

# Introducción al mejoramiento de procesos de Ingeniería usando R

Olga Cecilia Usuga Manco



II Simposio Data Analytics



<https://github.com/ousuga/Mejoramiento-de-procesos>



Olga Cecilia Usuga Manco

[olga.usuga@udea.edu.co](mailto:olga.usuga@udea.edu.co)

Ingeniera Industrial, MSc en Estadística, PhD en Ciencias-Estadística

Departamento de Ingeniería Industrial, Facultad de Ingeniería,  
Universidad de Antioquía

II Simposio Data Analytics

Septiembre 19-21, 2018

# Sesión I

Exploración de bases de datos

# ¿Qué es R y RStudio?



Motor



Tablero



# Instalación de R

Descargar e instalar R y Rstudio en el computador.

## 1. Bajar e instalar R

- Primero se debe instalar R.
- [Instalación de R.](#)

# Instalación de RStudio

## 2. Bajar e instalar Rstudio

- [Instalación de Rstudio.](#)

## R a través de RStudio

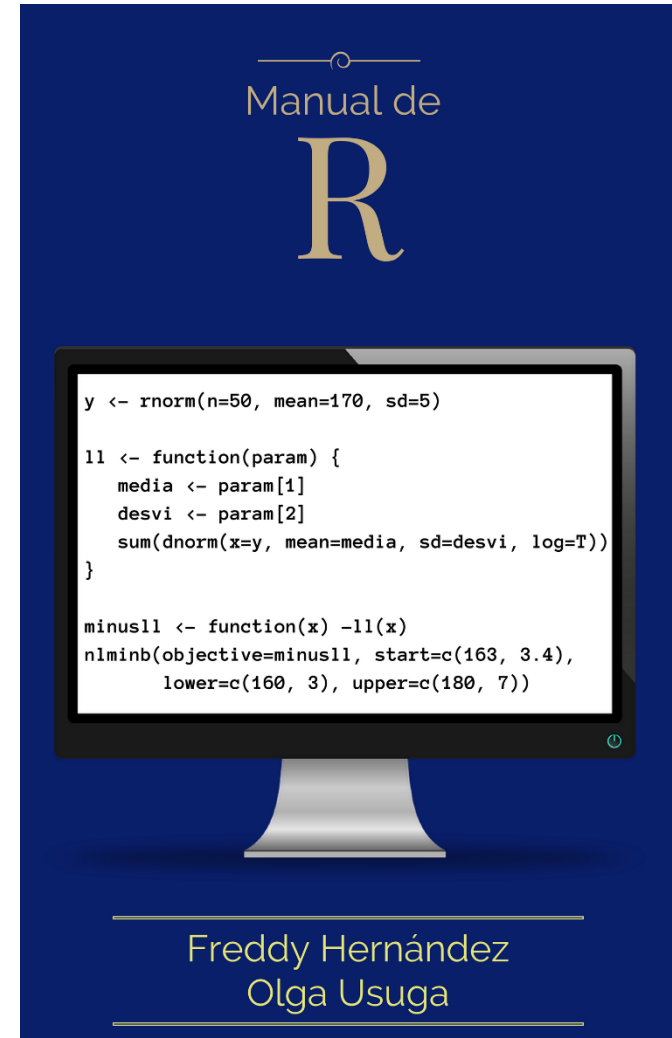




# Manual de R

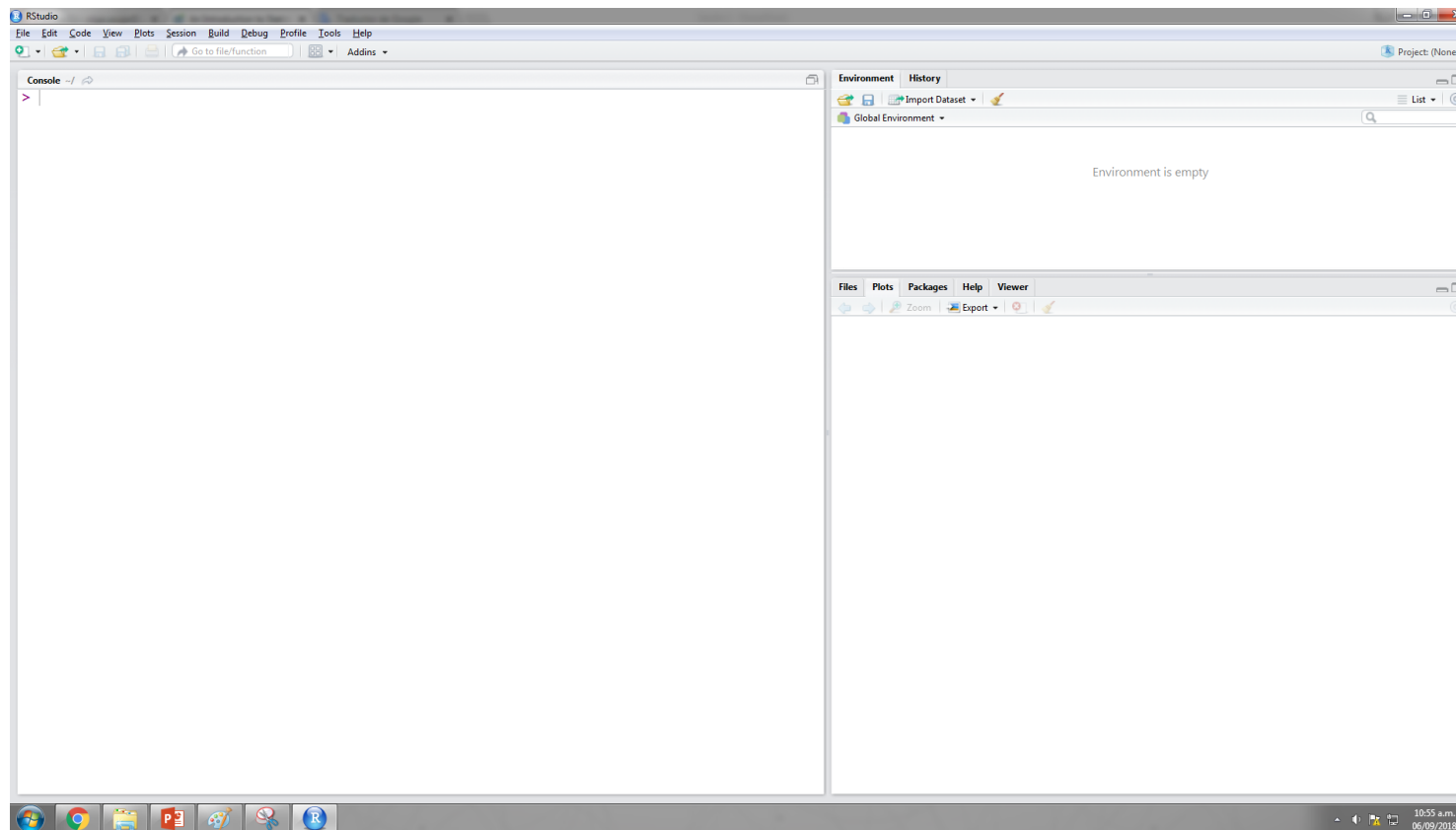
Hernández, F. y Usuga, O. Manual de R. 2018.

[https://bookdown.org/fhernanb/Manual\\_de\\_R](https://bookdown.org/fhernanb/Manual_de_R)



# R a través de RStudio

Luego de abrir Rstudio se vera esto.



# Introducción a R

Tipos de objetos

- **Vectores**
- Matrices
- Arreglos
- Listas
- **Data Frames**

# Operadores

Operadores aritméticos		Operadores comparativos		Operadores lógicos	
+	Adición	<	Menor que	! x	No x
-	Sustracción	>	Mayor que	x & y	x o y
*	Multiplicación	<=	Menor o igual que	x   y	x y y
/	División	>=	Mayor o igual que	isTRUE (x)	Prueba si x es cierta
^	Potencia	==	Igual		
%%	Módulo	!=	Diferentes		
%/%	División de enteros				

# ¿Qué son los paquetes?

R: Nuevo Teléfono



Paquetes de R: Aplicaciones que puedes descargar



# Instalación de paquetes

Dos formas para instalar un paquete, por ejemplo el paquete `qcc`

**Forma 1:** En el panel Files de Rstudio

- Click en la pestaña “Packages”
- Click en “Install”
- Escribir el nombre del paquete que se quiere instalar. En este caso `qcc`.
- Click en “Install”

**Forma 2:** En el panel de la consola correr `install.package("qcc")`

**Nota:** solo debe instalar un paquete una vez, a menos que desee actualizar un paquete ya instalado a la última versión.

