## COMPARACIÓN DE LOTES MEDIANTE BOXPLOTS

### EJEMPLO: LAS CIUDADES MÁS POBLADAS EN 16 PAÍSES

El *World Almanac* de 1967 lista 16 países que tienen 10 o más ciudades grandes; entre estas se han elegido las 10 más pobladas.

- Cómo se comparan las medianas a través de las naciones?
- Son las menores ciudades más grandes de China mayores que las más grandes ciudades de algunos de los otros países?
- Tienen los países con ciudades más grandes una tendencia a tener mayor variabilidad entre las poblaciones?
- Cuánta asimetría tienen los diferentes conjuntos de datos?

| (1) Sweden | | (2) Netherlands | | (3) Canada | | (4) France | |
|---|---|---|---|---|---|---|---|
| Stockholm | 7.87 | Amsterdam | 8.68 | Montreal | 11.91 | Paris | 28.11 |
| Goteborg | 4.22 | Rotterdam | 7.31 | Toronto | 6.72 | Marseilles | 7.83 |
| Malmo | 2.49 | The Hague | 6.02 | Vancouver | 3.84 | Lyon | 5.35 |
| Norrkoping | 0.94 | Utrecht | 2.64 | Edmonton | 2.81 | Toulouse | 3.30 |
| Vasteras | 0.89 | Eindhoven | 1.75 | Hamilton | 2.73 | Nice | 2.94 |
| Uppsala | 0.87 | Haarlem | 1.72 | Ottawa | 2.68 | Bordeaux | 2.54 |
| Orebro | 0.81 | Groningen | 1.51 | Winnipeg | 2.65 | Nantes | 2.46 |
| Halsingborg | 0.78 | Tilburg | 1.42 | Calgary | 2.49 | Strasbourg | 2.33 |
| Linkoping | 0.71 | Enschede | 1.31 | Quebec | 1.71 | St. Etienne | 2.03 |
| Boras | 0.69 | Arnhem | 1.29 | London | 1.69 | Lille | 1.99 |



| (5) Mexico | | (6) Argentina | | (7) Spain | | (8) England | |
|---|---|---|---|---|---|---|---|
| MexicoCity | 31.18 | BuenosAires | 29.66 | Madrid | 25.99 | London | 79.86 |
| Guadalajara | 10.12 | Rosario | 7.61 | Barcelona | 16.96 | Birmingham | 11.02 |
| Monterrey | 8.06 | Cordoba | 6.35 | Valencia | 5.01 | Liverpool | 7.22 |
| Juarez | 3.79 | La Plata | 4.10 | Seville | 4.74 | Manchester | 6.38 |
| Puebla | 3.46 | Avellaneda | 3.80 | Zaragoza | 3.57 | Leeds | 5.09 |
| Mexicali | 2.91 | Santa Fe | 2.75 | Bilboa | 3.34 | Sheffield | 4.88 |
| Leon | 2.71 | Mar del Plata | 2.70 | Malaga | 3.12 | Bristol | 4.30 |
| Torreon | 2.17 | Gral.SanMartin | 2.69 | Murcia | 2.64 | Coventry | 3.30 |
| Chihuahua | 2.06 | Tucuman | 2.51 | Cordoba | 2.14 | Nottingham | 3.10 |
| SanLuisPotosi | 1.86 | Lanus | 2.44 | Palma | 1.69 | Kingston | 2.99 |

| (9) Italy | | (10) West Germany | | (11) Brazil | | (12) Soviet Union | |
|---|---|---|---|---|---|---|---|
| Rome | 23.59 | WestBerlin | 21.92 | Sao Paulo | 49.81 | Moscow | 63.34 |
| Milan | 15.80 | Hamburg | 18.56 | RiodeJaneiro | 38.57 | Leningrad | 36.36 |
| Naples | 11.82 | Munich | 11.42 | Recife | 9.68 | Kiev | 13.32 |
| Turin | 11.14 | Cologne | 8.27 | BeloHorizonte | 9.52 | Baku | 11.37 |
| Genoa | 7.84 | Essen | 7.28 | Salvador | 8.08 | Tashkent | 10.90 |
| Palermo | 5.90 | Dusseldorf | 7.02 | Porto Alegre | 8.03 | Gorky | 10.84 |
| Florence | 4.54 | Frankfurt | 6.94 | Fortaleza | 6.99 | Kharkov | 10.70 |
| Bologna | 4.44 | Dortmund | 6.53 | Curitiba | 5.02 | Novosibirsk | 10.27 |
| Catania | 3.61 | Bremen | 5.84 | Belem | 4.95 | Kuibyshev | 9.50 |
| Venice | 3.36 | Hannover | 5.66 | Niterol | 2.78 | Sverdlovsk | 9.17 |

