal diferente de que se busca dividir los datos en dos grupos no homogéneos. Por ejemplo, el percentil 10 correspondería al dato que solo es superado por el 10% de los datos y su procedimiento a seguir es igual al de la mediana buscando al dato (percentil)*n/100.

Las medidas de dispersión describen la variabilidad interna de los datos, entendiendo que, entre menos variabilidad más representativa es la medida de posición si la comparamos con otra distribución con su misma medida posición.

- Varianza poblacional

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \quad (2)$$

- Desviación estándar poblacional

$$\sigma = \sqrt{\sigma^2} \quad (3)$$

- Varianza muestral

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad (4)$$

- Desviación estándar muestral

$$s = \sqrt{s^2} \quad (5)$$

### 1.1.2. Distribuciones De Probabilidad Continuas

El término probabilidad es una medida de la creencia de que un evento futuro pueda ocurrir, y con esto poder realizar inferencias para ayudar a tomar decisiones. Este concepto es muy útil en áreas donde no se pueden predecir eventos futuros con total certeza, además de donde se tienen dudas sobre la exactitud de las mediciones. Para hacer los cálculos de probabilidades se realiza un experimento, proceso por el cual realizamos una observación, ya sea simple, que solo se compone de un evento, o uno compuesto, que es consecuencia o prosigue después de otro evento, con todas estas posibles mediciones generamos un conjunto denominado espacio muestral, y este es a su vez el dominio de una función a la que llaman variable aleatoria que son las posibles medidas del experimento. La variable aleatoria se divide en dos grandes grupos, la discreta si solo puede tomar un número finito o contablemente infinito de valores distintos, o continúa si puede tomar cualquier valor en un intervalo, siendo infinitos sus valores posibles. La distribución de probabilidad de una variable aleatoria discreta puede darse para todos sus posibles valores ya que son contables y la suma de todas sus probabilidades pueden llegar al 100%, pero esto no aplica a las continuas ya que al ser infinitos sus valores, cada evento no puede tener una probabilidad asignada, por tanto, sus probabilidades se definen para un grupo/rango de valores de la variable aleatoria. Para describir las distribuciones de probabilidad continuas muchas veces se hace uso de funciones que facilitan los cálculos, funciones de densidad, para estas funciones se cumple que el área bajo la curva entre dos puntos de la función es igual a la probabilidad de que la variable tome un valor entre ellos dos, esto último se puede definir como una integral, pero para facilitar cálculos y evitar realizar el procedimiento de esa integral en cada oportunidad que deseemos calcular la probabilidad se define la función de distribución de probabilidad acumulada la cual nos da la probabilidad de que la función tome valores menores al punto evaluado. Al evaluar el valor más extremo de una función de probabilidad acumulada se debe obtener el 100% de probabilidad que es el máximo posible.

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \quad (6)$$

$$f(x) \rightarrow \text{Función de probabilidad}$$

$$F(x) \rightarrow \text{Función de probabilidad acumulada}$$

$$x \rightarrow \text{Valor a evaluar}$$

$$P(a < x < b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a) \quad (7)$$

$$P(a < x < b) \rightarrow \text{Probabilidad de que } x \text{ tome un valor entre } a \text{ y } b$$

$$E(x) = \int_{-\infty}^{\infty} x * f(x)\,dx \quad (8)$$

$$E(x) \rightarrow \text{Valor esperado}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 * f(x)\,dx \quad (9)$$

$$V(x) = E(x^2) - (E(x))^2 \quad (10)$$

$$V(x) \rightarrow \text{Varianza}$$

$$DE(x) = \sqrt{V(x)} \quad (11)$$

$$DE(x) \rightarrow Desviación\ estandar$$

- Propiedades de la función de densidad

$$F(x) = \int_{-\infty}^{\infty} f(x)\, dx = 1 \quad (12)$$

$$F(x) = \int_{-\infty}^{x} f(x)\, dx > 0 \quad (13)$$

### *1.1.3.* Distribución De Probabilidad Normal

Es la función de probabilidad continua más usada y conocida, es la base de toda la teoría inferencial, de hecho, algunas distribuciones continuas se derivan de esta, como la t-student, Chi cuadrado y Fisher.

En esta la media, mediana y moda coinciden, distribución simétrica. A 1 desviación estándar de la media se contiene el 68,27% de los datos, a 2 desviación el 95,45% y a 3 desviaciones el 99,73% (Regla empírica).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \quad (14)$$

$$E(x) = \mu \quad (15)$$

$$V(x) = \sigma^2 \quad (16)$$

$$\sigma \rightarrow Desviación\ estandar \qquad \sigma > 0$$

$$\mu \rightarrow Media$$

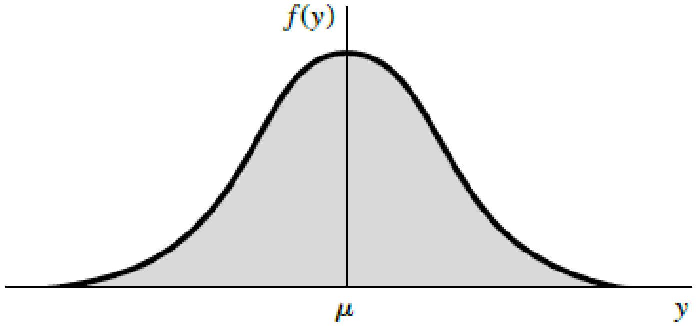


**Figura 1." Wackerly, D. (2009) Estadística matemáticas con aplicaciones 7 edición. Grafica función de densidad probabilidad normal Pág. 180 fig 4.10"**

La distribución es simétrica respecto a $\mu$, el área bajo la curva entre $[\mu, \infty)$ *es* 0,5. Numerosas variables continuas de fenómenos aleatorios tienden a comportarse probabilísticamente similar a esta distribución, además de que puede ser alcanzada de manera aproximada por otras distribuciones, algo descrito por el teorema del límite central. Por términos prácticos se genera una única distribución normal la cual se llama estándar, donde la $\mu = 0$ y $\sigma^2 = 1$, para esto se estandariza el valor de $\bar{x}$.

$$Z = \frac{\bar{x} - \mu}{\sigma} \quad (17)$$

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}} \quad (18)$$

Cuando lo que se analiza es una distribución muestral se tiene:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (19)$$

La integral de la función de probabilidad normal no se puede resolver de manera analítica, por tanto, para su solución se hace uso de integración numérica, esto quiere decir a partir de métodos numéricos, los cuales no se abarcan en este trabajo ya que el software estadístico lo calculan de manera automática.

#### *1.1.4.* Distribución T De Student

Para la distribución normal se conoce la desviación estándar poblacional y no importa el tamaño que se tome de la muestra para realizar inferencias, pero en la mayoría de los casos solo se posee solo una muestra de la población de interés. Cuando la muestra es lo suficientemente grande podemos reemplazar directamente la desviación típica poblacional por la muestral, pero solo es recomendable para un número de unidades muestrales mayor a treinta. Para muestras mayores a 30 se considera que la desviación estándar muestral es un buen estimador de la desviación estándar poblacional.

Cuando la muestra es pequeña, menor a 30, es poco posible que se hubieran incluido en ella valores extremos de la variable poblacional, y aunque la fórmula de desviación estándar muestral realiza una corrección para esto (Denominador es n-1 y no n en las ecuación (2) y (4)), la distribución de probabilidad de las medias muestrales no tiene un comportamiento igual a la distribución normal. Para tratar este problema surge la distribución t, más achatada y con más probabilidad en los extremos que la distribución normal, aunque se dice que no existe una distribución t como tal, sino una familia de distribuciones t, una por cada grado de libertad utilizado.

- Grados de libertad: “Los grados de libertad de un conjunto de observaciones, están dados por el número de valores que pueden ser asignados arbitrariamente, antes de que el resto de las variables queden completamente determinadas.” – Ronald Fisher

Para la distribución t de student debo asignar arbitrariamente el parámetro de la media poblacional, por tanto, se pierde un grado de libertad.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}\ con\ n - 1\ grados\ de\ libertad \quad (20)$$

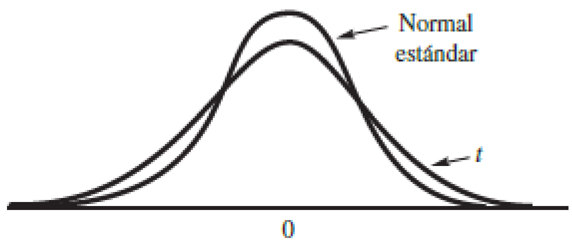


**Figura 2. "Wackerly, D. (2009) Estadística matemáticas con aplicaciones 7 edición. Grafica función de densidad probabilidad t de student VS normal estándar. Pag.360 fig 7.3"**

### *1.1.5.* Distribución Chi Cuadrado

Es la distribución más usada cuando el objetivo es monitorear la variabilidad de un proceso, es decir, controlar la calidad de los productos. Está distribución cuenta con un único parámetro, los grados de libertad, que es el parámetro de forma de la distribución. Para esta distribución se asigna arbitrariamente la desviación estándar poblacional, perdiendo un grado de libertad.

$$x^2 = \frac{(n-1)s^2}{\sigma^2}\ con\ n - 1\ grados\ de\ libertad \quad (21)$$

La distribución chi cuadrada no es simétrica y tiene dominio x>0, posee colas estrechas que se extienden a la derecha, esto quiere decir que es sesgada a la derecha.

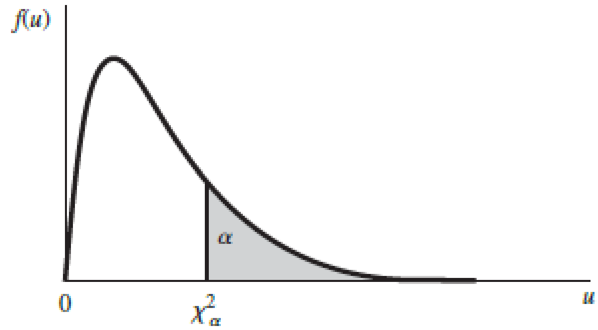


**Figura 3. “Wackerly, D. (2009) Estadística matemáticas con aplicaciones 7 edición. Grafica función de densidad probabilidad Chi cuadrado Pag.356 fig.7.2”**

### *1.1.6.* Distribución F De Fisher

Con esta distribución se pueden realizar inferencias sobre la varianza de dos grupos o poblaciones independientes y es una herramienta fundamental en análisis de regresión. Para esta distribución se asigna arbitrariamente las desviaciones estándar de las poblaciones del numerador y denominador, perdiendo un grado de libertad en cada uno.

$$F = \frac{S_1^2}{S_2^2} = \frac{Varianza\ mayor}{Varianza\ menor}\ con\ n_1 - 1\ gl\ en\ el\ númerador\ y\ n_2 - 1\ gl\ en\ el\ denominador \quad (22)$$

La distribución F de Fisher no es simétrica y tiene dominio x>0, posee colas estrechas que se extienden a la derecha, esto quiere decir que es sesgada a la derecha.

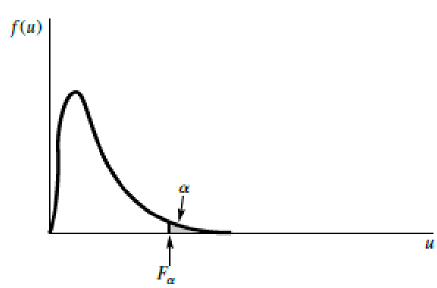


**Figura 4. “Wackerly, D. (2009) Estadística matemáticas con aplicaciones 7 edición. Grafica función de densidad probabilidad F de Fisher. Pág. 364 fig. 7.4”**

### 1.1.7. Distribuciones Muéstrales

Se debería considerar las distribuciones muestrales de $\bar{x}$ y $s^2$ como los mecánicos a partir de los cuales se pueden hacer inferencia de los parámetros poblacionales $\mu$ y $\sigma^2$, entendiendo los datos calculados sobre la muestra como un estadístico, lo que es una función de las variables aleatorias observadas en la muestra y algunas constantes conocidas, como es el tamaño de muestra. La bondad del ajuste de esos parámetros inferidos de esa distribución depende del comportamiento de las variables aleatorias sobre la cual se realiza la inferencia. La distribución muestral de los estadísticos proporciona un modelo teórico para la distribución de frecuencias relativas de los valores probables que se observarían del estadístico si se tomaran muestras repetidas, en otras palabras, en la distribución que resulta cuando un experimento se lleva a cabo una y otra vez, siempre con el mismo tamaño de muestra, y resultan los diversos valores de $\bar{x}$ y $s^2$, describiendo la variabilidad de los promedios muestrales alrededor de la media de la población.

### *1.1.8.* Distribución Muestral De Medias Y El Teorema Del Límite Central

De una población normal con media $\mu$ y varianza $\sigma^2$ se toma una muestra aleatoria de n observaciones. Cada observación $x_i, i = 1,2, \dots, n$ de la muestra aleatoria tendrá la misma distribución normal donde la media poblacional es igual a la media de la distribución de todas las medias muestrales posibles y la varianza muestral (de todas las medias muestrales) será igual a la varianza poblacional dividido por el tamaño de la muestra. Luego las muestras seguirás una distribución:

$$\bar{x} = (Media, Varianza) \sim Normal\left(\mu, \frac{\sigma^2}{n}\right) \quad (23)$$

Si $\bar{x}$ es la media de una muestra aleatoria de tamaño n, tomada de una población de con media $\mu$ y varianza $\sigma^2$ finita, entonces, la forma límite de la distribución muestral:

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma_x^2}} \quad (24)$$

$$\sigma_x^2 = \frac{\sigma^2}{n} \quad (25)$$

$$\lim_{n\to\infty} \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx Normal(0,1) \quad (26)$$

Esto último es lo es llamado el teorema del límite central, permitiendo realizar aproximaciones para la distribución muestral de la variable aleatoria de interés sin preocuparse por la distribución de la poblacional de la cual se tome la muestra ya que esta se aproximará a la normal. Si el tamaño de la muestra es grande.