# Lectura de bases de datos

## Ejemplo

```
datos1 <- read.table(file="https://tinyurl.com/y7hwtzu6",  
                     header=T, sep="\t", dec=".")
```

# Extracción de una variable de la base de datos

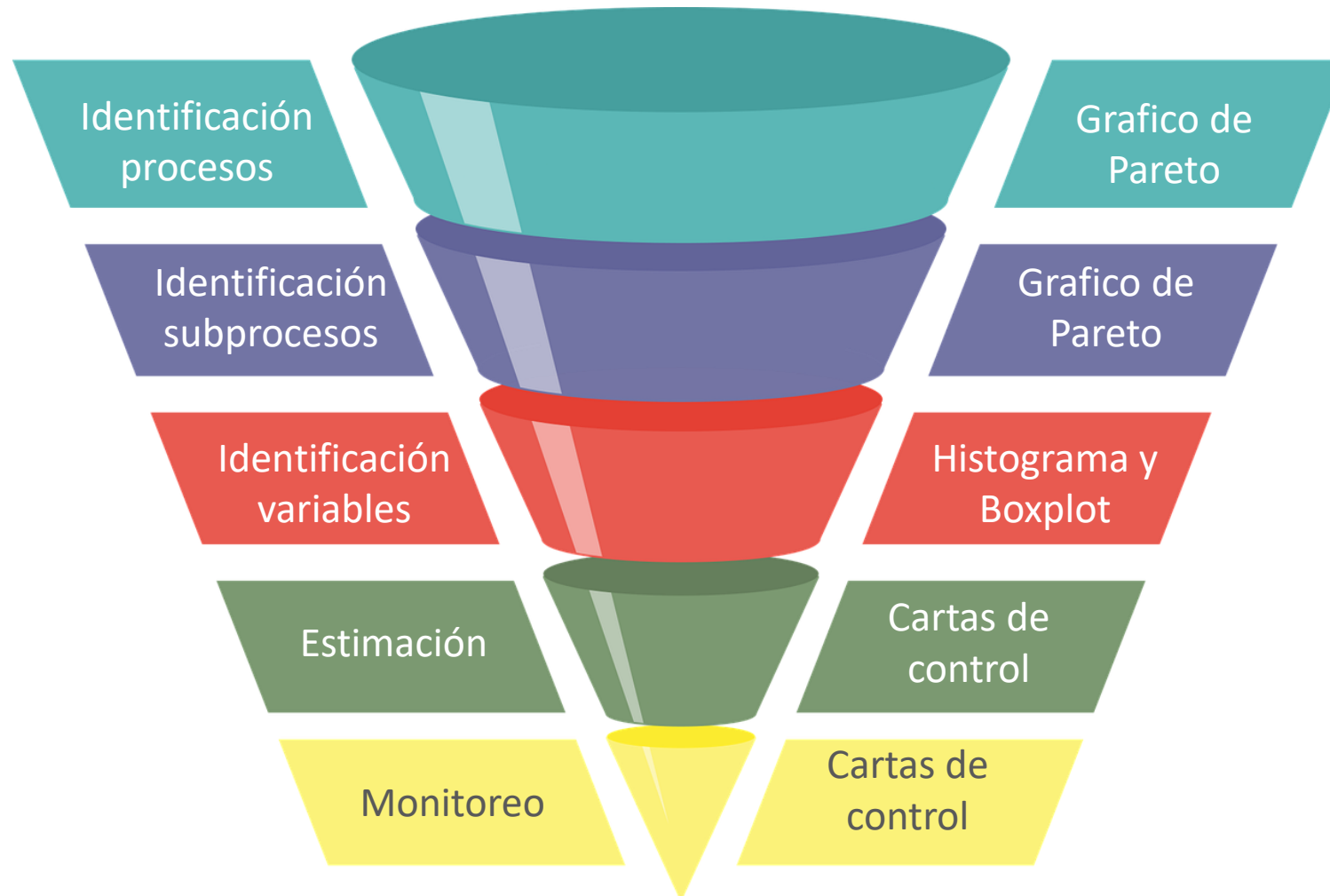
## Ejemplo

```
datos1$costos
```

```
datos1$cargo
```



# Control estadístico de procesos



# Control estadístico de procesos en empresa de envases de hojalata

**Objetivo:** Identificar y analizar las variables críticas que afecten a la calidad del producto para así proponer actividades que estén direccionadas a disminuir las causas de variación y al cumplimiento de los requerimientos.



## Bases de datos a usar



1. año
2. cargo
3. fase
4. lamina
5. referencia
6. uso
7. espesor
8. estaño
9. costo



1. variable
2. min
3. estandar
4. max
5. dia
6. mes
7. año
8. turno
9. inspector
10. pto\_control
11. muestra
12. oac
13. medicion



1. año
2. uso
3. referencia
4. lamina
5. producto
6. espesor
7. Estaño
8. barnizint
9. barnizext
10. costo

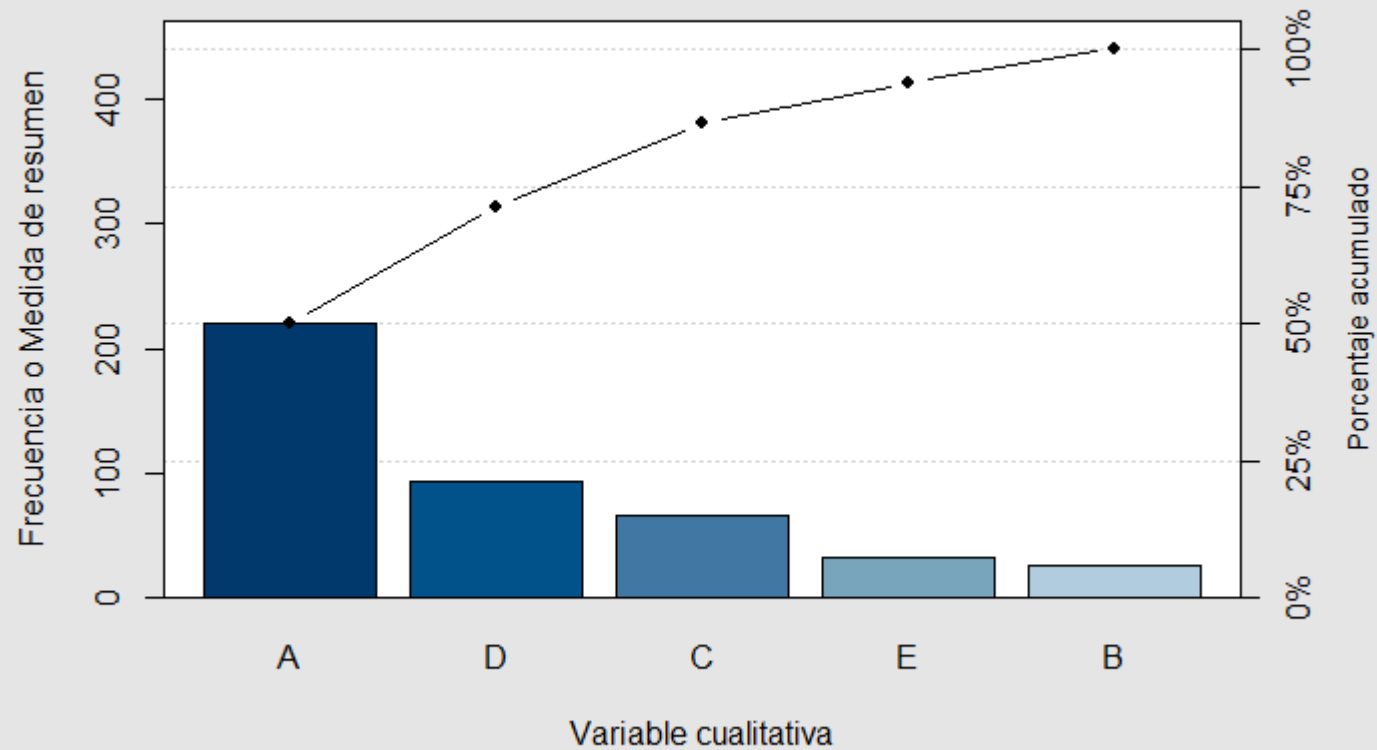
## Lectura de datos

```
datos1 <- read.table(file="https://tinyurl.com/y7hwtzu6",  
                     header=T, sep="\t", dec=".")
```

```
datos2 <- read.table(file="https://tinyurl.com/ybuunbt5",  
                     header=T, sep="\t", dec=".")
```