| (13) Japan | | (14) United States | | (15) India | | (16) China | |
|---|---|---|---|---|---|---|---|
| Tokyo | 110.21 | NewYork | 77.81 | Bombay | 45.37 | Shangai | 69.00 |
| Osaka | 32.14 | Chicago | 35.50 | Calcutta | 30.03 | Beijing | 40.10 |
| Nagoya | 18.88 | LosAngeles | 24.79 | Delhi | 22.98 | Hong Kong | 36.92 |
| Yokohama | 16.39 | Philadelphia | 20.02 | Hyderabad | 20.62 | Tianjin | 32.20 |
| Kyoto | 13.37 | Detroit | 16.70 | Madras | 17.25 | Shenyang | 24.11 |
| Kobe | 11.95 | Baltimore | 9.39 | Howrah | 16.11 | Wuhan | 21.46 |
| KitaKyush | 10.70 | Houston | 9.38 | Ahmedabad | 11.49 | Chongqing | 21.21 |
| Kawasaki | 7.89 | Cleveland | 8.76 | Kanpur | 9.47 | Canton | 16.50 |
| Fukuoka | 7.71 | Washington,DC | 7.63 | Bangalore | 9.07 | Xian | 15.00 |
| Sapporo | 7.04 | St. Louis | 7.50 | Poona | 7.21 | Nanjing | 11.13 |

**Los datos se encuentran en el archivo POBL16.txt**

```
> pobl16 <- read.table(file.choose(),header
=T)
> summary(pobl16)
```

Las respuestas a las preguntas planteadas pueden obtenerse a partir las siguientes medidas resumen

**Tabla 12: medidas resumen de las poblaciones de 10 ciudades mayores de 16 países**

| SWEDEN | NETHERLANDS | CANADA | FRANCE |
|---|---|---|---|
| Min. :0.6900 | Min. :1.290 | Min. : 1.690 | Min. : 1.990 |
| 1st Qu.:0.7875 | 1st Qu.:1.442 | 1st Qu.: 2.530 | 1st Qu.: 2.362 |
| Median :0.8800 | Median :1.735 | Median : 2.705 | Median : 2.740 |
| Mean :2.0270 | Mean :3.365 | Mean : 3.923 | Mean : 5.888 |
| 3rd Qu.:2.1020 | 3rd Qu.:5.175 | 3rd Qu.: 3.582 | 3rd Qu.: 4.837 |
| Max. :7.8700 | Max. :8.680 | Max. :11.910 | Max. :28.110 |
| MEXICO | ARGENTINA | SPAIN | ENGLAND |
| Min. : 1.860 | Min. : 2.440 | Min. : 1.690 | Min. : 2.990 |
| 1st Qu.: 2.305 | 1st Qu.: 2.692 | 1st Qu.: 2.760 | 1st Qu.: 3.550 |
| Median : 3.185 | Median : 3.275 | Median : 3.455 | Median : 4.985 |
| Mean : 6.832 | Mean : 6.461 | Mean : 6.920 | Mean :12.810 |
| 3rd Qu.: 6.992 | 3rd Qu.: 5.788 | 3rd Qu.: 4.942 | 3rd Qu.: 7.010 |
| Max. :31.180 | Max. :29.660 | Max. :25.990 | Max. :79.860 |
| ITALY | WGERMANY | BRAZIL | URSS |
| Min. : 3.360 1st | Min. :5.660 | Min. : 2.780 | Min. : 9.17 |
| Qu.: 4.465 | 1st Qu.:6.632 | 1st Qu.: 5.512 | 1st Qu.:10.38 |
| Median : 6.870 | Median :7.150 | Median : 8.055 | Median :10.87 |
| Mean : 9.204 | Mean :9.944 | Mean :14.340 | Mean :18.58 |
| 3rd Qu.:11.650 | 3rd Qu.:10.630 | 3rd Qu.: 9.640 | 3rd Qu.:12.83 |
| Max. :23.590 | Max. :21.920 | Max. :49.810 | Max. :63.34 |
| JAPAN | USA | INDIA | CHINA |
| Min. : 7.040 | Min. : 7.500 | Min. : 7.210 | Min. :11.13 |
| 1st Qu.: 8.592 | 1st Qu.: 8.915 | 1st Qu.: 9.975 | 1st Qu.:17.68 |
| Median : 12.660 | Median :13.040 | Median :16.680 | Median :22.78 |
| Mean : 23.630 | Mean :21.750 | Mean :18.960 | Mean :28.76 |
| 3rd Qu.: 18.260 | 3rd Qu.:23.600 | 3rd Qu.:22.390 | 3rd Qu.:35.74 |
| Max. :110.200 | Max. :77.810 | Max. :45.370 | Max. :69.00 |