Este concepto es muy importante ya que lleva a las distribuciones de muestra como se relaciona la distribución normal con las distribuciones anteriormente vistas.

Partiendo de la ecuación (17), con variables aleatorias normales estándar e independientes:

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2 tiene\ dist.\ x^2\ con\ n\ gl \quad (27)$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2\ tiene\ dist.\ x^2\ con\ n-1\ gl \quad (28)$$

Esta distribución es usualmente usada para realizar pruebas sobre la distribución de la varianza muestral, o cuando se desea saber si un grupo de datos sigue cierta distribución de probabilidad.

Sea Z una variable aleatoria normal estándar y sea W una variable con distribución $x^2$ con n grados de libertad. Entonces, si W y Z son independientes:

$$T = \frac{Z}{\sqrt{\frac{W}{v}}} \; tiene\ dist.t\ de\ student\ con\ v\ gl \quad (29)$$

Si W sigue una distribución $x^2$ como la ecuación (28) y Z una normal como la ecuación (26):

$$T = \frac{Z}{\sqrt{\frac{W}{v}}} = \frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)s^2}{\sigma^2}}{n-1}}} = \frac{\bar{Y}-\mu}{s/\sqrt{n}} \; tiene\ dist.t\ de\ student\ con\ n-1\ gl \quad (30)$$

Esta distribución es usualmente usada cuando se quiere realizar pruebas sobre la distribución de la media poblacional, pero se desconoce la varianza poblacional y se debe estimar con la varianza muestral.

Sean $W_1$ y $W_2$ variables aleatorias independientes con distribución $x^2$, con $v_1$ y $n_2$ grados de libertad, respectivamente. Entonces se dice que:

$$F = \frac{\frac{W_1}{v_1}}{\frac{W_2}{v_2}} \; tiene\ dist.F\ con\ v_1\ gl\ en\ el\ n\acute{u}merador\ y\ v_2\ gl\ en\ el\ denominador \quad (31)$$

$$F = \frac{\frac{W_1}{v_1}}{\frac{W_2}{v_2}} = \frac{\frac{\frac{(n_1-1)s^2}{\sigma^2}}{n_1-1}}{\frac{\frac{(n_2-1)s^2}{\sigma^2}}{n_2-1}}$$
$$= \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}\ tiene\ dist.F\ con\ n_1 - 1\ gl\ en\ el\ númerador\ y\ n_2 - 1\ gl\ en\ el\ denominador \quad (32)$$

Esta distribución es usualmente usada cuando se quiere realizar pruebas la relación entre dos la varianza de dos poblaciones normales, pero se desconoce las varianzas poblacionales y se debe estimar con las varianzas muestrales.

#### 1.1.9. Prueba De Hipótesis E Intervalos De Confianza

Las pruebas de hipótesis o de significación tienen como objetivo el realizar la validez de ciertas afirmaciones sobre valores estadísticos de la población, parámetros, a partir de valores estadísticos de la muestra, estadístico.  Esto se hace para ver si la variación de la muestra con respecto a la población es estadísticamente significativa, y no es solamente por el azar.

Las pruebas de hipótesis estadísticas inician a partir del planteamiento de una afirmación sobre un parámetro de la población, la mayoría de las veces con intención de confirmar si algún cambio encontrado en la población a partir de una muestra es indicio de un cambio en esta. La afirmación se divide en dos hipótesis, la primera es la hipótesis nula, que indica que continuidad en lo que creemos del parámetro, contra la segunda, la hipótesis alternativa, que indica un cambio en lo que creemos del parámetro, siendo contraria a la hipótesis nula.

Como estás hipótesis son refutadas con muestras y no directamente con la población, existe el riesgo de llegar a una conclusión incorrecta, dando lugar a dos tipos de error:

- Error tipo I: Aceptar la hipótesis nula cuando está es en realidad Falsa, también llamado falso verdadero.
- Error tipo II: Rechazar la hipótesis nula cuando es en realidad verdadera, también llamada falso.

Las hipótesis pueden plantear un supuesto sobre algún parámetro poblacional (Media, varianza, proporción, etc…), ya sea igualdad con un cierto valor, o una relación de menor/mayor, también la forma de la distribución de probabilidad, entre otras. Estas hipótesis se ubican sobre la distribución de probabilidad del parámetro a evaluar, generando dos zonas, la primera es de aceptación, donde afirmamos que no hay suficiente evidencia para probar que la hipótesis nula es falsa, y otra de rechazo, donde concluimos que se puede decir que la hipótesis nula es falsa, aceptando la afirmación de la hipótesis alternativa.

### ***1.1.10.*** **Prueba Bilateral Y Unilateral**

A su vez la zona de rechazo puede tener dos formas dependiendo del tipo hipótesis a probar:

- Prueba de hipótesis unilateral: La zona de rechazo está ubicada solo en uno de los dos extremos de la distribución, siempre usada cuando se prueba la bondad del ajuste a cierta distribución (Chi cuadrado o prueba F), ubicándose en esta última siempre en la zona derecha de la distribución (Cola derecha), o una relación menor/mayor con un parámetro de la población, siendo de cola izquierda y derecha respectivamente, entendiendo que es la relación ubicada en la hipótesis alternativa.
- Prueba de hipótesis bilateral: La zona de rechazo se ubica en los dos extremos de la distribución, siempre usada cuando se prueba una igualdad con un parámetro de la población.

### *1.1.11.* Nivel De Significancia Y Puntos Críticos

El nivel de significancia es un nivel dado por el investigador como el máximo probabilidad de cometer un error tipo I. Se simboliza con la letra griega Alpha (∝), siendo 1%, 5% o 10% sus valores más comunes, hablando de resultados poco significativo, significativo y altamente significativo, respectivamente, cuando esos valores son usados.

El nivel de significancia escogido será el usado para establecer la ubicación del, o de los, puntos críticos m, lugar en la distribución de probabilidad donde se separa la zona de rechazo y la de aceptación. Estos puntos críticos deben ser tales que asignen a la zona de rechazo una probabilidad igual a la del nivel de significancia.

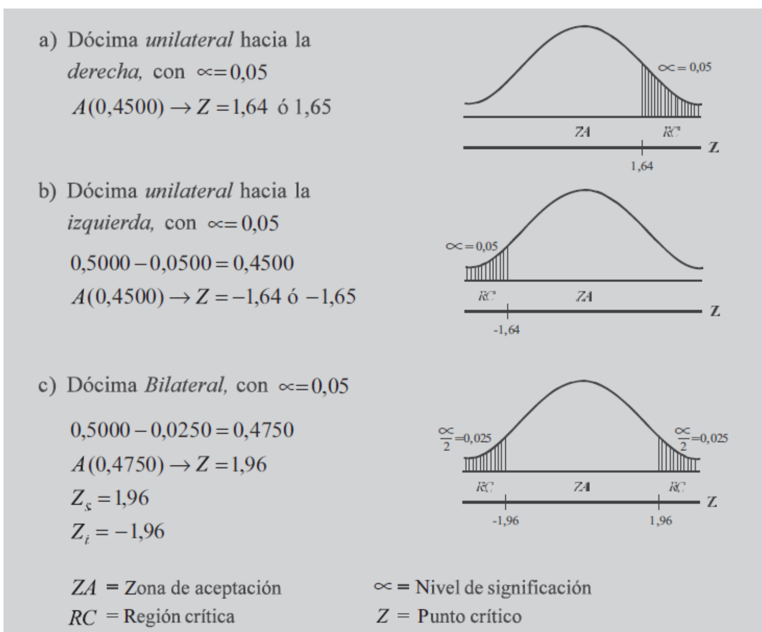


**Figura 5. "Ciro, B. (2012) Estadística y muestreo edición 13. Prueba de hipótesis en una distribución estándar (Z). Pág. 329 fig. 8.1"**

### *1.1.12.* Procedimiento Prueba De Hipótesis

Se enumera el paso a paso para resolver una prueba de hipótesis dado por Ciro Martinez en su libro de estadística y muestreo:

1) Formular la hipótesis nula y la alternativa

Para el planteamiento de algún supuesto sobre un parámetro poblacional se tiene:

$$H_0 : \theta = \theta_0.$$
$$H_a : \begin{cases} \theta > \theta_0 & \text{(alternativa de cola superior).} \\ \theta < \theta_0 & \text{(alternativa de cola inferior).} \\ \theta \neq \theta_0 & \text{(alternativa de dos colas).} \end{cases}$$

**Figura 6. "Wackerly, D. (2009) Estadística matemáticas con aplicaciones 7 edición. Hipótesis nula y alternativa. Pág. 500**

Acá se reemplazan los parámetros de interés, como la media, varianza o proporción poblacional, o la comparación entre estos mismos parámetros para dos poblaciones.

En el caso de la bondad de ajuste a cierta distribución, u otra hipótesis, se plantea una hipótesis nula y su contraria como la alternativa, por ejemplo, la prueba chi cuadrado en una tabla de contingencia coloca como si hipótesis nula la independencia de dos variables analizada, y la hipótesis alternativa la dependencia.

2) Se estable el riesgo: Nivel de significancia % ($\propto$)
3) Se establecen ciertos supuestos
    - La muestra es aleatoria.
    - La población es normal
    - Otros: La varianza poblacional es conocida (en la mayoría de los casos no se conoce, por lo tanto, debe ser estimada).
4) Se calcula el respectivo estadístico de prueba y su valor P:

El valor P se define como la probabilidad de que un valor estadístico calculado sea posible dada una hipótesis nula cierta. En una prueba de cola superior o derecha corresponde a la probabilidad en la distribución de obtener un valor superior al estadístico, en una prueba de cola inferior o izquierda corresponde a la probabilidad en la distribución de obtener un valor inferior al estadístico y en una prueba bilateral corresponde al doble de la probabilidad en la distribución de obtener un valor superior al valor absoluto del estadístico.

- Distribución de medias muestrales ($H_o:\ \mu = \theta_o$):

$$Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \ \ \sigma^2 conocida \ (33)$$

$$Z = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \ \ \sigma^2 \ desconocida \ muestra \ grande \ (34)$$

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \ \ \sigma^2 \ desconocida \ muestra \ pequeña, n - 1 \ gl \ (35)$$

- Distribución de varianzas ($H_o:\ \sigma^2 = \theta_o$):

$$x^2 = \frac{(n-1)s^2}{\sigma^2} \ con \ n - 1 \ grados \ de \ libertad \quad (36)$$

- Distribución de relación entre varianzas ($H_o:\ \sigma_1^2 = \sigma_2^2$):

$$F = \frac{S_1^2}{S_2^2} = \frac{Varianza \ mayor}{Varianza \ menor} \ con \ n_1 - 1 \ gl \ en \ el \ númerador \ y \ n_2 - 1 \ gl \ en \ el \ denominador \quad (37)$$

Nota: Otras pruebas de hipótesis serán estudiadas durante el análisis de los modelos de regresión.

5) Determinar los valores críticos y sus regiones de rechazo:

A partir del nivel de significancia se buscar un valor crítico que separará la región de rechazo y aceptación:

- Prueba de hipótesis cola superior: El punto crítico es aquel que tiene una probabilidad de obtener valores por encima de él, igual al nivel de significancia. Zona de rechazo = $[Critico, \infty)$
- Prueba de hipótesis cola inferior: El punto crítico es aquel que tiene una probabilidad de obtener valores por debajo de él, igual al nivel de significancia. Zona de rechazo = $(-\infty, Critico]$
- Prueba bilateral: Los puntos críticos son aquellos cuales por debajo del valor inferior y por encima del superior tienen una probabilidad total igual al nivel de significancia. Zona de rechazo $(-\infty, Critico\ inferior] \cup [Critico\ superior, \infty)$

6) Tomar la decisión estadística:

Si el estadístico se encuentra en la zona de aceptación o su valor P es mayor al nivel de significancia afirmamos que no se puede probar que la hipótesis nula es falsa, si el estadístico se encuentra en la zona de rechazo o su valor P es menor al nivel de significancia se afirma que la hipótesis alternativa es correcta.

7) Concluir y tomar una decisión:

Conociendo la decisión que se tomó en base a las hipótesis se concluye en relación con la afirmación que se había generado, igualdad o relación entre los parámetros estudiados.

#### 1.1.13. Modelos De Regresión

Un modelo matemático es una relación funcional entre variables donde un conjunto de variables de entrada (X) se relaciona con una variable de salida (Y). Un modelo de regresión es un procedimiento inferencial donde a través de unas variables independientes se trata de predecir un valor de una variable dependiente, por tanto, la variable dependiente está en función de las independientes. El modelo más simple, pero uno de los más poderosos es el lineal:

$$y = \beta_0 + \beta_1 x \quad (38)$$

En muchos casos las diferencias con el modelo pueden ser atribuibles a la variabilidad natural del proceso y no tiene explicación matemática, a esto se le denomina el error aleatorio ($\epsilon$)

$$y = \beta_0 + \beta_1 x + \epsilon \quad (39)$$

El análisis de regresión es apropiado en situaciones donde se sospecha o se asume que una variable está relacionada a una o más medidas hechas usualmente en un mismo individuo (objeto). El objetivo del análisis es usar los datos, valores observados de las variables, para estimar la forma de la relación.