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

## Diagrama Pareto



# Diagrama de Pareto

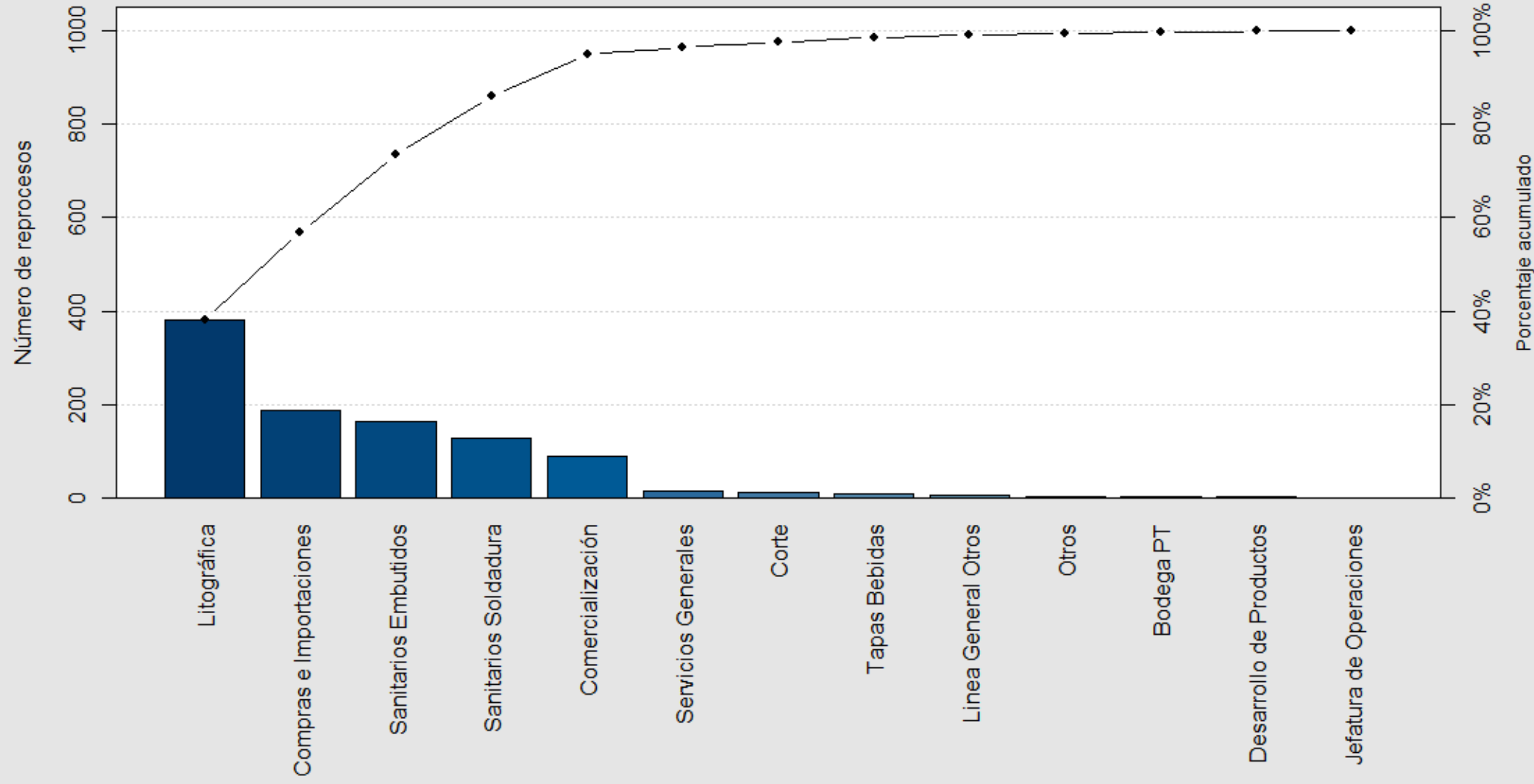
## Usos

- Identificar un producto o servicio para mejorar la calidad.
- Llamar la atención a los problema o causas de una forma sistemática.
- Identificar oportunidades para mejorar.
- Analizar las diferentes agrupaciones de datos (ej: por producto, por segmento, del mercado, etc.).
- Buscar las causas principales de los problemas y establecer la prioridad de las soluciones.
- Evaluar los resultados de los cambios efectuados a un proceso (antes y después).

## Diagrama de Pareto de Cargo vs Número de reprocesos en R

```
require(qcc)
par(mar=c(8,6,4,4))
datos1$cargo <- factor(datos1$cargo)
pareto.chart(table(datos1$cargo) , ylab="Número de reprocesos",
              ylab2="Porcentaje acumulado", cumperc=seq(0, 100, by=20),
              las=2, main="Diagrama de Pareto de Cargo vs Número de
              reprocesos", cex.axis=0.6)
```

Diagrama de Pareto de Cargo vs Número de reprocesos





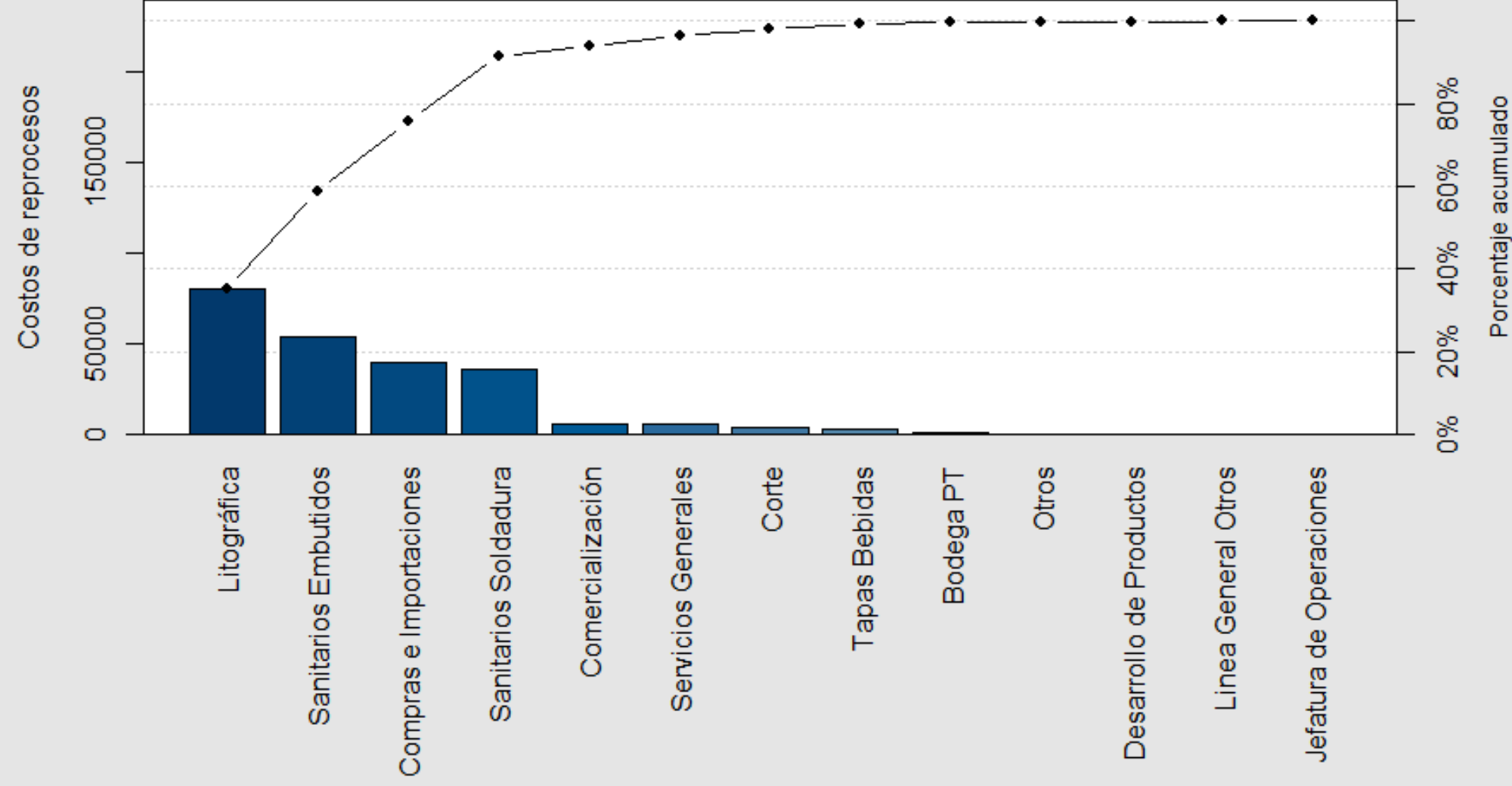
## Conclusión

Del diagrama de Pareto se observa que las áreas de litografía, compras e importaciones, sanitarios embutidos y sanitarios soldadura son las que representan el 86% de la cantidad total de productos en reproceso.