En la práctica, un gráfico con los boxplots en paralelo para los 16 grupos de datos hace que las respuestas a estas y otras preguntas similares aparezcan rápidamente .



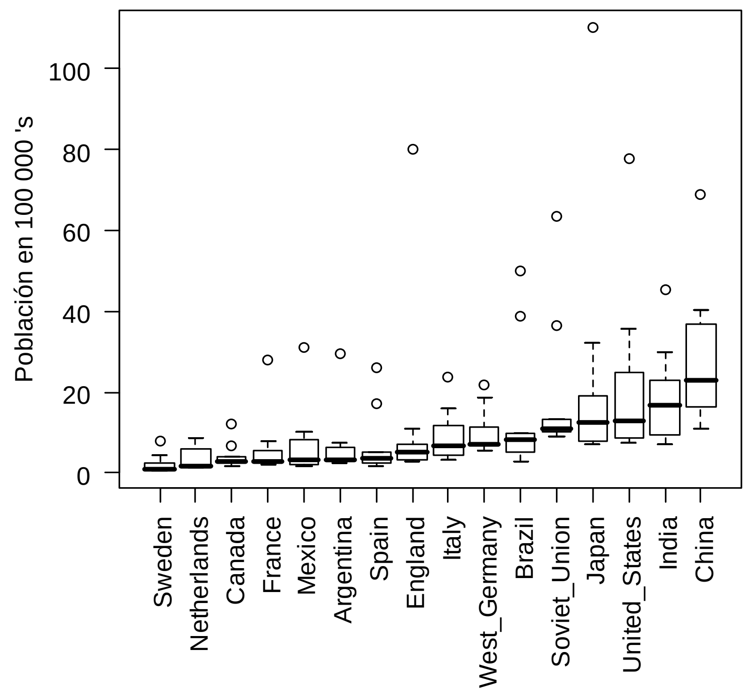



**FIGURA 12.** Boxplots ordenados de acuerdo a la mediana de las poblaciones de las 10 ciudades más grandes.

- Las ciudades más grandes de China tienden a ser más grandes que las de cualquier otra nación. todas las ciudades grandes de China son más grandes que todas las ciudades grandes de Suecia (Sweden) y Holanda (Netherlands).
- Comparamos la dispersión de estos 16 lotes mediante las longitudes de las cajas. Los datos de Canadá son los menos dispersos y los de China los más.

- La mayoría de los países presentan alguna asimetría en dirección a las grandes ciudades;

- Solamente India y Brasil tienen cajas que están sesgadas hacia la izquierda, pero ambos países tienen ciudades sustancialmente mayores que las representadas por las cajas.

- La mayor ciudad de todos los países, con la excepción de Holanda, es designada como un outlier; algunos tienen más de un outlier entre las 10 ciudades más pobladas.

Hemos detectado dos anormalidades en estos datos: **asimetría y outliers**.

Al haber ordenado los países en base a la mediana de los lotes podemos detectar otra característica:

**tendencia de aumento en la dispersión a medida que aumenta el nivel.**

Esta tendencia no es compatible con el supuesto de similar variabilidad entre lotes; cuando esto ocurre el análisis se simplifica.

Veremos transformaciones de los datos que permitan lograr homogeneidad de dispersiones y reducir la dependencia de éstas con el nivel.

**En R**

La función **`mar`** da los *márgenes* de los gráficos en *pulgadas* en el siguiente orden: margen inferior, margen izquierdo, margen superior y margen derecho respectivamente.
El valor por defecto es mar=c(5,4,4,2)+0.1.



`mar`

A numerical vector of the form `c(bottom, left, top, right)` which gives the number of lines of margin to be specified on the four sides of the plot.