Los errores son considerados como variables aleatorias no observables y usualmente nos referimos a ellos como errores aleatorios o residuales aleatorios. Los supuestos que se hacen sobre estas variables son:

- El valor esperado o promedio es cero
- La varianza es constante de observación a observación.
- La covarianza entre $\epsilon_i$ y $\epsilon_j$ con $i \neq j$ es igual a cero
- Para efectos de construir intervalos de confianza y probar hipótesis sobre los parámetros del modelo se asume que los errores tienen distribución normal.

***1.1.14.*** **Estimación Del Modelo**

La estimación por mínimos cuadrados para los parámetros del modelo involucra minimizar la suma de cuadrados de las desviaciones de los valores observados y su valor esperado (minimizar los errores), en otras palabras, se debe minimizar:

$$\epsilon'\epsilon = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left[Y_i - (Y_i)\right]^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

**Figura 7. "Téllez, C. (2016) Modelos estadísticos lineales con aplicación en R. Estimación mínimos cuadrados. Cap. 1.2.1"**

Los valores que minimizan $\epsilon'\epsilon$ son los estimadores mínimos cuadrados de $\beta_0$ y $\beta_1$, los cuales serán denotados por $\widehat{\beta_0}$ y $\widehat{\beta_1}$. La minimización se logra usando el cálculo diferencial de la manera usual:

- Derivando con respecto a $\beta_0$ y $\beta_1$ e igualando a cero.
- Las ecuaciones encontradas serán escritas en términos de $\widehat{\beta_0}$ y $\widehat{\beta_1}$
- Las soluciones para $\widehat{\beta_0}$ y $\widehat{\beta_1}$ son los estimadores mínimos cuadrados.

Omitiendo los pasos y demostraciones se obtiene:

$$S_{xy} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) \quad (40)$$

$$S_{xx} = \sum (X_i - \bar{X})^2 \quad (41)$$

$$S_{yy} = \sum (Y_i - \bar{Y})^2 \quad (42)$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \frac{cov(X,Y)}{var(X)} \quad (43)$$

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X} \quad (44)$$

Los estimadores de $\beta_0$ y $\beta_1$, lo que quiere decir que el valor esperado de $\widehat{\beta_0}$ y $\widehat{\beta_1}$ es igual a los parámetros a estimar $\beta_0$ y $\beta_1$, respectivamente.

El procedimiento de mínimos cuadrados no proporciona una estimación para la varianza de los errores, por analogía, usando $\widehat{y_i}$ como estimador de $E(Y_i)$, se puede estimar $\widehat{\sigma^2} = E[Yi - E(Y_i)]^2$, a partir de los datos mediante:

$$\widehat{\sigma^2} = s = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n-2} = \frac{\sum_{i=1}^{n}\left(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_i\right)^2}{n-2} = \frac{S_{yy} - \widehat{\beta_1}S_{xy}}{n-2} \quad (45)$$

Se divide por n-2 porque para para calcular $\widehat{y_i}$ se tienen dos parámetros estimados perdiendo dos grados de libertad, además logrando que $\widehat{\sigma^2}$ sea un estimador insesgado de $\sigma^2$.

Para medir numéricamente el ajuste del modelo de regresión comúnmente se usa el coeficiente determinación denotado por $R^2$, además que usar su raíz para medir el grado de correlación entre las variables $R$, llamado coeficiente de correlación.

$$R^2 = \frac{SCR}{SCT} = \left(\frac{S_{xy}}{\sqrt{S_{xx} * S_{yy}}}\right)^2 \quad (46)$$

$$R = \frac{S_{xy}}{\sqrt{S_{xx} * S_{yy}}} \quad (47)$$

### *1.1.15.* Verificación De Supuestos Sobre Los Errores

Los errores no son conocidos ya que no sabemos la verdadera relación entre las variables, pero los residuos del modelo si se pueden calcular.

- Media cero ($H_o$: $\mu = 0$)

Para probar este supuesto se hace uso de la prueba t:

Se calculan los residuos como la diferencia entre el dato de la variable dependiente y el valor estimado por el modelo para esa variable dependiente con sus respectivas variables independientes

$$\hat{\epsilon} = y_i - \widehat{y_i} \quad (48)$$

Luego se hace uso de la ecuación (35) para realizar la prueba de hipótesis sobre los errores.

- Prueba de normalidad ($H_o$: $Los\ datos\ sigun\ una\ distribución\ normal$):

Para probar este supuesto se hace uso de la prueba de Shapiro Wilks:

Se calculan los residuales de manera similar a la prueba anterior, luego se organizan de menor a mayor los datos, luego se calcula la diferencia entre este dato y el promedio de los datos y se eleva al cuadrado. A continuación, se lee los datos de la tabla de Shapiro Wilks para la cantidad de datos. En transcurso, se organizan los datos de mayor a menor y se calcula diferencia entre cada dato correspondiente entre los datos de menor a mayor y mayor a menor, pero solo los primeros p datos, donde p corresponde al número de coeficientes de Shapiro Wilks para esa cantidad de datos, y a esta diferencia se calcula la sumatoria de todos esos datos y eso se eleva al cuadrado. Después de esto se calcula la sumatoria de la multiplicación entre la diferencia y los coeficientes de Shapiro Wilks, y se eleva esto al cuadrado, y con estos dos se calcula el estadístico de prueba:

$$SW = \frac{\sum(X_i - \bar{X})^2}{(\sum(a_i) * (X_i - X_{i\,orden\,inverso}))^2} \quad (49)$$

Este estadístico se busca en la tabla de significancia para esa cantidad de datos obteniendo el Valor P. Si el valor P es mayor al nivel de significancia elegido para la prueba de hipótesis, no se puede afirmar que los datos no sigan una distribución normal.

- Prueba de varianza constante ($H_o$: $Homocedasticidad - Varianza\ constante$):

Para probar este supuesto se hace uso de la prueba de Breusch Pagan:

Se calculan los residuales de manera similar a la prueba anterior y se calcula $\widehat{\sigma^2}$, también llamado cuadrado medio del error (CME). Luego se ajusta la regresión $\frac{\hat{\epsilon}}{\widehat{\sigma^2}} = \delta Z + \mu$ y se obtiene la suma de cuadrados de la regresión (SCR). Se calcula el estadístico de prueba:

$$F_{BP} = \frac{SCR}{2} \quad (50)$$

Este estadístico de prueba sigue una distribución chi cuadrado con q grados de libertad, donde q es el número de variables representadas por Z.

- Prueba de independencia/autocorrelación ($H_o$: $\rho = 0\ No\ correlación$ ):

Para probar este supuesto se hace uso de la prueba de Durbin Watson:

En esta prueba se ajusta el modelo de regresión sin intercepto $e_t = \rho * e_{t-1} + \varepsilon$ y se mide la significancia de $\rho$ mediante una prueba t. Si $\rho$ es estadísticamente distinto de cero, entonces se concluye que los errores están correlacionados.

$$DW = \frac{\sum_{i=2}^{N}(\widehat{e_t} - \widehat{e_{t-1}})^2}{\sum_{i=1}^{N} \widehat{e_t}^2} \quad (51)$$

### *1.1.16.* Significancia Del Modelo

Si ya se verificó que el modelo cumple los supuestos sobre los errores, se procede a verificar si el modelo en general es significativo, y sus predicciones pueden ser tomadas en cuenta como buenas inferencias.

Nota: N corresponde al número de parejas de datos usadas para estimar la regresión.

- Prueba de hipótesis para la correlación con análisis de varianza F

Esta prueba de hipótesis permite determinar si la pendiente de modelo lineal es diferente a cero, o, en otras palabras, existe relación o dependencia lineal entre las variables. Para el cálculo del estadístico F se procede a llenar la siguiente tabla:

$$H_o:\ \beta = 0$$

$$H_1:\ \beta \neq 0$$

| Causa | Suma de los cuadrados | Grados de libertad | Cuadrados medios | F |
|---|---|---|---|---|
| Regresión | $SCR = \sum(\hat{y} - \bar{y})^2$ | 1 | $CMR = \frac{SCR}{1}$ | $\frac{CMR}{CME}$ |
| Residual | $SCE = \sum(y - \hat{y})^2$ | $n - 2$ | $CME = \frac{SCE}{n-2}$ | |
| Total | $SCR + SCE$ | $n - 1$ | | |

El estadístico F sigue una distribución F de Fisher con 1 grado de libertad en el numerador y n-2 grados de libertad en el denominador. Si el valor P (Cola derecha) es mayor que el nivel de significancia se afirma que no hay suficiente evidencia para probar que no hay dependencia lineal entre las variables.

- Prueba de hipótesis para la pendiente

Acá se prueba si la pendiente calculada para el modelo es diferente a cero, y, por tanto, significativa.

$$H_o:\ \beta_1 = 0$$

$$H_1:\ \beta_1 \neq 0$$

$$t = \frac{\widehat{\beta_1} - 0}{\widehat{\sigma^2}\sqrt{\frac{1}{S_{xx}}}} \quad (52)$$

El estadístico t sigue una distribución t de student con n-2 grados de libertad. Si el valor P (Dos colas) es mayor que el nivel de significancia se afirma que no hay suficiente evidencia para probar que la pendiente no es significativa.

- Prueba de hipótesis para el intercepto

Acá se prueba si el intercepto calculada para el modelo es diferente a cero, y, por tanto, significativo.

$$H_o:\ \beta_0 = 0$$

$$H_1:\ \beta_0 \neq 0$$

$$t = \frac{\widehat{\beta_0} - 0}{\widehat{\sigma^2}\sqrt{\frac{\sum x_i^2}{n * S_{xx}}}} \quad (53)$$

El estadístico t sigue una distribución t de student con n-2 grados de libertad. Si el valor P (Dos colas) es mayor que el nivel de significancia se afirma que no hay suficiente evidencia para probar que el intercepto no es significativo.

### 1.1.17. Diseño Experimental

En el caso de estudio de este trabajo se realizaron experimentos para caracterizar procesos, realizando una serie de pruebas en las que se hicieron cambios en las variables de entrada de un proceso para observar las razones de los cambios en la respuesta de salida.

Para entender el modelo del diseño experimental se deben tener claro ciertos conceptos:

- Factor: Variables controlables estudiadas dentro del experimento.
- Bloque: Variable que afecta los resultados del experimento, pero no atrae el interés principal dentro de la investigación.
- Nivel: Valor que puede tomar un factor.
- Tratamiento: Combinación de niveles de los factores.
- Unidad experimental: Unidad sobre la cual se toman mediciones.
- Replica: Observación de un tratamiento sobre diferentes unidades experimentales.
- Corrida: Proceso generado para obtener una observación sobre una unidad experimental.

La finalidad del diseño experimental de este trabajo es determinar que variables tienen influencia sobre las variables respuesta.

#### *1.1.18.* Diseño De Bloques Completo Balanceado

Se tiene a tratamiento y b bloques, donde todos los tratamientos se observan bajo todos los b bloques. Durante la fase de planeación del experimento se realizó una asignación aleatoria de los tratamientos dentro de cada bloque con un esquema de aleatorización para remover el sesgo de asignar tratamientos a unidades experimentales favorables y minimizando los efectos de factores no controlables, garantizando el supuesto de independencia estadística.

- Modelo estadístico

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad i = 1, \dots, a \quad j = 1, \dots, b \quad (54)$$

$$\mu \rightarrow Efecto\ medio\ común\ a\ todas\ las\ observaciones$$

$$\tau_i \rightarrow Efecto\ de\ cada\ tratamiento$$

$$\beta_j \rightarrow Efectod\ e\ cada\ bloque$$

$$\epsilon_{ij} \rightarrow Error\ aleatorio$$

$$y_{ij} \rightarrow Observación\ para\ el\ tratamiento\ i\ en\ el\ bloque\ j$$

$$y_i \quad \bar{y}_i. \rightarrow Total\ y\ promedio\ para\ el\ tratamiento\ i$$

$$y_j \quad \bar{y}_j. \rightarrow Total\ y\ promedio\ para\ el\ bloque\ i$$

$$\bar{y}.. \rightarrow Promedio\ general$$

- Estimación de los parámetros del modelo

Este modelo no tiene un estimador de parámetro único, pero bajo una restricción se puedes realizar algunas:

$$\sum \tau_i = 0 \; y \; \sum \beta_j = 0 \;\; Restricciones \quad (55)$$

$$\hat{\mu} = \bar{y}.. \quad (56)$$

$$\hat{\tau_i} = \bar{y}_i. - \bar{y}.. \quad (57)$$

$$\hat{\beta_j} = \bar{y}_j. - \bar{y}.. \quad (58)$$

- Estimados y residuales

$$\widehat{y_{ij}} = \bar{y}.. + \bar{y}_i. - \bar{y}.. + \bar{y}_j. - \bar{y}.. = \bar{y}_i + \bar{y}_j - \bar{y}.. \quad (59)$$

$$\hat{\epsilon} = y_{ij} - \widehat{y_{ij}} = y_{ij} - \bar{y}_i. - \bar{y}_j. + \bar{y}.. \quad (60)$$

- Ajuste del modelo

Para determinar si el tratamiento y bloques son significativos:

Nota: N corresponde al número de observaciones.

| Causa | Suma de los cuadrados | Grados de libertad | Cuadrados medios | F |
|---|---|---|---|---|
| Tratamientos | $SCT_r = b\sum_{i=1}^{a}(\bar{y}_i. - \bar{y}..)^2$ | $a-1$ | $CMT_r = \frac{SCT_r}{a-1}$ | $\frac{CMT_r}{CME}$ |
| Bloques | $SCB = a\sum_{j=1}^{b}(\bar{y}_j. - \bar{y}..)^2$ | $b-1$ | $CMB = \frac{SCB}{b-1}$ | $\frac{CMB}{CME}$ |
| Error | $SCE = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_i. - \bar{y}_j. + \bar{y}..)^2$ | $(a-1)(b-1)$ | $CME = \frac{SCE}{(a-1)(b-1)}$ | |
| Total | $SCT_r + SCB + SCE$ | $N-1$ | | |

Los estadístico F siguen una distribución F de Fisher, para los tratamientos se tiene a-1 grados de libertad en el numerador y (a-1)*(b-1) grados de libertad en el denominador, y para los bloques b-1 grados de libertad en el numerador y (a-1)*(b-1) grados de libertad en el denominador. Si el valor P (Cola derecha) es mayor que el nivel de significancia se afirma que no hay suficiente evidencia para probar que la respectiva variación estudiada es significativa. Si la causa de variación de los tratamientos es significativa se afirma que al menos uno de los tratamientos es diferente a los otros, en el caso de los bloques el análisis no suele ser relevante, pero se podría afirmar que, si su variación es significativa, estuvo bien plantear el diseño experimental separado por estos bloques.