## Diagrama de Pareto de Cargo vs Costos de reproceso en R

```
par(mar=c(8,6,4,4))
datos1$cargo <- factor(datos1$cargo)
costos_rep <- tapply(datos1$costo, datos1$cargo, FUN=sum)
pareto.chart(costos_rep , ylab="Costos de reprocesos", ylab2="Porcentaje
acumulado", cumperc=seq(0, 100, by=20), las=2, main="Diagrama
de Pareto de Cargo vs Costos de reproceso", cex.axis=0.6)
```

Diagrama de Pareto de Cargo vs Costos de reproceso



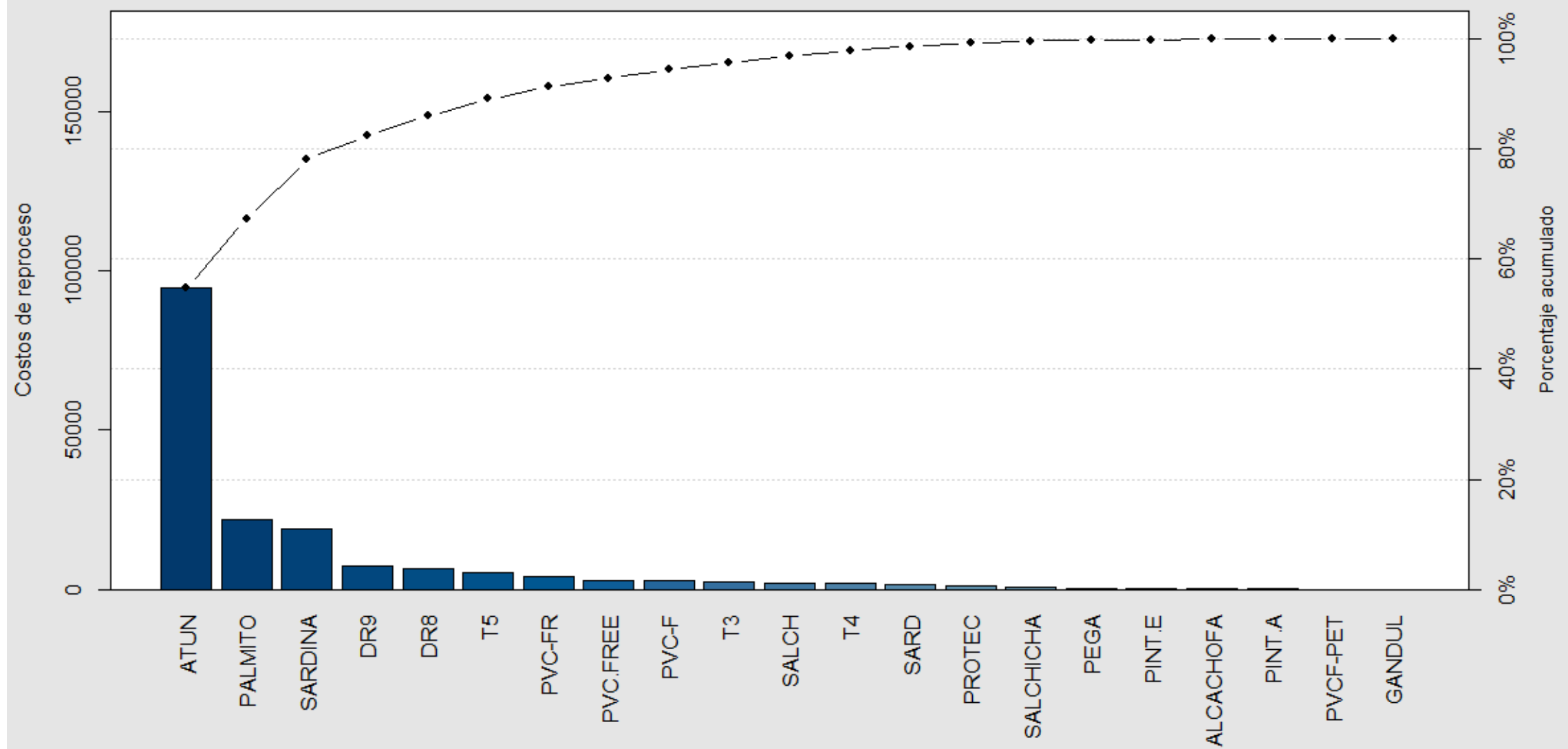
## Conclusión

Del diagrama de Pareto se observa que las áreas de litografía, sanitarios embutidos y compras e importaciones son las que representan el 75.8% de los costos totales de productos en reproceso.

## Diagrama de Pareto de Sub\_Cargo vs Costos de reproceso en R

```
par(mar=c(5, 4, 4, 2))
datos1_sub1<- subset(datos1, cargo=="Litográfica" |
                      cargo=="Sanitarios Embutidos" |
                      cargo=="Compras e Importaciones")
datos1_sub1$uso <- factor(datos1_sub1$uso)
costos_uso <- tapply(datos1_sub1$costo, datos1_sub1$uso, FUN=sum)
pareto.chart(costos_uso , ylab="Costos de reproceso", ylab2="Porcentaje
acumulado", cumperc=seq(0, 100, by=20), las=2,
main="Diagrama de Pareto de Sub_Cargo vs Costos de
reproceso", cex.axis=0.6)
```

Diagrama de Pareto de Sub\_Cargo vs Costos de reproceso



## Conclusión

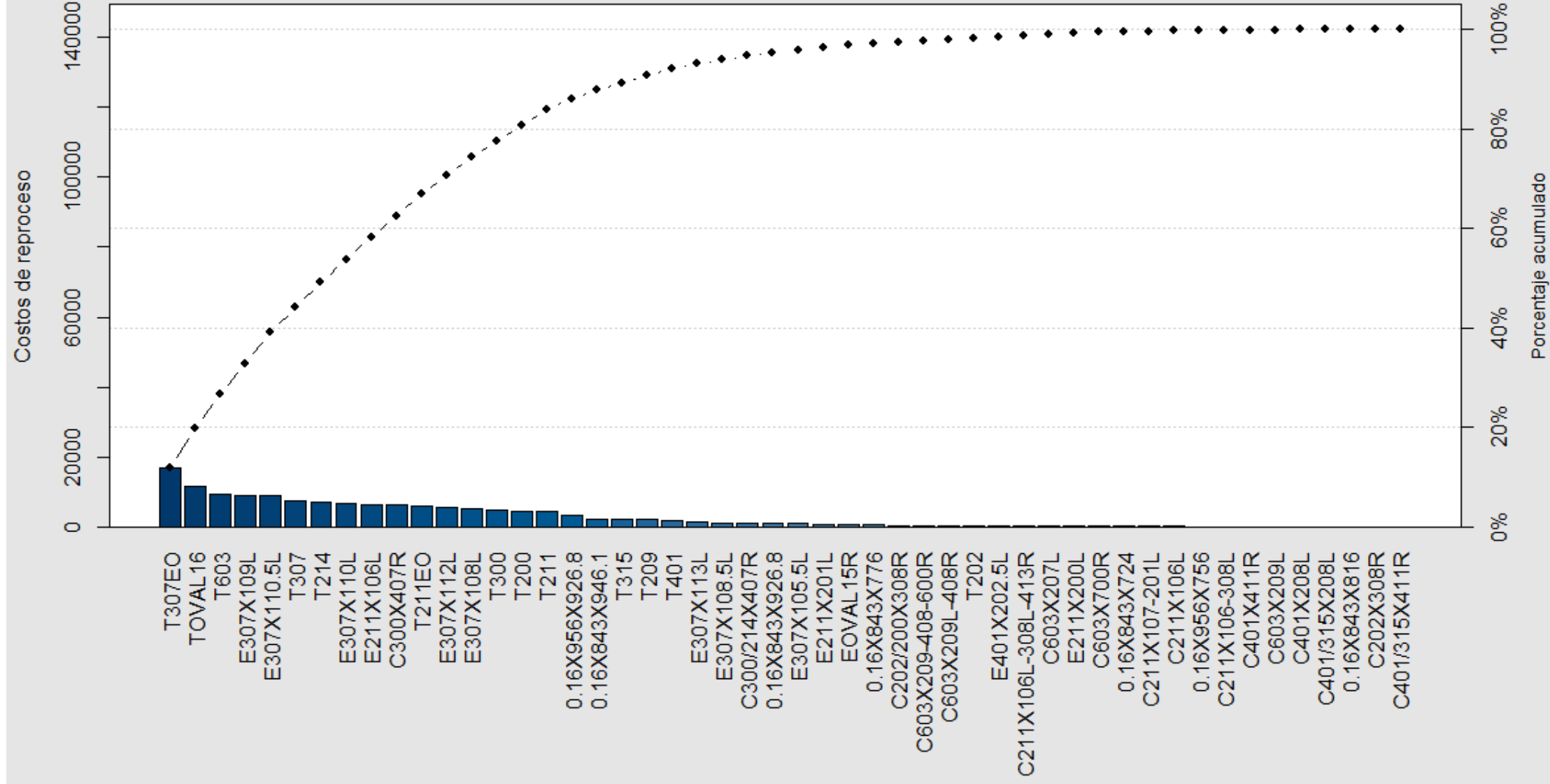
Del diagrama de Pareto se observa que de las áreas de Litografía, Sanitarios embutidos y Compras e importaciones, los usos que representan el 82,3% de los costos totales de productos en reproceso corresponden a Atún, Palmito, Sardina y DR9.

