**Taller de Análisis exploratorio de bases de datos**

A partir del análisis de la base de datos de **devoluciones**

1. Seleccionar los usos que representan el 80% de las devoluciones y los usos que representan el 80% de los costos de devoluciones.
2. De los usos seleccionados en el literal 1. seleccionar las referencias que representan el 80% de los costos de devoluciones.
3. De las referencias seleccionadas en el ítem 2. (T307EO) seleccionar las variables que tienen mayor variabilidad.
4. Caracterizar las variables que podrían tener mayor incidencia en el mejoramiento del proceso.

**Solución**

datos3 <- read.table(file="https://tinyurl.com/y9oeddzg",header=T, sep="\t", dec=".")

# Usos vs devoluciones

require(qcc)

par(mar=c(8,6,4,4))

pareto.chart(table(datos3$uso) , ylab="Número de devoluciones" ,

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Usos vs Número de devoluciones",

cex.axis=0.6)

# Usos vs Costos de devoluciones

datos3$uso <- factor(datos3$uso)

costos\_uso <- tapply(datos3$costo, datos3$uso, FUN=sum)

pareto.chart(costos\_uso , ylab="Costos de devoluciones",

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Usos vs Costos de devoluciones",

cex.axis=0.6)

# Referencias vs Costos - Atun

datos3\_sub1<- subset(datos3, uso=="ATUN")

datos3\_sub1$referencia <- factor(datos3\_sub1$referencia)

costos\_ref1 <- tapply(datos3\_sub1$costo, datos3\_sub1$referencia, FUN=sum)

pareto.chart(costos\_ref1 , ylab="Costos de devoluciones",

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Referencias de Atun vs Costos de devoluciones",

cex.axis=0.6)

datos3\_sub2<- subset(datos3, uso=="SARDINA")

datos3\_sub2$referencia <- factor(datos3\_sub2$referencia)

costos\_ref2 <- tapply(datos3\_sub2$costo, datos3\_sub2$referencia, FUN=sum)

pareto.chart(costos\_ref2 , ylab="Costos de devoluciones",

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Referencias vs Costos de devoluciones",

cex.axis=0.6)

datos3\_sub3<- subset(datos3, uso=="PEGA")

datos3\_sub3$referencia <- factor(datos3\_sub3$referencia)

costos\_ref3 <- tapply(datos3\_sub3$costo, datos3\_sub3$referencia, FUN=sum)

pareto.chart(costos\_ref3 , ylab="Costos de devoluciones",

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Referencias vs Costos de devoluciones",

cex.axis=0.6)

datos3\_sub4<- subset(datos3, uso=="DILUY")

datos3\_sub4$referencia <- factor(datos3\_sub4$referencia)

costos\_ref4 <- tapply(datos3\_sub4$costo, datos3\_sub4$referencia, FUN=sum)

pareto.chart(costos\_ref4 , ylab="Costos de devoluciones",

ylab2="Porcentaje acumulado",

cumperc=seq(0, 100, by=20), las=2,

main="Diagrama de Pareto de Referencias vs Costos de devoluciones",

cex.axis=0.6)

aggregate(datos2$medicion ~ datos2$variable, FUN=function(x) { c(mean(x) , sd(x) , (sd(x)/mean(x))\*100 )})

# Ejemplo para caracterizar Apertura Anillo (POP) L-D

datos2\_sub2 <- subset(datos2, variable=="Apertura Anillo (POP) L-D" )

hist(datos2\_sub2$medicion, col="coral",

main="Histograma de Apertura Anillo (POP) L-D",

xlim=c(10,26), ylim=c(0,2000), xlab="Apertura Anillo (POP) L-D",

ylab="Número de mediciones")

abline(v=c(11, 18, 25), lty=3)

boxplot(datos2\_sub2$medicion, col="coral",

main="Boxplot de Apertura Anillo (POP) L-D",

ylim=c(11,25), ylab="Apertura Anillo (POP) L-D")

abline(h=c(11, 18, 25), lty=3)

summary(datos2\_sub2$medicion)

boxplot(datos2\_sub2$medicion ~ datos2\_sub2$turno, col="coral",

main="Boxplot de Apertura Anillo (POP) L-D",

ylim=c(11,25), ylab="Apertura Anillo (POP) L-D")

abline(h=c(11, 18, 25), lty=3)

plot(datos2\_sub2$medicion, type="l", main="Muestra vs Apertura Anillo (POP) L-D",

xlab="Muestra", ylab="Apertura Anillo (POP) L-D",

ylim=c(11, 25), col="coral")

abline(h=c(11, 18, 25), lty=3)