## 1.2. Generalidades De Salmueras

### 1.2.1. ¿Qué es la presión?

La presión es la fuerza que se ejerce en una cierta unidad de área, y en la industria del petróleo es normalmente medida en la cantidad de libras por cada pulgada cuadrada (psi). Estas presiones en la industria son causadas por fluidos, el yacimiento petrolífero, fricciones o generadas mecánicamente. Las presiones de interés en este trabajo son las causadas por el yacimiento, que son las que se quieren controlar, y las causadas por los fluidos, con lo que se controlarán.

Un fluido es aquel elemento que se encuentra en un estado de la materia líquido o gaseoso, siendo de interés para este trabajo el fluido más abundante en la tierra, el agua. Los fluidos generan una cierta presión sobre lo que se apoyan, en el caso de un pozo petrolero, ejercen una presión sobre el recipiente que lo contiene, el pozo, siendo esta presión mayor a medida que más profundo es medido con respecto a la superficie del fluido, entendiendo esto como la columna de fluido, medido en pies (ft). La presión que es causada por el fluido también depende de la densidad del fluido, que es la cantidad de masa de fluido por cada unidad de volumen, entendiendo que el agua es incompresible, por tanto, su variación no es mucha a cambios grandes de temperatura o presión, y en la industria normalmente es medio en libras de fluido por cada galón de volumen (ppg). A mayor densidad, mayor es la presión que ejerce.

Nota: Se debe entender que el fluido se mueve por una diferencia de presión, tendiendo a fluir hacia el punto de menor presión.

### *1.2.2.* **Gradiente De Presión**

Como se dijo anteriormente, la presión ejercida por un fluido es función de dos variables, sin incluir la gravedad, la altura de columna de fluido, o profundidad, y la densidad del fluido, asumiendo esta última como una constante. Como la función tiene dos constantes, se generó un concepto que relaciona la presión que ejerce el fluido por cada pie de profundidad, el cual es llamado gradiente de presión, normalmente medido en libras por pulgada cuadrada, por cada pie de profundidad (psi/ft).

$$\frac{lb}{gal} * \frac{7{,}48\, gal}{1\, ft^3} * \left(\frac{1\, ft}{12\, in}\right)^2 = 0{,}052\,\frac{lb}{in^2} * \frac{1}{ft} = \frac{0{,}052\, psi}{ft} \quad (61)$$

Este factor de conversión dice que por cada libra por galón de densidad del fluido se espera un gradiente de presión de 0,052. Por ejemplo, para el agua se tendría

$$\Delta P = 8{,}33 * 0{,}052\,\frac{psi}{ft} = 0{,}433\,\frac{psi}{ft} \quad (62)$$

Por cada pie de profundidad vertical de agua se espera que la presión aumente 0,433 psi.

### *1.2.3.* Profundidad Vertical VS Medida

Por diversos motivos fuera del alcance de este libro, los pozos petroleros no son totalmente verticales, por lo tanto, la distancia que se mide desde superficie no es igual a la distancia vertical hasta esta misma, sucediendo esto porque las herramientas de perforación, entre otras, solo miden la distancia recorrida, llamada profundidad medida. Esto es importante porque la formula anterior de gradiente, entre otras, usan la profundidad vertical, entendiendo que antes de usar las formulas debe calcular la profundidad vertical por medio de trigonometría, lo cual no se explica acá.

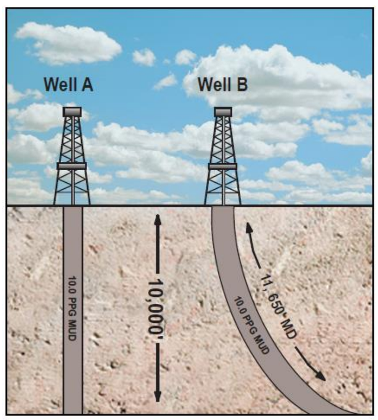


**Figura *8*. "Institution Well Control International. (2017) Manual de control de pozos WCS. Profundidad vertical VS profundidad medida Cap. 1"**

### *1.2.4.* Presión De Formación

Cuando se perfora un pozo petrolero y se llega al punto de interés, el yacimiento de petróleo, este tiene una presión confinada en el espacio poroso de la roca, presión causada por el peso de la sobrecarga de las capas de rocas superiores, y los fluidos contenida en esta.

### *1.2.5.* Presión Hidrostática

Se define como presión hidrostática a la presión causada por una columna de fluido debido a su densidad, la altura vertical.

$$PH = 0{,}052 * profundidad\ vertical\ (ft) * densidad\ del\ fluido\ (ppg) \quad (63)$$

Los pozos petroleros para controlar el fluido del yacimiento suelen estar llenos de fluido que generen esta presión hidrostática.

### 1.2.6. Densidad Equivalente

Cualquier aumente presión puede generar un aumento en la presión total medida en un punto, y si esta presión es conocida, se puede realizar el cálculo de una densidad equivalente si se quisiera alcanzar esa misma presión a esa misma profundad vertical con un cierto fluido.

La densidad equivalente de lodo (EMW) es igual a la sumatoria de todas las presiones que afecta a un pozo, las cuales no se nombran acá porque se entraría a otros temas, a una profundad o zona dada.

$$EMW = \frac{Presión\ (psi)}{0{,}052 * Profundidad\ vertical\ (ft)} + Densidad\ de\ fluido\ actual\ (ppg) \quad (64)$$

### 1.2.7. Surgencias

Una surgencia es una entrada no deseada de fluidos de yacimiento hacia el pozo, generando un grupo de sobrecostos operativos o incluso generando riesgos a la integridad de las personas si esta surgencia se convierte en un reventón.

Si el fluido de yacimiento observa una presión menor dentro el pozo, tenderá a fluir a este, y si esto no se realiza de manera controlada, se ocasionará un reventón, debido a esto la ecuación (64) es importante porque al momento de detectar una surgencia, se suele realizar un cierre de pozo para medir la presión en una prueba, y con esto calcular la presión nueva de fluido necesaria para contrarrestar la presión del yacimiento.

### 1.2.8. Fluidos De Control

Para la mayoría de los métodos de control de pozo usados para controlar las surgencias se requiere el uso de un fluido de control, dentro el pozo, con una densidad equivalente para realizar una igualación de la presión de yacimiento con la presión hidrostática, generando un equilibrio. Como la profundidad del pozo permanece constante, la única forma de alcanzar la igualación de presión con el fluido de control es el aumento de la densidad de este.

Para producir este fluido de control se inicia con un fluido base, agua, y se le agregan sales que se disuelven en este aumentando su densidad final, generando una salmuera. Las sales más comúnmente usadas son el cloruro de sodio, clorato de sodio, formiato de potasio, cloruro de potasio y formiato de sodio, siendo estos dos últimos los de interés para este trabajo.

### *1.2.9.* Características Requeridas:

Un buen fluido debería:

- Ser lo suficientemente denso como para controlar la presión del yacimiento, sin ser tan denso que fluya hacia el yacimiento generando perdida de fluido y pudiendo generar daños en la formación petrolífera.

- Ser efectivo en su relación costo/beneficio, siendo menos costoso y causando poco o ningún daño en la formación.
- Estar libre de partículas solidas que puedan taponar los poros de la formación, y filtrado.
- Ser no corrosivo, para evitar daños en las tuberías que sostienen al pozo.
- Ser estable en largos tiempos, sin decantarse, perdiendo densidad, si se planea dejarlo dentro del pozo por mucho tiempo.

#### 1.2.10. Barril Equivalente

Para poder realizar experimentos a nivel de laboratorio, ahorrando costos debido a la cantidad de reactivos necesarios, se hace uso de un barril equivalente, definido en glosario de la empresa Schlumberger de la siguiente manera:

En la interpretación se sugiere que "Un volumen de 350 cm3. En los experimentos de laboratorio con lodos, 350 cm3 es el volumen elegido para representar 42 galones estadounidenses (1 barril de petróleo) [0,159 m3], de modo que una masa de 1,0 gramo representa 1,0 lbm. Este es un concepto útil para los técnicos de lodos cuando mezclan o hacen ensayos de pruebas con muestras de lodos. Por ejemplo, en la preparación de una formulación o para los fines de una prueba piloto, añadir 1,0 gramo a 350 cm3 de fluido es el equivalente experimental de añadir 1,0 lb a 42 galones (1,0 bbl) de fluido. (Schlumberger, 2021)

#### 1.2.11. Medición De La Densidad

Para medir la densidad se usaron dos instrumentos, la balanza de lodos y el picnómetro para confirmar. Aunque el picnómetro tiene mayor exactitud, es cierto que en campo normalmente no se tiene acceso a él, además de que la exactitud medida del lodo siempre se hace con máximo un decimal, por tanto, la exactitud del picnómetro no es necesaria.

#### *1.2.11.1. Medición Balanza De Lodos*

La balanza cosiste en un brazo graduado y un contrapeso para medir la densidad. Los procedimientos para medir la densidad con este instrumento es el siguiente:

1. Ubicar la base del elemento de tal modo que quede nivelado.
2. Llenar el recipiente con la salmuera asegurando que el recipiente esté limpio y seco.
3. Colocar la tapa del recipiente y asegurando secar la salmuera del recipiente que salga por el hoyo de la tapa.
4. Instale la cuchilla del brazo en el apoyo.
5. Mover el contrapeso del brazo graduado hasta que el recipiente y el brazo estén en balance.
6. Se lee la densidad en la que quedó el contrapeso por medio de la regleta del brazo (ppg).

Nota: La balanza debe ser calibrada con agua fresca o destilada, siguiendo los pasos anteriores. Si se obtiene 8.33 ppg, si no, se debe ajustar la balanza.

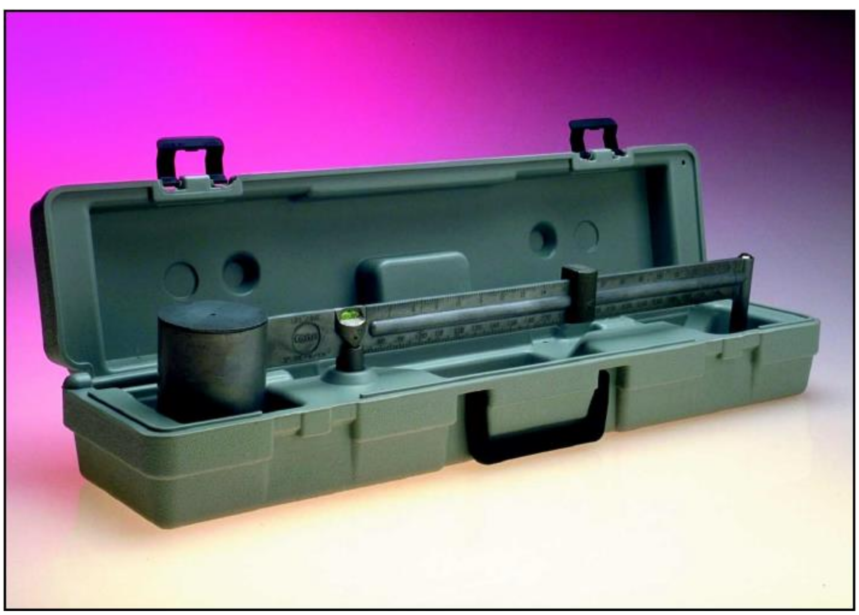

**Figura *9*. "Institution Well Control International. (2017) Manual de control de pozos WCS. Balanza de lodos Cap. 9"**

#### *1.2.11.2. Medición En Picnómetro*

El picnómetro es un método sencillo por medio del cual sabiendo la densidad del agua fresca se puede calcular la densidad de otro fluido para un volumen fijo que es el picnómetro.

$$v_{agua} = v_{salmuera} \quad (65)$$

$$\frac{m_{agua}}{\rho_{agua}} = \frac{m_{salmuera}}{\rho_{salmuera}} \quad (66)$$

$$\rho_{salmuera} = \frac{m_{salmuera} * \rho_{agua}}{m_{agua}} \quad (67)$$

Teniendo en cuenta esta formula configuramos la masa del picnómetro en cero para la balanza digital, se mide la masa del picnómetro lleno de agua fresca, se hace lo mismo para la salmuera, luego se reemplaza en la formula (67) teniendo en cuenta que la densidad del agua fresca es conocida. Realizando las conversiones necesarias la ecuación (67) para el picnómetro que se tiene en el laboratorio del proyecto se obtiene:

$$\rho_{salmuera}(ppg) = \frac{m_{salmuera}(gramos) * (8{,}33)}{25{,}064} \quad (68)$$

# 2. Descripción Y Desarrollo Del Problema

## 2.1. Planteamiento Del Problema

Actualmente no existen unas tablas suministradas por la academia para la preparación de salmueras tipo KCl (Cloruro de potasio), NaCOOH (Formiato de sodio) y mezclas entre ellas; algunas empresas han creado sus propias tablas, pero hay una variación muy grande entre las concentraciones usadas para alcanzar dichas densidades cuando se incluye el formiato de sodio, por tanto, se requiere por medio de pruebas de laboratorio y de campo generar unas tablas y/o correlaciones que al ponerse a prueba cumplan con repetibilidad y reproducibilidad.