## Diagrama de Pareto de Sub\_Uso vs Costos de reproceso en R

```
par(mar=c(6, 4, 4, 2))
datos1_sub2<- subset(datos1, (cargo=="Litográfica" |
                             cargo=="Sanitarios Embutidos" |
                             cargo=="Compras e Importaciones") &
                             (uso=="ATUN" | uso=="PALMITO" | uso=="SARDINA" |
                             uso=="DR9"))
datos1_sub2$referencia <- factor(datos1_sub2$referencia)
costos_uso <- tapply(datos1_sub2$costo, datos1_sub2$referencia, FUN=sum)
pareto.chart(costos_uso , ylab="Costos de reproceso", ylab2="Porcentaje
acumulado", cumperc=seq(0, 100, by=20), las=2,
main="Diagrama de Pareto de Sub_Uso vs Costos de reproceso",
cex.axis=0.6)
```



Diagrama de Pareto de Sub\_Uso vs Costos de reproceso



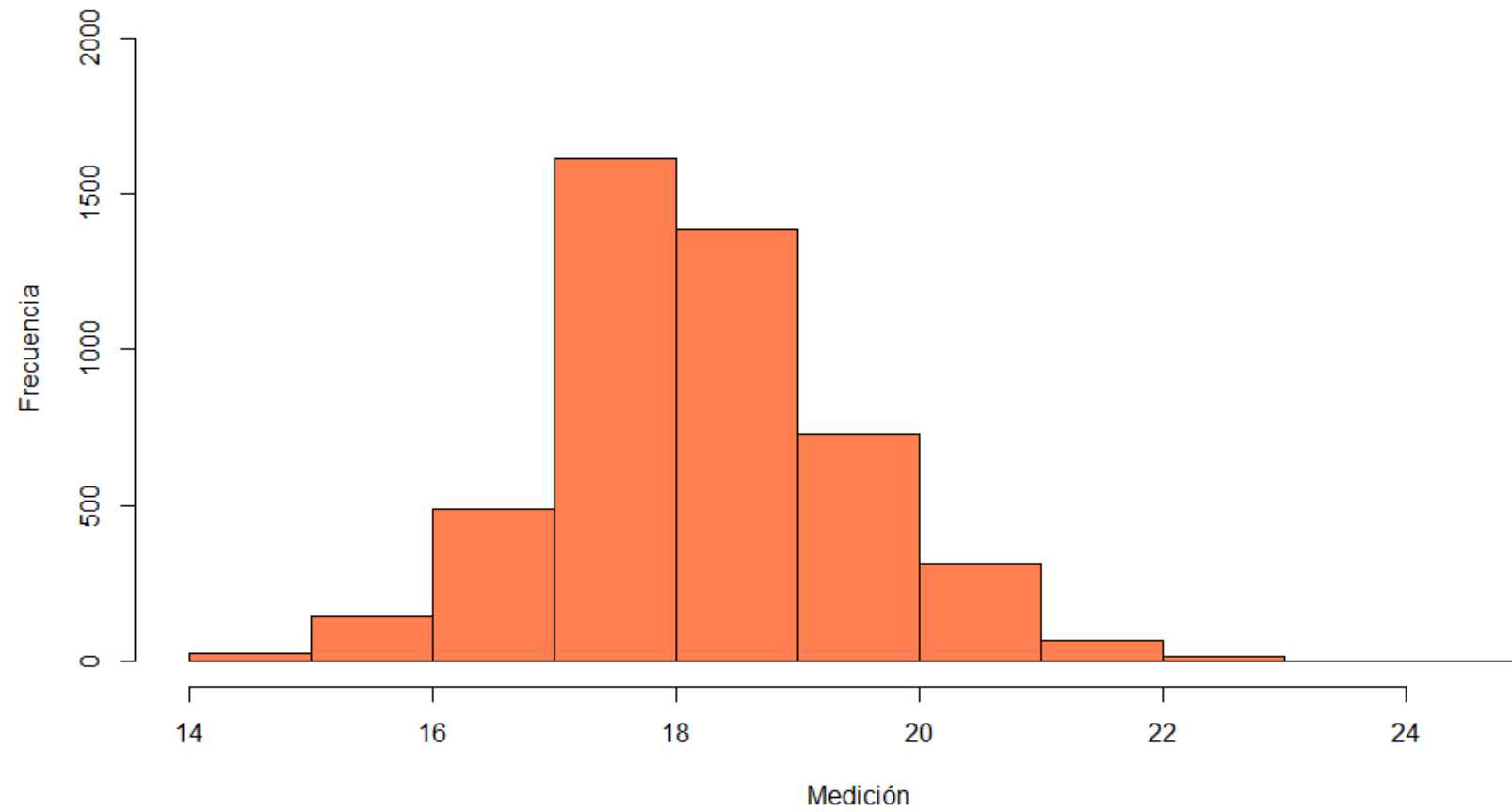
## Conclusión

Del diagrama de Pareto se observa que de las áreas de Litografía, Sanitarios embutidos y Compras e importaciones y los usos de Atún, Palmito, Sardina y DR9, las referencias que representan el 80.77% de los costos totales de productos en reproceso corresponden a 15 Referencias, siendo las 4 primeras T307EO, TOVAL16, T603 y E307X109L.

T307EO



# Histograma



# Histograma

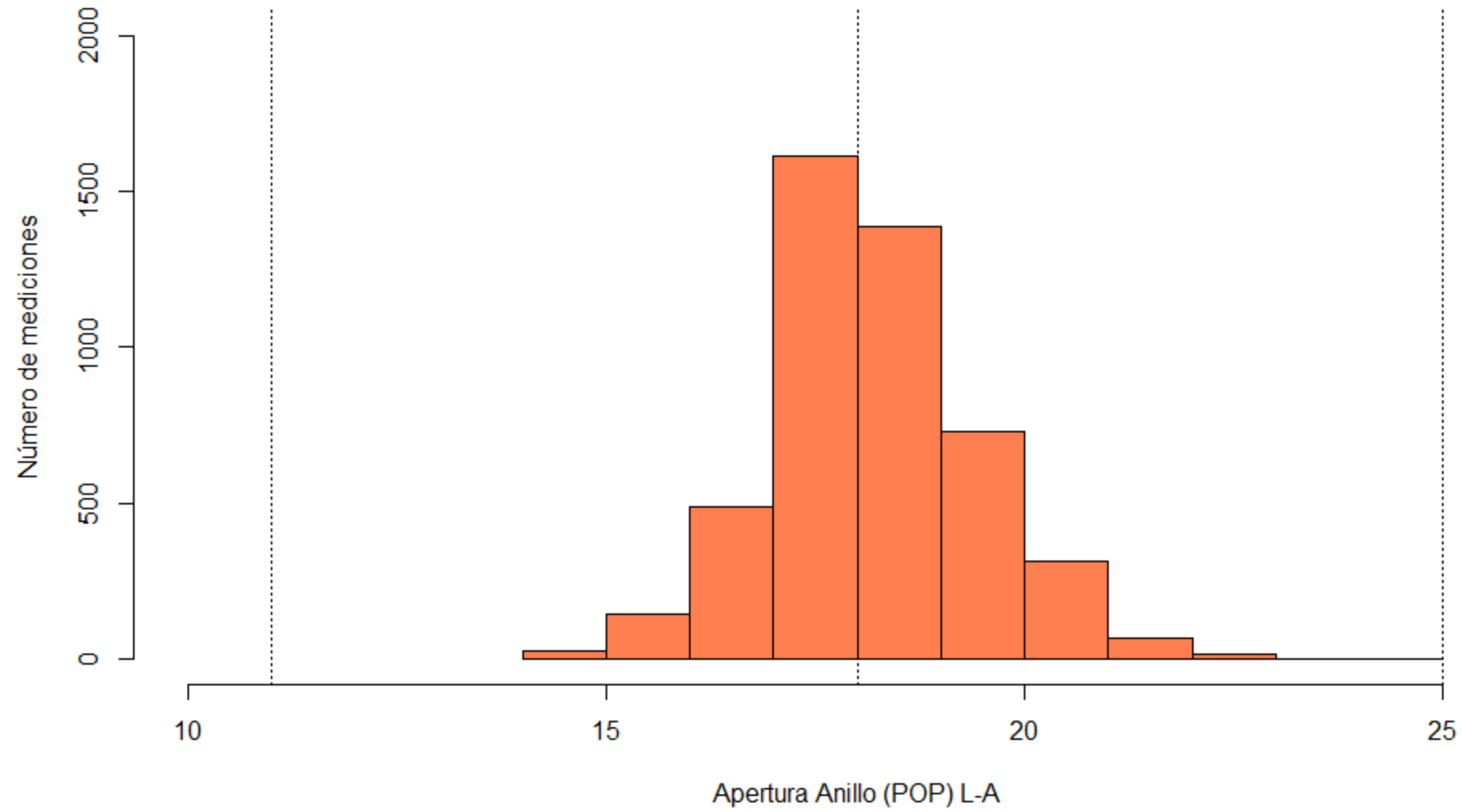
## Usos:

1. Visualizar la tendencia central de mediciones.
2. Estudiar el centrado del proceso.
3. Examinar la variabilidad del proceso
4. Analizar la forma del histograma.
5. Identificar mediciones raras.
6. Clasificar y analizar mediciones de acuerdo con distintas fuentes de procedencia.

