1. Las variables escogidas fueron: Calidad del vino, alcohol, pH

Calidad del vino: Toma valores entre 0 y 10

0-3: Calidad mala

4-7: Caliad normal

8-10: Calidad buena

Media: 5.636023. El vino en promedio tiene una calidad de 6 puntos aproximadamente lo cual representa una calidad normal.

Mediana: 6. El 50% del vino analizado tiene una calidad igual o menor a 6 puntos.

Moda: La mayoría de las muestras de vino analizadas tienen una calidad de 5 puntos lo cual significa que la calidad de la mayor parte del vino analizado es normal

Varianza: 0.6519644

Desviación Típica: 0.8074431

Coeficiente de variación: 14.3264. Los datos se consideran homogéneos

**2.** Genere una Población Normal de tamaño 500, seleccione 8 muestras de tamaños varios (Mucho mayor que 30, mayor que 30, 30, 20), 4 muestras con remplazo y 4 sin remplazo.

Código en R:

#generate a normal population of size 500

p <- rnorm(500, 0, 1)

#generate 4 samples without replace

s1 <- sample(p, size = 300)

s2 <- sample(p, size = 40)

s3 <- sample(p, size = 30)

s4 <- sample(p, size = 20)

#generate 4 samples with replace

sr1 <- sample(p, size = 300, replace = TRUE)

sr2 <- sample(p, size = 40, replace = TRUE)

sr3 <- sample(p, size = 30, replace = TRUE)

sr4 <- sample(p, size = 20, replace = TRUE)

Generar una muestra con reemplazo significa que puedes escoger el mismo elemento de una población varias veces a la hora de escoger una muestra.

* 1. **a.** Calcule para cada una de las muestras los Estadísticos Descriptivos, de la Conferencia 1.
  2. **b. Calcúlelos en la población inicial. Analice las diferencias.**

1. La función measures definida en el script code.R es la encargada de calcular las métricas de la Conf 1(media, varianza, etc):

measures = function(x)

{

list(

Mean = mean(x),

Median = median(x),

Mode = get\_mode(x),

Variance = var(x),

Standard\_Deviation = sd(x),

Coefficient\_of\_variation = (sd(x)/abs(mean(x))) \* 100

)

}

La diferencia entre las muestras y la población inicial es que en las muestras las métricas son mas alejadas del valor real de la población a medida que decrece el tamaño de las muestras.

* 1. **c.** Grafique los resultados
  2. La función display es la encargada de construir el gráfico de caja y bigotes y el histograma de frecuencias.

display = function(x)

{

boxplot(x, horizontal = TRUE, col = "orange")

hist(x)

}

* 1. **d. Para cada muestra calcule los intervalos de confianza para la media y la varianza**

Para calcular los intervalos de confianza se procede exactamente como se esplica en la conferencia 2:

mean\_confidence\_interval = function(s, alpha)

{

# s: sample

# alpha: confidence level

n <- length(x)

m <- mean(s)

sd <- sd(s)

if (n > 30)

{

z <- qnorm(1 - alpha/2)

left <- m - z \* sd/sqrt(n)

right <- m + z \* sd/sqrt(n)

}

else

{

t <- qt(1 - alpha/2, n - 1)

left <- m - t \* sd/sqrt(n)

right <- m + t \* sd/sqrt(n)

}

glue("Mean confidence interval with confidence level {alpha}:\nLeft bound: {left}\nRight bound: {right}")

}

variance\_confidence\_interval = function(s, alpha)

{

# s: sample

# alpha: confidence level

n <- length(x)

v <- var(s)

x1 <- qchisq(1 - alpha/2, n - 1)

x2 <- qchisq(alpha/2, n - 1)

left <- (n - 1) \* v / x1

right <- (n - 1) \* v / x2

glue("Variance confidence interval with confidence level {alpha}:\nLeft bound: {left}\nRight bound: {right}")

}

* **glue es un paquete de R**
  1. e) Analice las diferencias en las muestras de tamaños similares: Por ver

1. 3-) ¿Se puede afirmar que el vino de color blanco tiene en promedio menor acidez fija que el vino de color rojo?
2. *Sugerencia:* Asuma que todas las observaciones provienen de una distribución normal.

Se tiene una prueba de hipótesis de comparación de la media entre dos poblaciones. La prueba de hipótesis seria la siguiente:

H0: u1 >= u2

H1: u2 < u1

El siguiente código permite hacer dicha prueba de hipótesis en R, con un nivel de confianza de 99.5 %:

white\_wine <- subset(wine\_table, color == "white")

red\_wine <- subset(wine\_table, color == "red")

t.test(red\_wine$fixed.acidity, white\_wine$fixed.acidity, conf.level = 0.995, alternative = “l”)

Al observer el resultado del test, el p-valor = 0.5, es el menor para el cual se rechaza H0. Pero alfa es la probabilidad de rechazar H0 siendo Ho verdadera, o sea, la prob de cometer un error de tipo 1. Por tanto no es factible tomar alfa >= 0.5 pues al menos en la mitad de los casos se cometeria un error de tipo 1 y por eso no se puede rechazar la hipótesis nula.