Las densidades finales alcanzadas por cada sal varían dependiendo de su solubilidad en el agua, teniendo un estándar a continuación:

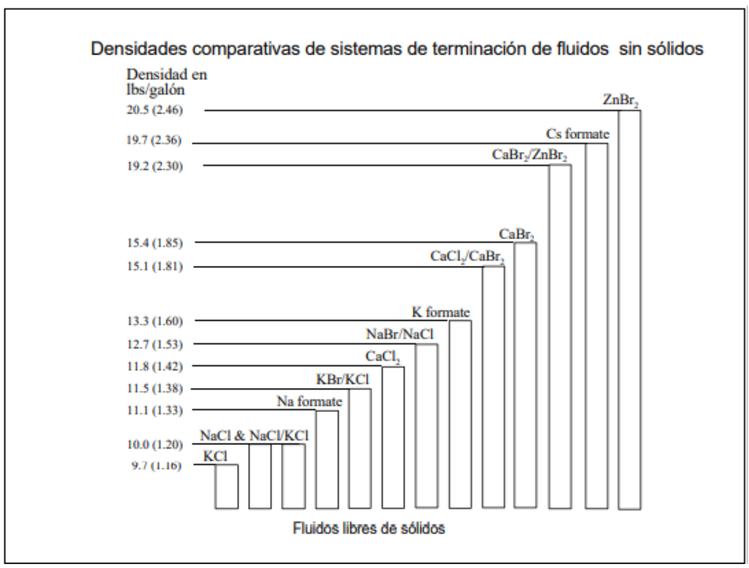


**Figura 10. "Baroid. (2000) Manual de fluidos. Densidades máximas alcanzadas para diferentes sales Cap. 1 Pág. 1-5 fig. 11"**

La mezcla de KCl y NaCOOH pueden alcanzar un peso máximo igual al alcanzado por el NaCOOH solo (11,1 ppg).

La importancia de este proyecto radica en la estandarización de una metodología que permita la preparación de salmueras monovalentes (formiato de sodio) y bivalente (cloruro de potasio y formiato de sodio), debido a que este es un proceso operativo en el que muchas veces durante su preparación o tratamiento encontramos una incertidumbre de las concentraciones requeridas cuando se realiza una mezcla de estas, lo cual no permite tener una planeación y logística adecuada cuando se requieren estos productos para las actividades de control del pozo especialmente. Con este trabajo se busca solucionar este inconveniente y así facilitar a las diversas empresas de servicios de fluidos una estandarización en las medidas o cantidades de los productos a utilizar. La mezcla de sales en la preparación de salmueras se realiza por diversos motivos, entre ellos el peso máximo posible alcanzado por la salmuera, los costos de la materia primera y los tiempos de preparación. De aquí figuran varios aspectos que mejoran la aplicación de la técnica como disminución de costos, disponibilidad de productos químicos, especialmente formiato de sodio, y mejoramiento en la salud y seguridad de los trabajadores involucrados en la preparación de la salmuera debido a la necesidad de uso excesivo de fuerza durante tiempos prolongados cuando se requieren concentraciones considerables de formiato de sodio debido a su difícil manejo en comparación con otros productos. Por otro lado, una utilidad destacada es la búsqueda de nuevas concentraciones de los productos puros en las mezclas que permitan alcanzar mayores densidades con una menor cantidad de producto.

Nota: Todos los productos usados en la empresa donde se realizaron los experimentos de este proyecto deben pasar una prueba inicial para unos límites de pH, solidos suspendidos, dureza total, turbidez, cloruros, hierro total, hierro ferroso, sulfatos y nitratos exigidos por las empresas operadoras que solicitan las salmueras. Debido a esto no fue necesario realizar pruebas de esto de nuevo para todos los experimentos ya que los químicos usados se aseguran de que estén libre de partículas solidas que puedan taponar los poros de la formación.

## 2.2. Antecedentes

Algunas empresas han creado sus propias tablas, las cuales son confidenciales para uso de estas mismas o para investigaciones privadas. Sin embargo, existen algunas investigaciones que muestran algún tipo de relación con el trabajo a desarrollar. Baroid en el año 2000 en su libro “manual de fluidos” estableció algunas tablas para productos como Cloruro de sodio, Cloruro de potasio, Cloruro de calcio, Bromuro de sodio, Bromuro de sodio/cloruro de sodio, Bromuro de calcio, Bromuro de calcio/cloruro de calcio. Sin embargo, no hacen énfasis en las salmueras de formiato que son de bastante aplicabilidad hoy en día ni de las mezclas que incluyan cloruro de potasio y formiatos.

Otra empresa que ha dado aportes significativos en el tema, incluso sus tablas son utilizadas por otras empresas para la preparación de las salmueras es Brinadd, que incluye las tablas de la preparación de salmueras de formiato de sodio, cloruro de potasio y mezclas entre estas (se pueden apreciar en sección de generalidades). Sin embargo, al momento de utilizar estas tablas se tiene una variación considerable en la práctica, por ende, lo que se busca es mejorar la exactitud de las tablas.

Es de resaltar que no se encuentra más información relevante frente a la preparación de salmueras, la mayoría de los estudios y literatura que se encuentra hace referencia a la aplicabilidad de estas salmueras en casos específicos más no en su preparación.

Es de importancia aclarar que la selección de la salmuera adecuada depende de muchos factores como temperatura, presión, compatibilidad con la formación, compatibilidad con los fluidos de formación, pH, viscosidad, entre otros. Este trabajo no se centrará en la selección de la salmuera en relación con estos parámetros si no en la preparación de ésta de acuerdo con las concentraciones de cada sal para lograr un peso requerido.

A continuación, se adjuntan las tablas que suministra Brinadd las cuales sirvieron como base:

| Densidad lbs/gal | Porcentaje peso | Concent lbs/bbl | Volumen agua bbl |
|---|---|---|---|
| 8.4 | 1.50 | 4.00 | 0.995 |
| 8.5 | 3.50 | 11.60 | 0.986 |
| 8.6 | 5.50 | 18.90 | 0.976 |
| 8.7 | 7.40 | 26.10 | 0.969 |
| 8.8 | 9.30 | 33.40 | 0.960 |
| 8.9 | 11.10 | 40.70 | 0.950 |
| 9.0 | 12.90 | 47.90 | 0.943 |
| 9.1 | 14.70 | 55.20 | 0.933 |
| 9.2 | 16.40 | 62.40 | 0.924 |
| 9.3 | 18.00 | 69.70 | 0.917 |
| 9.4 | 19.70 | 76.90 | 0.907 |
| 9.5 | 21.20 | 84.20 | 0.898 |
| 9.6 | 22.70 | 91.50 | 0.890 |
| 9.7 | 24.20 | 98.70 | 0.881 |

**Tabla 1 "Baroid. (2000) Manual de fluidos. Solución de cloruro de potasio requerida para hacer 1barril (42 gal). Cap. 1 Pág. 1-13 Tabla. 1-3"**

| Densidad lbs/gal | Porcentaje peso | Concent lbs/bbl | Volumen agua bbl |
|---|---|---|---|
| 8.4 | 1.25 | 4.4 | 0.99 |
| 8.5 | 3.01 | 10.7 | 0.98 |
| 8.6 | 4.75 | 17.2 | 0.97 |
| 8.7 | 6.48 | 23.7 | 0.96 |
| 8.8 | 8.18 | 30.2 | 0.95 |
| 8.9 | 9.87 | 36.9 | 0.94 |
| 9.0 | 11.54 | 43.6 | 0.93 |
| 9.1 | 13.20 | 50.5 | 0.92 |
| 9.2 | 14.84 | 57.4 | 0.90 |
| 9.3 | 16.47 | 64.4 | 0.89 |
| 9.4 | 18.09 | 71.4 | 0.88 |
| 9.5 | 19.70 | 78.6 | 0.87 |
| 9.6 | 21.29 | 85.9 | 0.85 |
| 9.7 | 22.17 | 89.9 | 0.84 |
| 9.8 | 24.45 | 100.7 | 0.83 |
| 9.9 | 26.02 | 108.2 | 0.81 |
| 10.0 | 27.58 | 115.8 | 0.80 |
| 10.1 | 29.13 | 123.6 | 0.79 |
| 10.2 | 30.68 | 131.4 | 0.78 |
| 10.3 | 32.22 | 139.4 | 0.77 |
| 10.4 | 33.75 | 147.4 | 0.76 |
| 10.5 | 35.28 | 155.6 | 0.76 |
| 10.6 | 36.81 | 163.9 | 0.75 |
| 10.7 | 38.33 | 172.3 | 0.75 |
| 10.8 | 39.85 | 180.8 | 0.74 |
| 10.9 | 41.37 | 189.4 | 0.74 |
| 11.0 | 42.88 | 198.1 | 0.73 |
| 11.1 | 45.00 | 210.6 | 0.73 |

**Tabla 2 "Baroid. (2000) Manual de fluidos. Solución de formiato de sodio requerida para hacer 1barril (42 gal). Cap. 1 Pág. 1-14 Tabla. 1-4"**

| K C L | | | |
|---|---|---|---|
| Densidad lbs/gal | Porcentaje peso | Concent lbs/bbl | Volumen agua bbl |
| 8.4 | 1.3 | 4.00 | 0.995 |
| 8.5 | 3.1 | 11.60 | 0.986 |
| 8.6 | 5.0 | 18.90 | 0.976 |
| 8.7 | 6.8 | 26.10 | 0.969 |
| 8.8 | 8.6 | 33.40 | 0.960 |
| 8.9 | 10.3 | 40.70 | 0.950 |
| 9.0 | 12.1 | 47.90 | 0.943 |

| FORMIATO DE SODIO + KCL | | | |
|---|---|---|---|
| Densidad lbs/gal | Porcentaje peso, Fna | Concent Fna lbs/bbl | KCl 9,0 PPG bbl |
| 9.10 | 1.7 | 6.55 | 0.99 |
| 9.20 | 3.4 | 13.19 | 0.99 |
| 9.30 | 5.1 | 19.90 | 0.98 |
| 9.40 | 6.8 | 26.70 | 0.97 |
| 9.50 | 8.4 | 33.59 | 0.97 |
| 9.60 | 10.1 | 40.56 | 0.96 |
| 9.70 | 11.7 | 47.63 | 0.95 |
| 9.80 | 13.3 | 54.78 | 0.94 |
| 9.90 | 14.9 | 62.03 | 0.94 |
| 10.00 | 16.5 | 69.37 | 0.93 |
| 10.10 | 18.1 | 76.80 | 0.92 |
| 10.20 | 19.7 | 84.34 | 0.91 |
| 10.30 | 21.3 | 91.97 | 0.91 |
| 10.40 | 22.8 | 99.71 | 0.90 |
| 10.50 | 24.4 | 107.55 | 0.89 |
| 10.60 | 25.9 | 115.50 | 0.88 |
| 10.70 | 27.5 | 123.56 | 0.87 |
| 10.80 | 29.0 | 131.72 | 0.87 |
| 10.90 | 30.6 | 140.00 | 0.86 |
| 11.00 | 32.1 | 148.40 | 0.85 |
| 11.10 | 33.7 | 156.91 | 0.84 |
| 11.20 | 35.2 | 165.55 | 0.83 |
| 11.30 | 36.7 | 174.31 | 0.82 |
| 11.40 | 38.3 | 183.19 | 0.81 |
| 11.45 | 39.0 | 187.68 | 0.81 |

**Tabla 3 "Baroid. (2000) Manual de fluidos. Solución combinada de cloruro de potasio (9ppg) y formiato de sodio requeridos para hacer 1barril (42 gal). Cap. 1 Pág. 1-15 Tabla. 1-5"**

## 2.3. Diseño estadístico

Para abordar el trabajo se inició realizando una prueba de ANOVA de dos vías, diseño por bloques completamente balanceados aleatorizados, para comprobar si hay diferencias significativas en la densidad alcanzada por la salmuera por tres proveedores distintos, iniciando en todos con un volumen X de salmuera de KCL al 9 ppg y terminando siempre en 350 ml de salmuera, igual a un barril equivalente.

Luego se establece una regresión para la preparación de un barril equivalente de salmuera monovalente de NaCOOH a diferentes densidades, iniciando con un volumen de agua, volumen descrito también por una regresión.