## En R

```
datos2_sub1 <- subset(datos2, variable=="Apertura Anillo (POP) L-A" )  
hist(datos2_sub1$medicacion, col="coral",  
      main="Histograma de Apertura Anillo (POP) L-A",  
      xlim=c(10,26), ylim=c(0,2000), xlab="Apertura Anillo (POP) L-A",  
      ylab="Número de mediciones")  
abline(v=c(11, 18, 25), lty=3)
```

Histograma de Apertura Anillo (POP) L-A

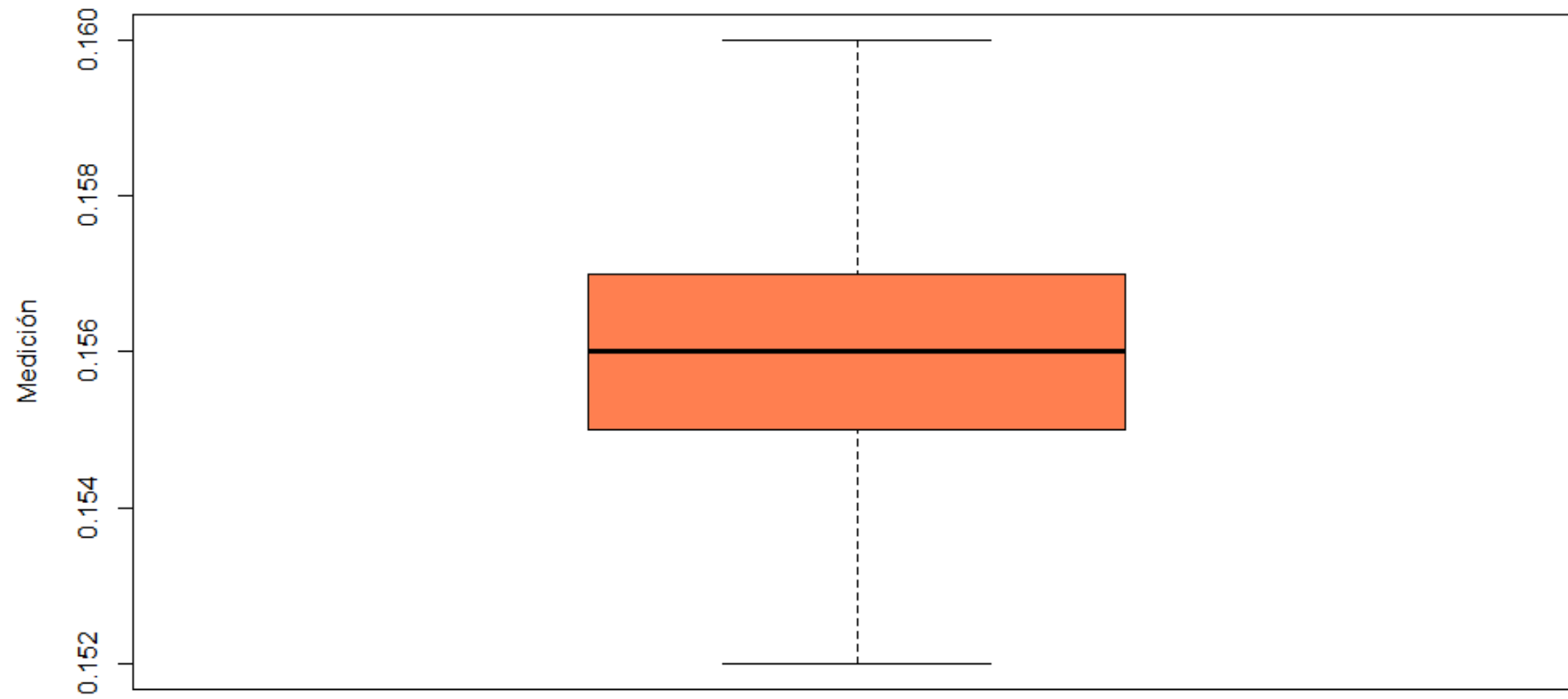


## Conclusión

A partir del gráfico se observa que el 33.67% de las mediciones de apertura del anillo (POP) lado A se encuentran entre 17 y 18. Las mediciones se encuentran centradas en el intervalo de especificación, entre 11 y 25, con variación moderada y no se observan mediciones raras.



# Boxplot



# Boxplot

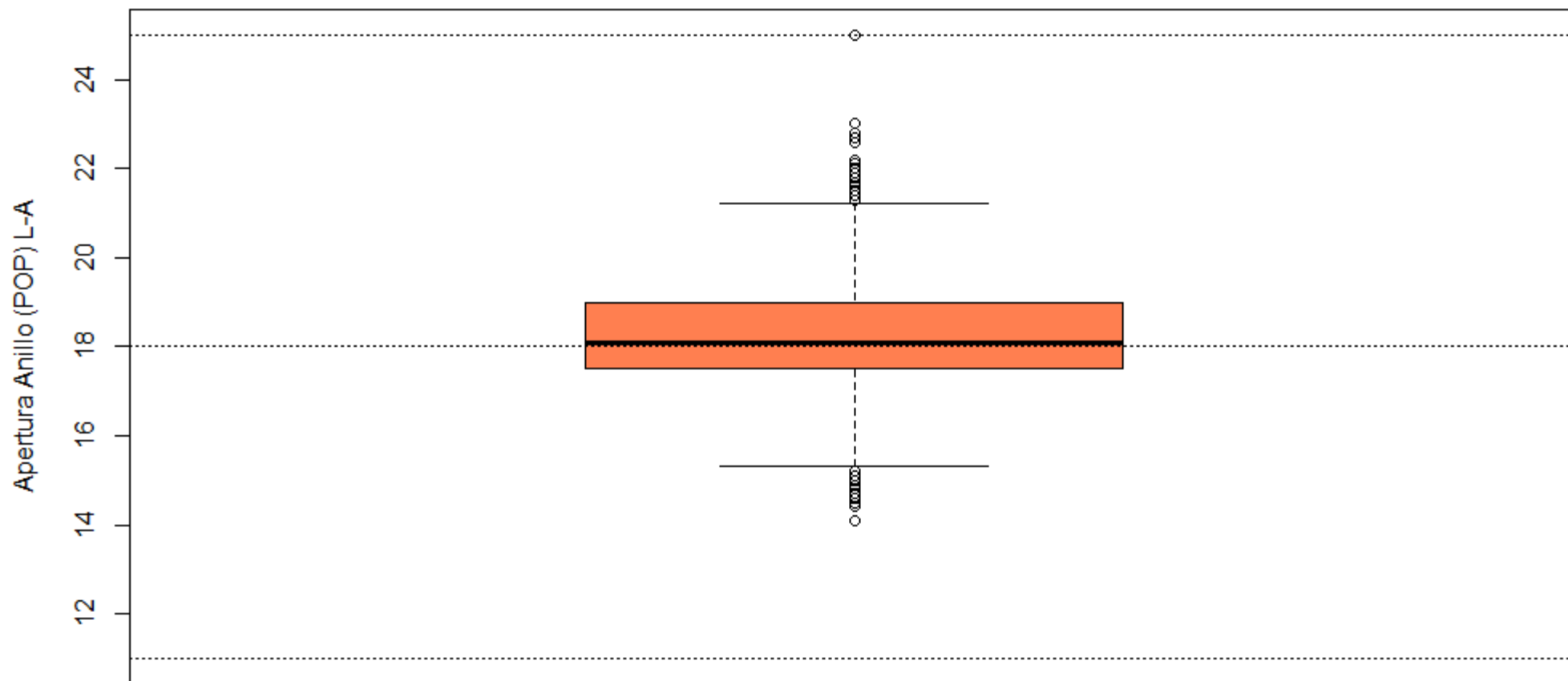
## Usos

1. Describir el comportamiento de un proceso.
2. Examinar la variabilidad del proceso.
3. Identificar mediciones raras.

## En R

```
boxplot(datos2_sub1$medicacion, col="coral",  
        main="Boxplot de Apertura Anillo (POP) L-A",  
        ylim=c(11,25), ylab="Apertura Anillo (POP) L-A")  
abline(h=c(11, 18, 25), lty=3)  
  
summary(datos2_sub1$medicacion)
```