A continuación, se genera una regresión para la preparación de un barril equivalente de una salmuera bivalente de KCl y NaCOOH a diferentes densidades, iniciando con un volumen de salmuera monovalente de KCl al 9 pp, volumen de salmuera descrito también por una regresión. Se debe aclarar que la tabla de preparación de salmuera monovalente KCl suministrada por Brinadd se ajusta bastante bien a lo observado en laboratorio y campo, por eso no se explorará más esto, sino que se trabajará con la tabla suministrada por ellos.

Nota: No se describirá todo el proceso realizado para obtener los datos aquí usados ya que fueron bastantes los experimentos que no obtuvieron el resultado deseado. Todas las preparaciones finales de salmuera deben dar como resultado un barril equivalente ya que se estableció esto con estándar para poder comparar las respuestas. Se sabe que la relación se conserva, y que, si se desea preparar más de un barril equivalente, por ejemplo, para dos barriles equivalentes, debo usar el doble de volumen inicial y el doble de sal, en otras palabras, la cantidad es multiplicativa.

# 3. Resultados Obtenidos

## 3.1. ¿Existen diferencias dependiendo del lote o del proveedor?

Las sales requeridas para la preparación de salmueras puedes ser suministrada por diferentes proveedores que obtienen esta materia prima de diferentes lugares del globo, por tanto, se debe verificar si hay diferencias significativas entre proveedores.

Para este experimento se llega siempre a un barril equivalente de salmuera iniciando con un cierto volumen de salmuera al 9 ppg que asegure que al final se obtenga un barril equivalente. Se analizaron 3 proveedores (tratamiento) y 5 adiciones de NaCOOH para cada uno (bloqueo), estos bloques son iguales para cada tratamiento y el volumen inicial de salmuera al 9 ppg es igual en cada bloque para los mismos tratamientos. Para cada cruce tratamiento y bloque se midió la densidad final alcanzada por la salmuera en el picnómetro.

Para iniciar el experimento se realizó una aleatorización del proveedor y bloque usado en R.

| | plots | r | tratamiento |
|---|---|---|---|
| 1 | 101 | 1 | Proveedor3 |
| 2 | 102 | 1 | Proveedor2 |
| 3 | 103 | 2 | Proveedor2 |
| 4 | 104 | 1 | Proveedor1 |
| 5 | 105 | 2 | Proveedor1 |
| 6 | 106 | 3 | Proveedor2 |
| 7 | 107 | 3 | Proveedor1 |
| 8 | 108 | 4 | Proveedor1 |
| 9 | 109 | 2 | Proveedor3 |
| 10 | 110 | 3 | Proveedor3 |
| 11 | 111 | 4 | Proveedor3 |
| 12 | 112 | 4 | Proveedor2 |
| 13 | 113 | 5 | Proveedor3 |
| 14 | 114 | 5 | Proveedor2 |
| 15 | 115 | 5 | Proveedor1 |

**Tabla 4. Aleatorización.**

Los bloques son, en orden, 189.4, 162, 116 y 70.4 gramos de NaCOOH adicionados a la salmuera de 9 ppg.

Después de realizar el experimento y organizar los resultados se obtiene:

| Proveedor | NaCOOH | Densidad |
|---|---|---|
| Proveedor 1 | 189,4 | 11,1 |
| Proveedor 1 | 162 | 10,82 |
| Proveedor 1 | 116 | 10,31 |
| Proveedor 1 | 70,4 | 9,83 |
| Proveedor 1 | 24 | 9,3 |
| Proveedor 2 | 189,4 | 11,1 |
| Proveedor 2 | 162 | 10,8 |
| Proveedor 2 | 116 | 10,29 |
| Proveedor 2 | 70,4 | 9,81 |
| Proveedor 2 | 24 | 9,29 |
| Proveedor 3 | 189,4 | 11,08 |
| Proveedor 3 | 162 | 10,79 |
| Proveedor 3 | 116 | 10,27 |
| Proveedor 3 | 70,4 | 9,77 |
| Proveedor 3 | 24 | 9,34 |

**Tabla 5. Resumen salmueras bivalentes de cloruro de potasio (9 ppg) y formiato de sodio requeridas para hacer 1 barril equivalente a diferentes densidades.**

Al ingresar esto a R para realizar en análisis de varianza se obtiene:

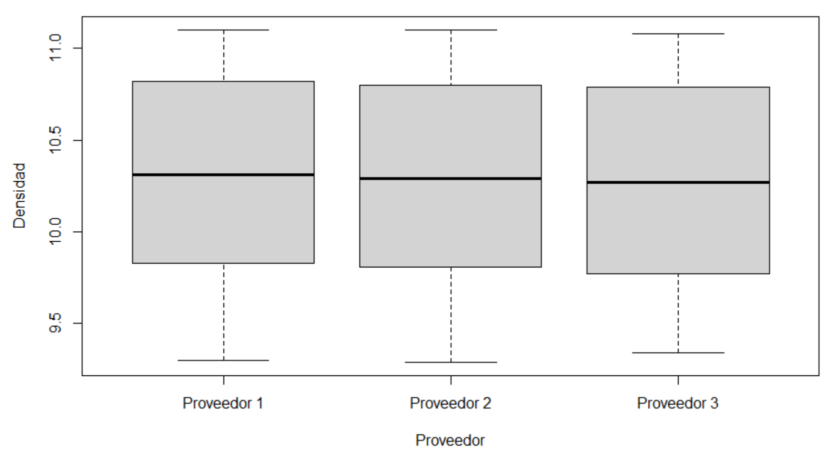


**Figura 11. Boxplot densidad VS proveedor.**

No se observan diferencias significativas gráficamente.

```
Analysis of Variance Table

Response: Densidad
                   Df Sum Sq Mean Sq   F value    Pr(>F)    
factor(Proveedor)   2 0.0012 0.00062    1.3933    0.3026    
factor(NaCOOH)      4 6.3048 1.57620 3542.0225 5.069e-13 ***
Residuals           8 0.0036 0.00045                        
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Figura 12. ANOVA densidad vs proveedor.**

Se confirma por medio del ANOVA que el proveedor no afecta significativamente la densidad final obtenida ya que su Valor P es 0.2093, por tanto, es mayor a 0.05.

```
> t.test(m$residuals)

        One Sample t-test

data:  m$residuals
t = -5.6176e-17, df = 14, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.008830788  0.008830788
sample estimates:
    mean of x 
-2.312965e-19 

> shapiro.test(m$residuals)

        Shapiro-Wilk normality test

data:  m$residuals
W = 0.93319, p-value = 0.3043

> lmtest::bptest(m)

        studentized Breusch-Pagan test

data:  m
BP = 10.491, df = 6, p-value = 0.1054

> dwt(m) #Solo revisa la autocorrelación de orden 1
 lag Autocorrelation D-W Statistic p-value
   1      -0.3179775      2.178527    0.96
 Alternative hypothesis: rho != 0
```

**Figura 13. Supuestos del error ANOVA densidad vs proveedor.**

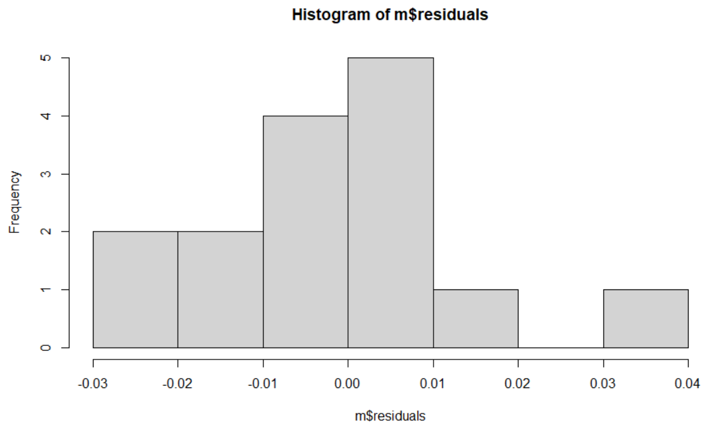


**Figura 14. Histograma residual.**

Se cumplen los supuestos del error de media cero, distribución normal, varianza constante y no autocorrelación, por tanto, el modelo ANOVA es confiable y la conclusión hecha es correcta.

Ya habiendo comprobado que no hay diferencias entre proveedores, se usarán estos mismos proveedores para realizar el cálculo de las siguientes regresiones.

## 3.2. Salmuera Monovalente De Formiato De Sodio

Para una salmuera monovalente de formiato de sodio se requiere conocer la cantidad de NaCOOH y volumen inicial de agua necesaria para obtener como resultado un barril equivalente, a cierta densidad final, después del aumento de volumen generado por la sal.

| Salmuera final 350 ml | | |
|---|---|---|
| Densidad (ppg) | Adición NaCOOH (gr) | Volumen de Agua inicial (ml) |
| 9,14 | 64 | 325 |
| 11,07 | 244 | 248 |
| 9,14 | 64 | 325 |
| 9,95 | 145 | 290 |
| 11,1 | 244 | 247 |

**Tabla 6. Resumen barril equivalente salmuera monovalente.**

Se realizarán dos regresiones, una que relacione densidad final de la salmuera con la adición de NaCOOH necesaria, y otra con densidad final de la salmuera y el volumen de agua inicial.

```
Call:
lm(formula = NaCOOH ~ Densidad)

Residuals:
      1       2       3       4       5 
-1.4120  0.3956 -1.4120  4.8026 -2.3743 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -778.463     17.163  -45.36 2.36e-05 ***
Densidad      92.328      1.696   54.43 1.37e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.309 on 3 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.9987 
F-statistic:  2962 on 1 and 3 DF,  p-value: 1.366e-05
```

**Figura 15. Regresión lineal salmuera monovalente Densidad Vs cant. NaCOOH.**

Se observa que tanto el intercepto como la pendiente del modelo son significativas, además de que el coeficiente de determinación es casi perfecto.

```
> t.test(resid)

        One Sample t-test

data:  resid
t = 3.4653e-17, df = 4, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -3.558067  3.558067
sample estimates:
   mean of x
4.440892e-17

> shapiro.test(resid)

        Shapiro-Wilk normality test

data:  resid
W = 0.82523, p-value = 0.1281

> lmtest::bptest(MODNaCOOH)

        studentized Breusch-Pagan test

data:  MODNaCOOH
BP = 0.0026153, df = 1, p-value = 0.9592

> hist(MODNaCOOH$residuals)
> qqPlot(MODNaCOOH$residuals)
[1] 4 5
> dwtest(MODNaCOOH,alternative="two.sided")

        Durbin-Watson test

data:  MODNaCOOH
DW = 2.9429, p-value = 0.1717
alternative hypothesis: true autocorrelation is not 0
```

**Figura 16. Supuestos del error regresión salmuera monovalente Densidad Vs cant. NaCOOH.**

Se cumplen los supuestos del error de media cero, distribución normal, varianza constante y no autocorrelación, por tanto, el modelo de regresión lineal de la Densidad final de la salmuera (ppg) Vs la cantidad de NaCOOH requerida (gr).

```
> summary(MODNaCOOH2)

Call:
lm(formula = VolAgua ~ Densidad)

Residuals:
      1       2       3       4       5 
 0.6332  0.3544  0.6332 -2.1677  0.5469 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 687.6989     7.2902   94.33 2.63e-06 ***
Densidad    -39.7519     0.7205  -55.17 1.31e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.405 on 3 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.9987 
F-statistic:  3044 on 1 and 3 DF,  p-value: 1.312e-05
```

**Figura 17. Regresión lineal salmuera Densidad Vs vol. de agua inicial.**

Se observa que tanto el intercepto como la pendiente del modelo son significativas, además de que el coeficiente de determinación es casi perfecto.

```
> t.test(resid)

        One Sample t-test

data:  resid
t = 4.0793e-17, df = 4, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.511286  1.511286
sample estimates:
   mean of x
2.220446e-17

> shapiro.test(resid)

        Shapiro-Wilk normality test

data:  resid
W = 0.62791, p-value = 0.001436

> lmtest::bptest(MODNaCOOH)

        studentized Breusch-Pagan test

data:  MODNaCOOH
BP = 0.0026153, df = 1, p-value = 0.9592

> hist(MODNaCOOH$residuals)
> qqPlot(MODNaCOOH$residuals)
[1] 4 5
> dwtest(MODNaCOOH,alternative="two.sided")

        Durbin-Watson test

data:  MODNaCOOH
DW = 2.9429, p-value = 0.1717
alternative hypothesis: true autocorrelation is not 0
```

**Figura 18. Supuestos del error regresión salmuera monovalente Densidad Vs vol. De agua inicial.**

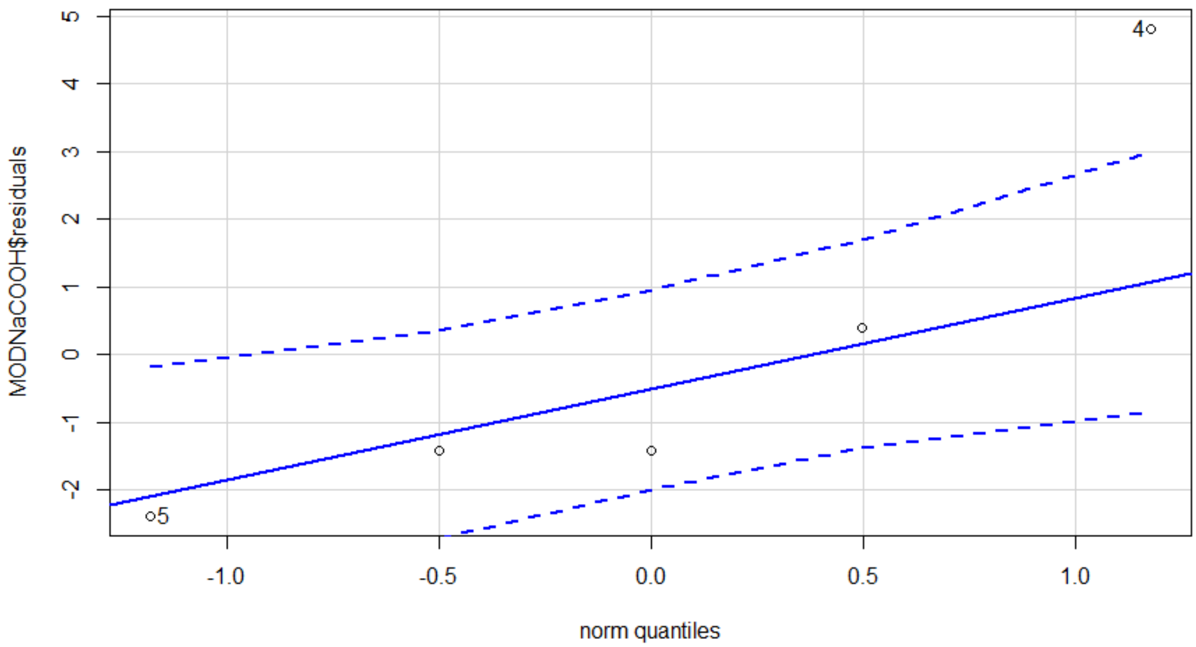


**Figura 19. Gráfico cuantiles residuales salmuera monovalente Densidad Vs vol. de agua inicial.**

Se cumplen los supuestos del error de media cero, varianza constante y no autocorrelación y aunque no pasa la prueba de normalidad de Sahpiro Wilk, al observar el grafico de cuantiles muestra que solo hay problemas con un dato. Al analizar este dato individualmente se llega a la conclusión de que corresponde al proveedor que peor comportamiento tiene, y que a este proveedor para que pueda realizar la disolución completa se requieren de trabajos adicionales que a nivel industrial se hace, pero que en prueba de laboratorio no se puede. Esto genera que existan problemas con este dato pero que pueden ser resueltos sin problemas cuando deba ser usado en campo, por tanto, el modelo de regresión lineal de la Volumen de agua inicial (ml) Vs la cantidad de NaCOOH requerida (gr) es útil.

### 3.3. Salmuera Bivalente De Cloruro De Potasio Y Formiato De Sodio

Para una salmuera bivalente de cloruro de potasio y formiato de sodio se requiere conocer la cantidad de NaCOOH y volumen inicial de salmuera KCl al 9 ppg necesaria para obtener como resultado un barril equivalente, a cierta densidad final, después del aumento de volumen generado por la sal.

| Salmuera final 350 ml | | |
|---|---|---|
| Densidad | NaCOOH | VolKCl |
| 189,4 | 11,1 | 260 |
| 162 | 10,82 | 275 |
| 116 | 10,31 | 297 |
| 70,4 | 9,83 | 320 |
| 24 | 9,3 | 343 |
| 189,4 | 11,1 | 260 |
| 162 | 10,8 | 275 |
| 116 | 10,29 | 297 |
| 70,4 | 9,81 | 320 |
| 24 | 9,29 | 343 |
| 189,4 | 11,08 | 260 |
| 162 | 10,79 | 275 |
| 116 | 10,27 | 297 |
| 70,4 | 9,77 | 320 |
| 24 | 9,34 | 343 |

**Tabla 7. Resumen barril equivalente salmuera bivalente.**

Se realizarán dos regresiones, una que relacione densidad final de la salmuera con la adición de NaCOOH necesaria, y otra con densidad final de la salmuera y el volumen de KCl al 9 ppg.

```
> summary(MODNaCOOH)

Call:
lm(formula = NaCOOH ~ Densidad)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3309 -0.7882 -0.2684  1.0602  3.3273 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -835.8997     7.6461  -109.3   <2e-16 ***
Densidad      92.4230     0.7438   124.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.868 on 13 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.9992,	Adjusted R-squared:  0.9991 
F-statistic: 1.544e+04 on 1 and 13 DF,  p-value: < 2.2e-16
```

**Figura 20. Regresión lineal salmuera bivalente Densidad Vs cant. NaCOO.**

Se observa que tanto el intercepto como la pendiente del modelo son significativas, además de que el coeficiente de determinación es casi perfecto.

```
> t.test(resid)

        One Sample t-test

data:  resid
t = 1.5985e-17, df = 14, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.9969558  0.9969558
sample estimates:
   mean of x
7.430399e-18

> shapiro.test(resid)

        Shapiro-Wilk normality test

data:  resid
W = 0.98048, p-value = 0.9726

> lmtest::bptest(MODNaCOOH)

        studentized Breusch-Pagan test

data:  MODNaCOOH
BP = 2.4131, df = 1, p-value = 0.1203

> hist(MODNaCOOH$residuals)
> qqPlot(MODNaCOOH$residuals)
[1] 15 14
> dwtest(MODNaCOOH,alternative="two.sided")

        Durbin-Watson test

data:  MODNaCOOH
DW = 1.4926, p-value = 0.2745
alternative hypothesis: true autocorrelation is not 0
```

**Figura 21. Supuestos del error regresión salmuera bivalente Densidad Vs cant. NaCOOH.**

Se cumplen los supuestos del error de media cero, distribución normal, varianza constante y no autocorrelación, por tanto, el modelo de regresión lineal de la Densidad final de la salmuera (ppg) Vs la cantidad de NaCOOH requerida (gr).

```
> summary(MODNaCOOH2)

Call:
lm(formula = VolKCl ~ Densidad)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5966 -0.6743 -0.2630  0.6718  1.8247 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 772.1444     4.5105   171.2   <2e-16 ***
Densidad    -46.1154     0.4387  -105.1   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.102 on 13 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9987 
F-statistic: 1.105e+04 on 1 and 13 DF,  p-value: < 2.2e-16
```

**Figura 22. Regresión lineal salmuera bivalente Densidad Vs vol. de salmuera KCl 9 ppg.**

Se observa que tanto el intercepto como la pendiente del modelo son significativas, además de que el coeficiente de determinación es casi perfecto.

```
> t.test(resid)

        One Sample t-test

data:  resid
t = -1.4187e-16, df = 14, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.5881103  0.5881103
sample estimates:
    mean of x
-3.890117e-17

> shapiro.test(resid)

        Shapiro-Wilk normality test

data:  resid
W = 0.96416, p-value = 0.7641

> lmtest::bptest(MODNaCOOH)

        studentized Breusch-Pagan test

data:  MODNaCOOH
BP = 2.4131, df = 1, p-value = 0.1203

> hist(MODNaCOOH$residuals)
> qqPlot(MODNaCOOH$residuals)
[1] 15 14
> dwtest(MODNaCOOH,alternative="two.sided")