**Boxplot de Apertura Anillo (POP) L-A**

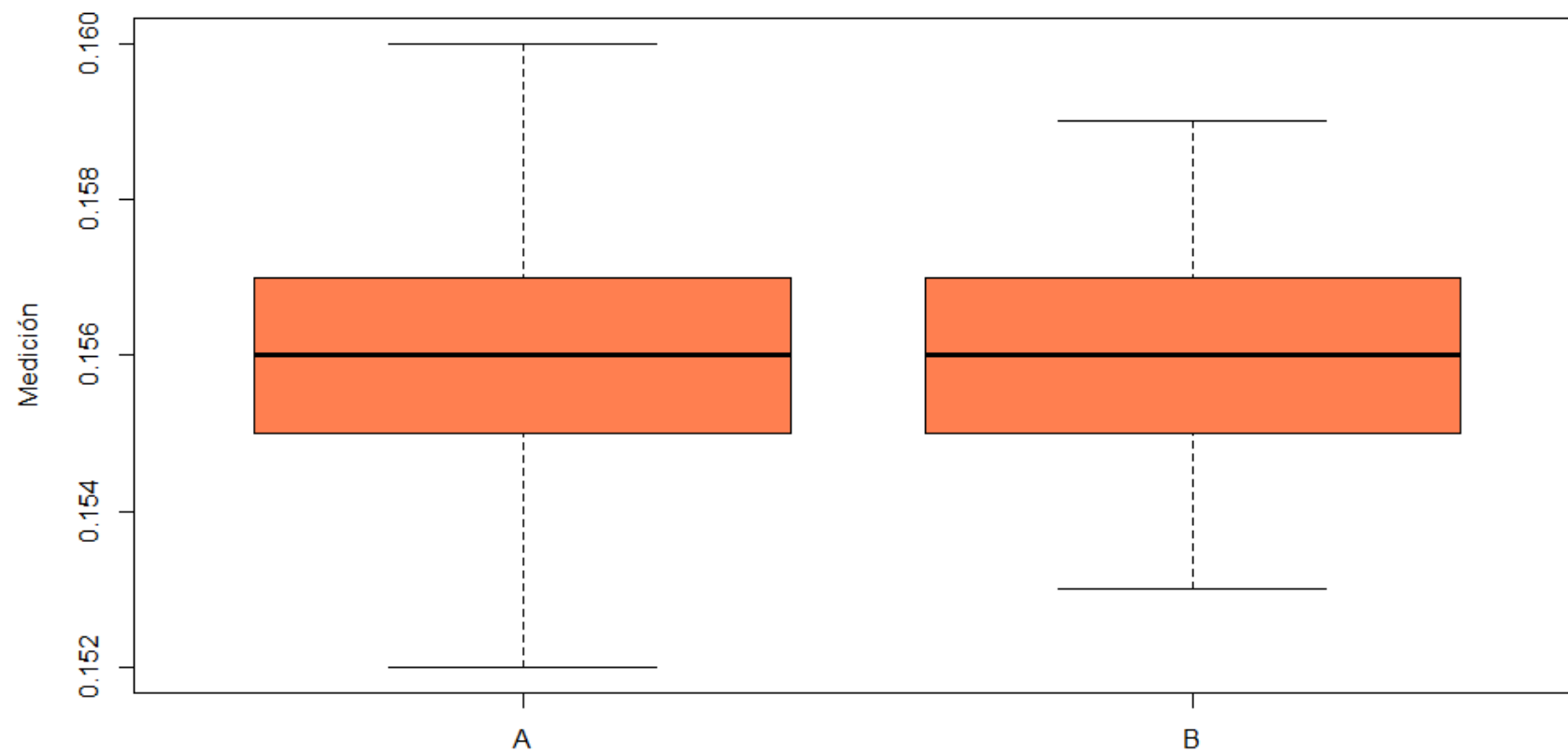


Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.1	17.5	18.1	18.3	19.0	25.0

## Conclusión

A partir del grafico y de las medidas numéricas se observa que el 25% de las mediciones se encuentran por debajo del 17.5, el 50% de las mediciones por debajo de 18.1 y el 75% por debajo de 19. Además, se observa que la menor y mayor medición tomaron el valor de 14.1 y 25.

## Boxplot comparativo



# Boxplot comparativo

## Usos

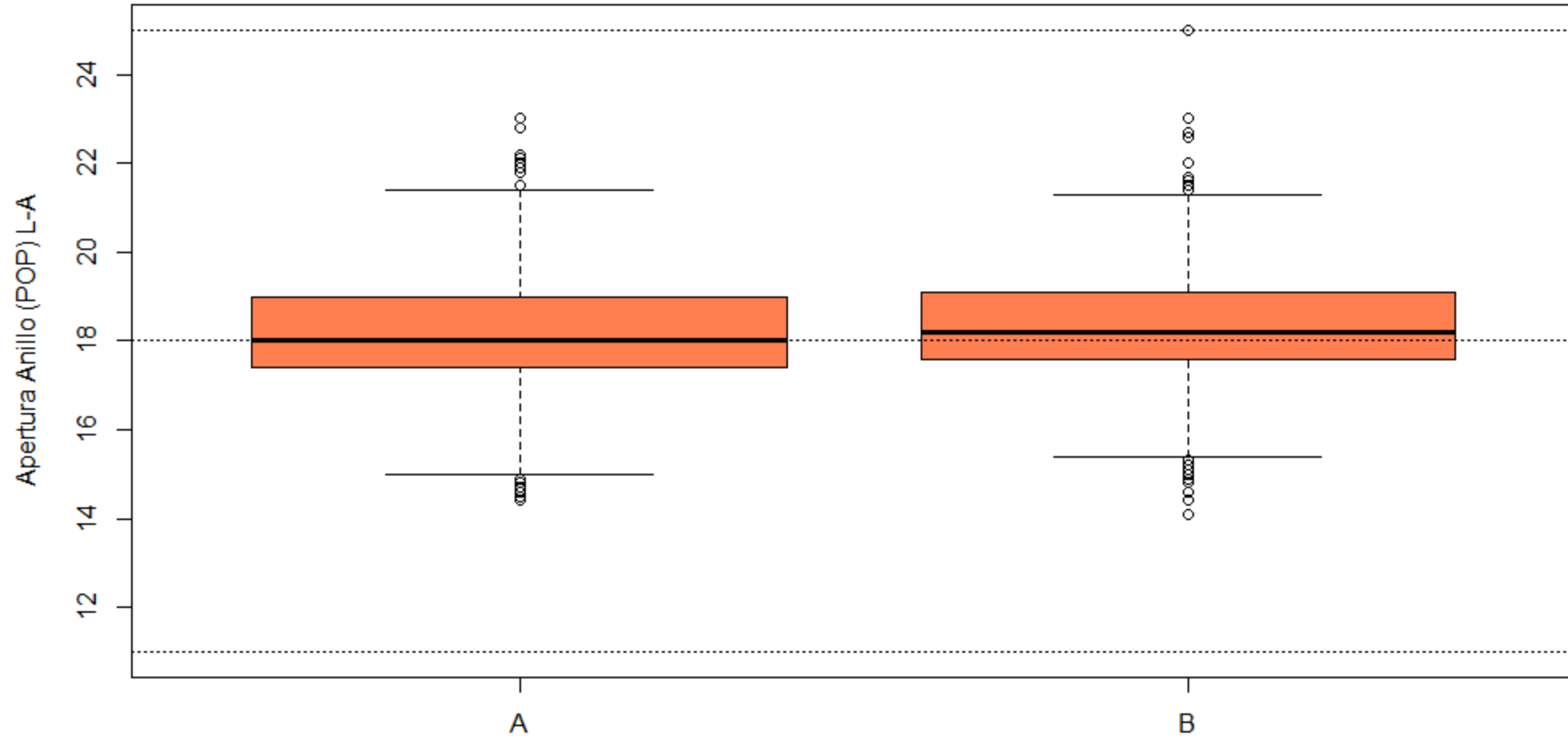
1. Describir el comportamiento de un proceso de acuerdo a las fuentes de procedencia.
2. Examinar la variabilidad del proceso de acuerdo a las fuentes de procedencia.
3. Identificar mediciones raras de acuerdo a las fuentes de procedencia.
4. Comparar procesos para hacer análisis por lotes, proveedores, turnos, máquinas, etc.

## En R

```
boxplot(datos2_sub1$medicacion ~ datos2_sub1$turno, col="coral",  
        main="Boxplot de Apertura Anillo (POP) L-A vs Turno",  
        ylim=c(11,25), ylab="Apertura Anillo (POP) L-A")  
abline(h=c(11, 18, 25), lty=3)
```



Boxplot de Apertura Anillo (POP) L-A



## Conclusión

A partir del grafico se observa que la medición no presenta diferencias significativas entre los dos turnos en los que se tomaron las mediciones. Además, el valor mas alto de la medición fue registrada en el turno B.

# Gráfico de dispersión

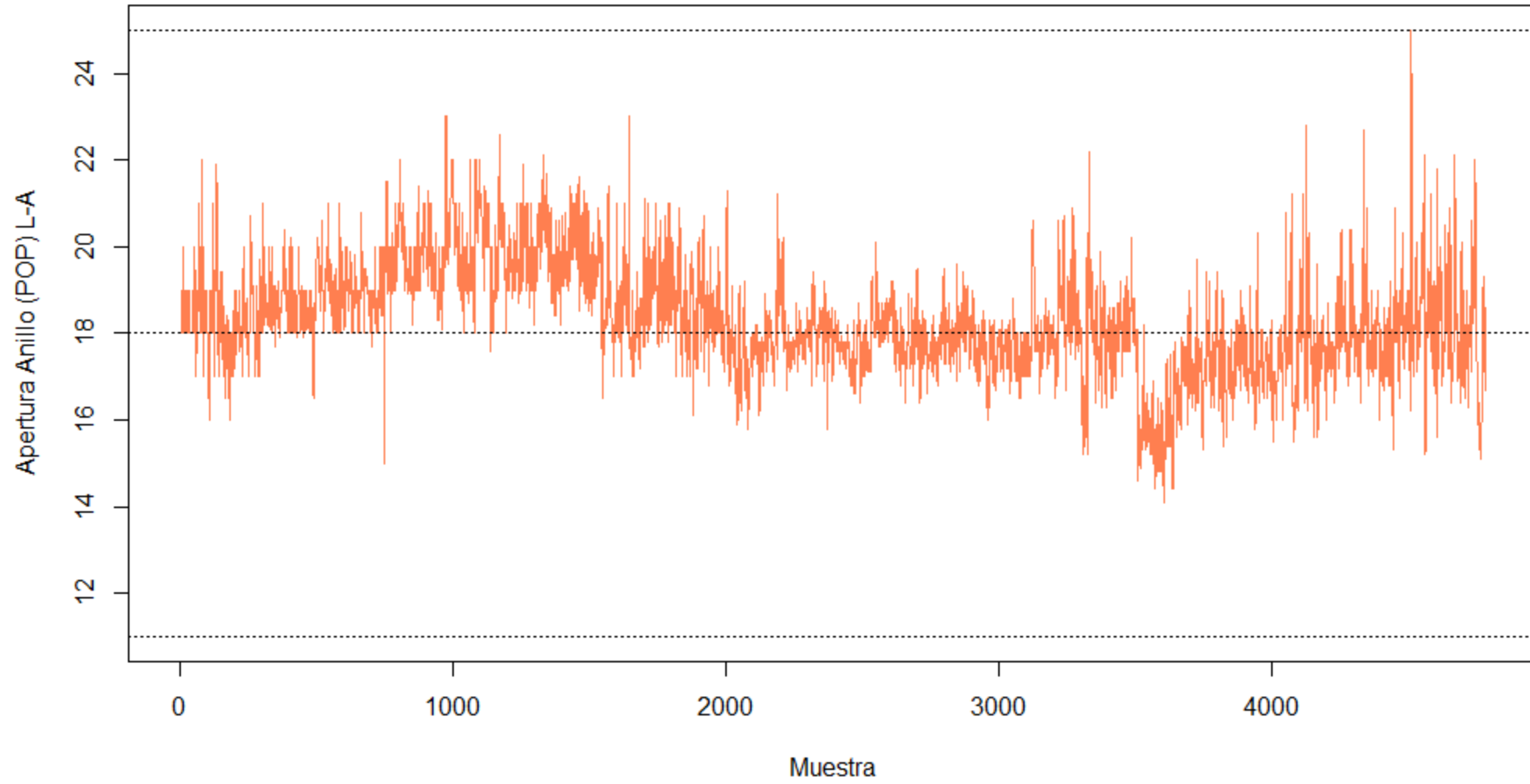
## Usos

1. Identificar relaciones entre dos mediciones: directas o inversas.
2. Identificar la fuerza de la relación entre dos mediciones: Débil, Moderada o Fuerte.
3. Identificar patrones de las relaciones entre las mediciones.

## En R

```
plot(datos2_sub1$medicacion, type="l", main="Muestra vs Apertura Anillo (POP)  
L-A", xlab="Muestra", ylab="Apertura Anillo (POP) L-A",  
      ylim=c(11, 25), col="coral")  
abline(h=c(11, 18, 25), lty=3)
```

**Muestra vs Apertura Anillo (POP) L-A**



## Conclusión

A partir del grafico se observa que la medición muestra variabilidad por encima, sobre y por debajo del valor central de la medición.

# Medidas de resumen por mediciones

## Usos:

Calcular de forma numérica las siguientes características de las mediciones:

1. El punto central: Promedio
2. La variabilidad: Desviación estándar y Coeficiente de variación

## En R

```
tapply(X=datos2$medicion, INDEX=datos2$variable, FUN=mean)
tapply(X=datos2$medicion, INDEX=datos2$variable, FUN=sd)
tapply(X=datos2$medicion, INDEX=datos2$variable,
      FUN=function(x) { c(mean(x) , sd(x) , sd(x)/mean(x)) })

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



## En R

```
$`Apertura Anillo (POP) L-A`  
[1] 18.301077  1.266785  6.921914
```

```
$`Apertura Anillo (POP) L-B`  
[1] 19.176797  1.687472  8.799549
```

```
$`Apertura Anillo (POP) L-C`  
[1] 15.807860  1.118791  7.077432
```

```
$`Apertura Anillo (POP) L-D`  
[1] 17.371409  1.783157 10.264894
```

Variable	Media	Desviación estándar	CV(%)	Especificación
Apertura del anillo (POP) L-A	18.30108	1.266785	6,921914	18.000000 ± (11.000000,25.000000)
Apertura del anillo (POP) L-B	19.17679	1.687471	8.799549	18.000000 ± (11.000000,25.000000)
Apertura del anillo (POP) L-C	15.80786	1.118791	7,077432	18.000000 ± (11.000000,25.000000)
Apertura del anillo (POP) L-D	17.37141	1.783157	10.26489	18.000000 ± (11.000000,25.000000)
Apertura TOTAL (TEAR) L-A	58.90427	3.690048	6,264484	48.000000 ± (33.000000,63.000000)
Apertura TOTAL (TEAR) L-B	58.59055	3.615772	6,171255	48.000000 ± (33.000000,63.000000)
Apertura TOTAL (TEAR) L-C	50.10517	5.293844	10.56546	48.000000 ± (33.000000,63.000000)
Apertura TOTAL (TEAR) L-D	55.40676	4.867363	8.784782	48.000000 ± (33.000000,63.000000)
Explosión L-A	35.70113	2.124610	5,951099	40.000000 ± (32.000000,48.000000)
Explosión L-B	35.67336	2.214299	6,207151	40.000000 ± (32.000000,48.000000)
Explosión L-C	35.67391	5.167307	14,48483	40.000000 ± (32.000000,48.000000)
Explosión L-D	36.41169	2.901078	7,967433	40.000000 ± (32.000000,48.000000)

Variable	Media	Desviación estándar	CV(%)	Especificación
Peso Barniz Post Repair, Especificación:14-18	15.60432	2.471926	15,84129	16.000000 ± (14.000000,18.000000)
Peso Barniz Post Repair, Especificación: 20-24	22.35741	1.890292	8,454876	22.000000 ± (20.000000,24.000000)
Peso Barniz Post Repair, Especificación: 28-36	29.70625	1.703112	5,733177	32.000000 ± (28.000000,36.000000)
Peso Barniz Post Repair, Especificación: 37-45	43.31038	2.360978	5,451299	41.000000 ± (37.000000,45.000000)
Peso Barniz Post Repair, Especificación: 40-48	42.40263	2.362978	5,572716	44.000000 ± (40.000000,48.000000)
Peso Barniz Post Repair L-A, (Ext)	23.89120	5.812197	24,32777	21.000000 ± (18.000000,24.000000)
Peso Barniz Post Repair L-B, (Ext)	23.97222	6.291022	26,24296	21.000000 ± (18.000000,24.000000)
Profundidad de la incisión L-A	0.1097031	0.01089701	9,933181	0.070000 ± (0.070000,0.130000)
Profundidad de la incisión L-B	0.1054709	0.01114396	10,56592	0.070000 ± (0.070000,0.130000)
Profundidad de la incisión L-C	0.1053207	0.01424448	13,52486	0.070000 ± (0.070000,0.130000)
Profundidad de la incisión L-D	0.0958758	0.00906008	9,449807	0.070000 ± (0.070000,0.130000)

Taller

<https://tinyurl.com/y8gwqmu5>

## Referencias

Hernández, F. y Usuga, O. Manual de R. 2018. [https://bookdown.org/fhernanb/Manual\\_de\\_R](https://bookdown.org/fhernanb/Manual_de_R)

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. R News 4/1, 11-17. <https://cran.r-project.org/doc/Rnews>.