        Durbin-Watson test

data:  MODNaCOOH
DW = 1.4926, p-value = 0.2745
alternative hypothesis: true autocorrelation is not 0
```

**Figura 23. Supuestos del error regresión salmuera bivalente Densidad Vs vol. de salmuera KCl 9 ppg.**

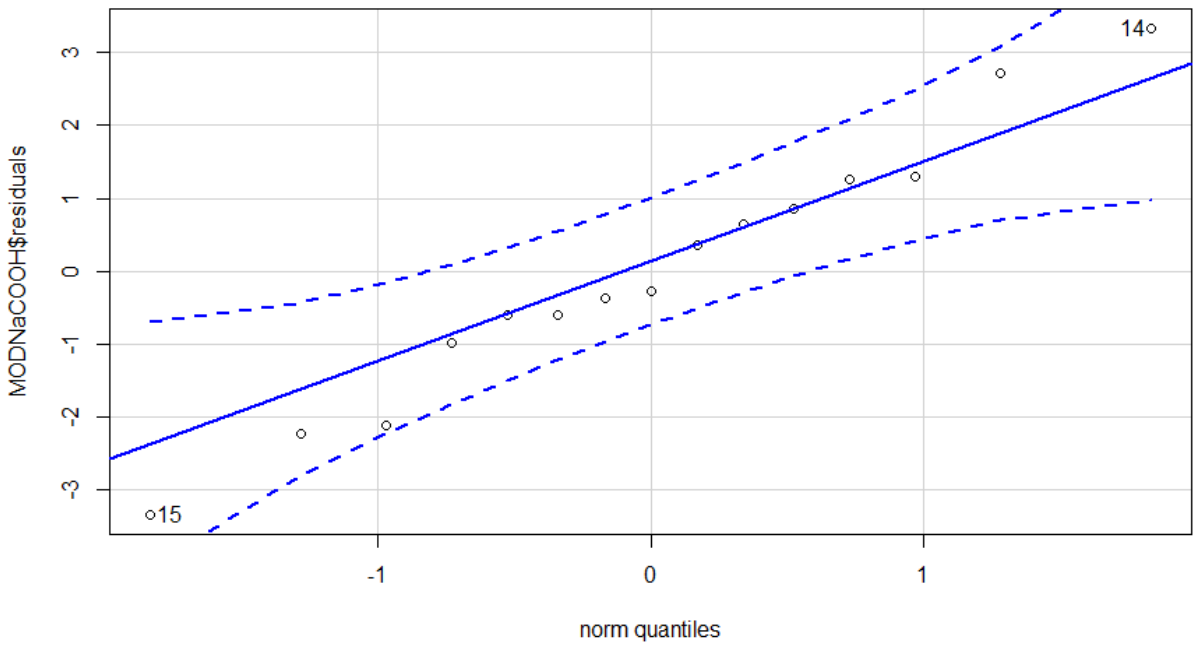


**Figura 24. Gráfico cuantiles residuales salmuera bivalente Densidad Vs vol. de agua inicial.**

Se cumplen los supuestos del error de media cero, distribución normal, varianza constante y no autocorrelación, por tanto, el modelo de regresión lineal de la Volumen de salmuera KCl al 9 ppg a inicial (ml) Vs la cantidad de NaCOOH requerida (gr) es útil. Aunque el modelo no pase prueba de normalidad Shapiro Wilk, en la gráfica de cuantiles no se observan problemas serios de normalidad, y curiosamente se observa problema con el mismo proveedor con el que se tuvo problemas para la salmuera monovalente.

### 3.4. Preparación En Campo

Los resultados acá obtenidos para la salmuera bivalente KCl + NaCOOH fueron puestos a prueba en tres preparaciones. Se prepararon 103 bbl al 11.1 ppg, 200 bbl a 11 ppg y 120 bbl a 10.7 ppg obteniendo una diferencia de 1 a 5 sacos de 25 kg de NaCOOH máximo, sin superar un 2% de error, una diferencia bastante aceptable considerando las cantidades que fueron preparadas, y en el volumen no se encontraron diferencias medibles. No fue posible poner a prueba los resultados obtenidos para la salmuera monovalente de NaCOOH pero se espera no difiera mucho por la exactitud de la bivalente.

# 4. Metodología

## 4.1. Preparación Salmuera Monovalente NaCOOH

Como resultado de la regresión generada en los numerales anteriores se genera la siguiente tabla realizando la conversión de ml a barriles equivalentes (350 ml):

| NaCOOH | | |
|---|---|---|
| **Densidad lbs/gal** | **Concentración lbs/bbl** | **Volumen de agua bbl** |
| 8,4 | 2,00 | 0,99 |
| 8,5 | 6,33 | 0,99 |
| 8,6 | 15,56 | 0,99 |
| 8,7 | 24,79 | 0,98 |
| 8,8 | 34,02 | 0,97 |
| 8,9 | 43,26 | 0,95 |
| 9 | 52,49 | 0,94 |
| 9,1 | 61,72 | 0,93 |
| 9,2 | 70,95 | 0,92 |
| 9,3 | 80,19 | 0,91 |
| 9,4 | 89,42 | 0,90 |
| 9,5 | 98,65 | 0,89 |
| 9,6 | 107,89 | 0,87 |
| 9,7 | 117,12 | 0,86 |
| 9,8 | 126,35 | 0,85 |
| 9,9 | 135,58 | 0,84 |
| 10 | 144,82 | 0,83 |
| 10,1 | 154,05 | 0,82 |
| 10,2 | 163,28 | 0,81 |
| 10,3 | 172,52 | 0,80 |
| 10,4 | 181,75 | 0,78 |
| 10,5 | 190,98 | 0,77 |
| 10,6 | 200,21 | 0,76 |
| 10,7 | 209,45 | 0,75 |
| 10,8 | 218,68 | 0,74 |
| 10,9 | 227,91 | 0,73 |
| 11 | 237,15 | 0,72 |
| 11,1 | 246,38 | 0,70 |

**Tabla 8. Preparación barril salmuera tipo Formiato de sodio.**

Esta tabla funciona para la preparación de un barril de salmuera tipo Formiato de sodio a cualquier densidad entre 8.4 a 11.1 ppg, donde la segunda columna corresponde a la cantidad de libras que se deben agregar por cada barril a preparar, y la tercera columna la cantidad de barriles de agua necesarios.

Para calcular lo necesario una cantidad de barriles a cierta densidad, se leen los valores en la fila correspondientes a la fila de esa densidad, luego se multiplica la cantidad de barriles finales a preparar por el valor de la segunda columna, obteniendo las libras de Formiato de sodio necesarias, finalmente se multiplica la cantidad de barriles finales a preparar por el valor de la tercera columna, obteniendo la cantidad de agua a la que se le debe adicionar esta sal.

Por ejemplo, si se necesitara preparar 100 bbl de salmuera NaCOOH al 11.1 ppg se requerirían 24638 lb, aproximadamente 447 sacos de 25 kg, y 70 bbl de agua.

$$NaCOOH = 100 * 246.38 = 24638\ lb \quad (69)$$

$$Vol_{agua} = 100 * 0.7 = 70\ bbbl \quad (70)$$

### 4.2. Preparación Salmuera Bivalente KCl + NaCOOH

Como resultado de la regresión generada en los numerales anteriores se genera la siguiente tabla realizando la conversión de ml a barriles equivalentes (350 ml):

| KCl | | |
|---|---|---|
| **Densidad lbs/gal** | **Concentración lbs/bbl** | **Volumen de agua bbl** |
| 9 | 47,900 | 0,943 |
| KCl 9 ppg + NaCOOH | | |
| **Densidad lbs/gal** | **Concentración lbs/bbl** | **Volumen de KCl 9 ppg bbl** |
| 9,1 | 5,150 | 0,997 |
| 9,2 | 14,392 | 0,994 |
| 9,3 | 23,634 | 0,981 |
| 9,4 | 32,877 | 0,968 |
| 9,5 | 42,119 | 0,954 |
| 9,6 | 51,361 | 0,941 |
| 9,7 | 60,603 | 0,928 |
| 9,8 | 69,846 | 0,915 |
| 9,9 | 79,088 | 0,902 |
| 10 | 88,330 | 0,889 |
| 10,1 | 97,573 | 0,875 |
| 10,2 | 106,815 | 0,862 |
| 10,3 | 116,057 | 0,849 |
| 10,4 | 125,300 | 0,836 |
| 10,5 | 134,542 | 0,823 |
| 10,6 | 143,784 | 0,809 |
| 10,7 | 153,026 | 0,796 |
| 10,8 | 162,269 | 0,783 |
| 10,9 | 171,511 | 0,770 |
| 11 | 180,753 | 0,757 |
| 11,1 | 189,996 | 0,744 |

**Tabla 9. Preparación barril salmuera tipo Cloruro de potasio y Formiato de sodio.**

Esta tabla funciona para la preparación de un barril de salmuera tipo Cloruro de potasio y Formiato de sodio a cualquier densidad entre 9.1 a 11.1 ppg, donde la segunda columna corresponde a la cantidad de libras que se deben agregar por cada barril a preparar, y la tercera columna la cantidad de barriles de agua o salmuera KCl al 9 ppg necesarios.

Para calcular lo necesario una cantidad de barriles a cierta densidad, se leen los valores en la fila correspondientes a la fila de esa densidad, luego se multiplica la cantidad de barriles finales a preparar por el valor de la segunda columna, obteniendo las libras de Formiato de sodio necesarias, a continuación, se multiplica la cantidad de barriles finales a preparar por el valor de la tercera columna, obteniendo la cantidad de salmuera tipo KCl al 9 ppg a la que se le debe adicionar esta sal. Para preparar estos barriles de salmuera tipo KCl al 9 ppg se multiplican los barriles de salmuera KCl al 9 ppg obtenidos en los pasos anteriores por 47.9 lb, obteniendo la cantidad de KCl necesario, y multiplicando estos mismos barriles por 0.943, obteniendo los barriles de agua necesarios.

Si se necesitara preparar 100 bbl de salmuera KCl + NaCOOH al 11.1 ppg se requerirían 19000 lb, aproximadamente 345 sacos de 25 kg de NaCOOH, y 74.4 bbl de KCl al 9 ppg, y para preparar los 74.4 bbl de salmuera KCl al 9 ppg se requiere de 3564 lb, aproximadamente 65 sacos de 25 kg de KCl, y 70.16 bbl de agua.

$$NaCOOH = 100 * 189.996 = 18999.6\ lb \quad (71)$$

$$Vol_{KCl\ 9ppg} = 100 * 0.744 = 74.4\ bbbl \quad (72)$$

$$KCl = 74.4 * 47.9 = 3563.76\ lb \quad (73)$$

$$Vol_{agua} = 74.4 * 0.943 = 70.16\ bbl \quad (74)$$

# 5. Análisis Económico

Se comparará la cantidad de dinero necesaria para producir una cierta cantidad de salmuera a cierta densidad con la salmuera tipo NaCOOH y tipo KCl + NaCOOH para analizar cuál es más viable económicamente. Aunque se tienen los resultados de la cantidad de sal necesaria en campo para el segundo tipo de salmuera en las tres preparaciones, se van a utilizar los datos de laboratorio ya que para el primer tipo no se conoce su valor real en campo.

Cuando se realizó el análisis económico el kg de KCl costaba $1690 COP y el kg de NaCOOH $1915 COP.

## 5.1. Preparación 103 bbl Salmuera Al 11.1 Ppg

- **Tipo NaCOOH**

| | |
|---|---|
| Kg NaCOOH | 11535 |
| Sacos 25 Kg | 462 |
| Costo NaCOOH | $ 22.118.250 |

- **Tipo KCl + NaCOOH**

| | |
|---|---|
| Vol KCl 9,0 ppg | 76,51 |
| Kg KCl | 1666 |
| Sacos 25 Kg | 67 |
| Costo KCl | $ 2.830.750 |
| Kg NaCOOH | 8895 |
| Sacos 25 Kg | 356 |
| Costo NaCOOH | $ 17.043.500 |
| Costo TOTAL | $ 19.874.250 |

Se esperaría obtener un ahorro de $2.244.000 COP preparando 103 bbl salmuera al 11.1 ppg.

## 5.2. Preparación 200 bbl Salmuera Tipo Al 11 ppg

- **Tipo NaCOOH**

| | |
|---|---|
| Kg NaCOOH | 21559 |
| Sacos 25 Kg | 863 |
| Costo NaCOOH | $ 41.316.125 |

- **Tipo KCl + NaCOOH**

| | | |
|---|---|---|
| Vol KCl 9,0 ppg | | 151,21 |
| Kg KCl | | 3292 |
| Sacos 25 Kg | | 132 |
| Costo KCl | $ | 5.577.000 |
| Kg NaCOOH | | 16432 |
| Sacos 25 Kg | | 658 |
| Costo NaCOOH | $ | 31.501.750 |
| Costo TOTAL | $ | 37.078.750 |

Se esperaría obtener un ahorro de $4.237.375COP preparando 200 bbl salmuera al 11 ppg.

### 5.3. Preparación 100 bbl Salmuera Tipo Al 10.7 ppg

- **Tipo NaCOOH**

| | | |
|---|---|---|
| Kg NaCOOH | | 11425 |
| Sacos 25 Kg | | 457 |
| Costo NaCOOH | $ | 21.878.875 |

- **Tipo KCl + NaCOOH**

| | | |
|---|---|---|
| Vol KCl 9,0 ppg | | 94,8 |
| Kg KCl | | 2064 |
| Sacos 25 Kg | | 83 |
| Costo KCl | $ | 3.506.750 |
| Kg NaCOOH | | 8347 |
| Sacos 25 Kg | | 334 |
| Costo NaCOOH | $ | 15.990.250 |
| Costo TOTAL | $ | 19.497.000 |

Se esperaría obtener un ahorro de $2.381.875 COP preparando 100 bbl salmuera al 10.7 ppg.

En todas las preparaciones se obtuvo un costo menor al preparar una salmuera bivalente de KCl + NaCOOH en contra de la monovalente de NaCOOH, esto sin tener en cuenta que la disponibilidad de NaCOOH suele ser menor y más difícil de conseguir en ciudades lejos del puerto, además de que el uso de NaCOOH en preparaciones requiere mucho mas trabajo dado a que tiende a endurecerse formando “rocas”, lo que genera aumentos de tiempos operativos y desgaste a los operadores ya que es necesario realizar trabajos físicos fuertes sobre el NaCOOH.

## 6. Conclusiones

- Se logro obtener las cantidades de sal y volúmenes de agua y salmuera necesarios para preparar a distintas densidades salmueras monovalentes de Formiato de sodio y bivalente de Formiato de sodio con Cloruro de potasio.
- Se probaron los resultados en campo con grandes preparaciones de salmuera tipo KCl + NaCOOH obteniendo diferencias aceptables con el modelo obtenido.

## 7. Recomendaciones

- Se recomienda realizar pruebas de laboratorio a densidades mas bajas para la salmuera monovalente de NaCOOH ya que se observaron ciertas discrepancias cuando se buscan estas densidades en campo, pero en este trabajo no se le dieron prioridad ya que en campo no se realizan estas preparaciones de NaCOOH a bajas densidad ya que el KCl es más económico y fácil de preparar cumpliendo la misma función.
- Aunque se llegó a la conclusión de que no hay diferencias significativas entre proveedores, el NaCOOH de ciertos proveedores requiere de un manejo especial generando un mayor trabajo.
- Se observó que iniciando con salmueras de densidad de hasta 9.3 ppg, tipo KCl, se puede lograr alcanzar densidades de hasta 11.1 ppg adicionando NaCOOH, por tanto, se recomienda en futuros trabajos realizar un análisis de estas salmueras para confirmar si hay un cambio en el costo total considerable.

# 8. CODIGO EN R

1. Aleatorización experimento

```
require(agricolae)
tratamiento=c("Proveedor1","Proveedor2","Proveedor3")
rep=c(5,5,5)
ran=design.crd(trt=tratamiento,r=rep,seed=123)
ran
```

2. Análisis del diseño experimental

```
require(readxl)
setwd('D:\\Educación\\Estadistica ESPECIALIZACIÓN\\Trabajo de grado')
base=read_excel("CALCULOS PREPARACIÓN KCL+NACOOH.xlsx","ANOVA")
attach(base)
names(base)
table(Proveedor,NaCOOH)
boxplot(Densidad~Proveedor)
boxplot(Densidad~NaCOOH)
m=lm(Densidad~factor(Proveedor)+factor(NaCOOH))
anova(m)
require(agricolae)
require(car)
t.test(m$residuals)
shapiro.test(m$residuals)
qqPlot(m$residuals)
hist(m$residuals)
lmtest::bptest(m)
dwt(m) #Solo revisa la autocorrelación de orden 1

detach(base)
```

3. Regresión salmuera monovalente NaCOOH

```
require(readxl)
require(agricolae)
require(car)
library(lmtest)
setwd('D:\\Educación\\Estadistica ESPECIALIZACIÓN\\Trabajo de grado')
base=read_excel("CALCULOS PREPARACIÓN KCL+NACOOH.xlsx","NaCOOH")
attach(base)
names(base)
```

```
MODNaCOOH<- lm(NaCOOH~Densidad)
summary(MODNaCOOH)
resid<-residuals(MODNaCOOH)
t.test(resid)
shapiro.test(resid)
lmtest::bptest(MODNaCOOH)
hist(MODNaCOOH$residuals)
qqPlot(MODNaCOOH$residuals)
dwtest(MODNaCOOH,alternative="two.sided")
MODNaCOOH2<- lm(VolAgua~Densidad)
summary(MODNaCOOH2)
resid<-residuals(MODNaCOOH2)
t.test(resid)
shapiro.test(resid)
lmtest::bptest(MODNaCOOH)
hist(MODNaCOOH$residuals)
qqPlot(MODNaCOOH$residuals)
dwtest(MODNaCOOH,alternative="two.sided")
```

4. Regresión salmuera bivalente KCl (9 ppg ) + NaCOOH

```
require(readxl)

require(agricolae)

require(car)

library(lmtest)

setwd('D:\\Educación\\Estadistica ESPECIALIZACIÓN\\Trabajo de grado')

base=read_excel("CALCULOS PREPARACIÓN
KCL+NACOOH.xlsx","KCl+NaCOOH")

attach(base)

names(base)

MODNaCOOH<- lm(NaCOOH~Densidad)

summary(MODNaCOOH)

resid<-residuals(MODNaCOOH)

t.test(resid)

shapiro.test(resid)

lmtest::bptest(MODNaCOOH)
```

```
hist(MODNaCOOH$residuals)
qqPlot(MODNaCOOH$residuals)
dwtest(MODNaCOOH,alternative="two.sided")
MODNaCOOH2<- lm(VolKCl~Densidad)
summary(MODNaCOOH2)
resid<-residuals(MODNaCOOH2)
t.test(resid)
shapiro.test(resid)
lmtest::bptest(MODNaCOOH)
hist(MODNaCOOH$residuals)
qqPlot(MODNaCOOH$residuals)
dwtest(MODNaCOOH,alternative="two.sided")
```

# 9. Bibliografía

Bencardino, C. M. (2012). *Estadística y muestreo.* Bogotá: Ecoe ediciones.

Ortiz, A. F. (s.f.). *Diseño experimental (Diapositivas de clase).*

Piñerez, C. T. (2016). *Modelos estadísticos lineales. Con aplicaciones en R.* Ediciones de la U.

School, W. C. (s.f.). *Manual de control de pozos.*

Wackerly, D. D., & Mendenhall III, W. y. (2010). *Estadística matemática con aplicaciones.* México: Cengage Learning Editores, S.A.

Swartwout R, Pearcy R, Tetra tecnologies, (1996). Design and Application of Brine-Based Drilling Fluids. Society of Petroleum Engineers (SPE 35332)

Howard Siv K, Shell Research BV. (1995). Formate Brines for Drilling and Completion: State of the Art. Society of Petroleum Engineers (SPE 30498)

Baroid, a Halliburton company. (2000). MANUAL DE FLUIDOS. Capítulo 1. Fluidos de terminación

Brinadd. Tablas de salmueras. Obtenido de: https://www.academia.edu/18599993/134081676_Tablas_Salmueras_Completas