Hemos obtenido el boxplot de las poblaciones de las 10 ciudades más pobladas mediante las siguientes instrucciones.

```
> par(las=2,cex=1,mar=c(7.2,4,2,2))
> boxplot(pobl16,
   ylab= "Población en 100 000 's" )
```

Para <u>graficar los boxplots</u> en <u>orden creciente</u> de las medianas

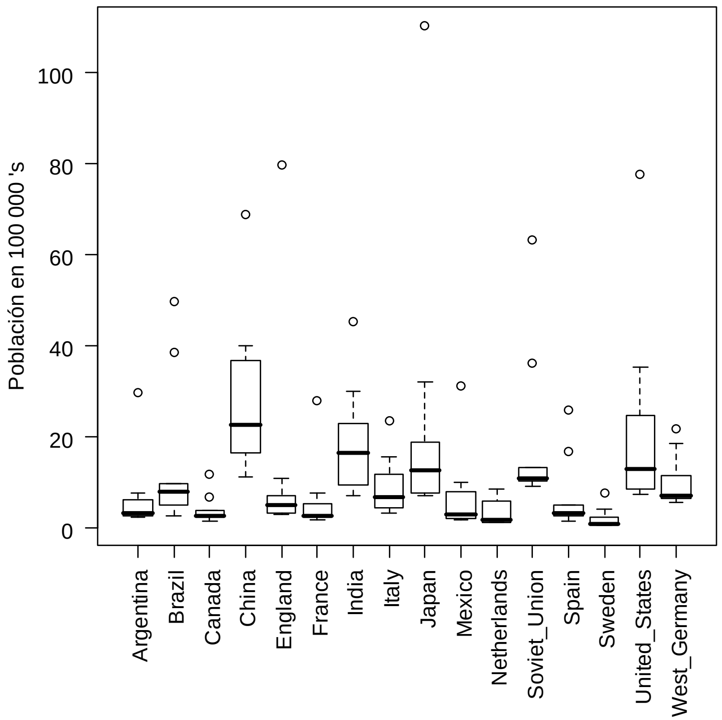
```
> orden.med <-
sort.list(sapply(pobl16,median))
> boxplot(pobl16[orden.med])
```

En este caso **`sapply(pobl16,median)`**

es lo mismo que **`apply(pobl16,2, median)`**

<u>Para graficar en orden alfabético</u>
Vemos que el data frame no tiene a los países con sus nombres en orden alfabético

```
> names(pobl16)
 [1] "Sweden"        "Netherlands"
 [3] "Canada"        "France"
 [5] "Mexico"        "Argentina"
 [7] "Spain"         "England"
 [9] "Italy"         "West_Germany"
[11] "Brazil"        "Soviet_Union"
[13] "Japan"         "United_States"
[15] "India"         "China"
```




Primero hallamos **los índices** de los nombres ordenados de acuerdo con el orden alfabético

```
> sort.list(names(pobl16))
 [1]  6 11  3 16  8  4 15  9 13  5  2 12  7
1
[15] 14 10
```

```
> orden.alfabetico <-
                sort.list(names(pobl16))
```



```
> names(pobl16[orden.alfabetico])
 [1] "Argentina"     "Brazil"        "Canada"
 [4] "China"         "England"       "France"
 [7] "India"         "Italy"         "Japan"
[10] "Mexico"    "Netherlands"    "Soviet_Union"
[13] "Spain"      "Sweden"        "United_States"
[16] "West_Germany"
```

```
> x11()
> par(las=2,cex=1,mar=c(7.2,4,2,2))
> boxplot(pobl16[orden.alfabetico ],
      ylab= "Población en 100 000 's" )
```

## NIVEL VERSUS DISPERSIÓN

Nos interesa hallar una ***transformación*** de los datos que reduzca o elimine el crecimiento, o el decrecimiento, de la dispersión con el nivel.
Los datos re-expresados serán más adecuados tanto para exploración visual como para aplicar técnicas usuales de comparación de grupos.

Por ejemplo el análisis de varianza de un factor es más simple y más efectivo cuando hay, exacta o aproximadamente, igualdad de varianzas entre grupos.

# Transformaciones de potencia

Definimos la transformación de potencia con potencia (ó exponente) $p$ como la transformación que reemplaza $x$ por $x^p$.

Para $p$=0 utilizamos log $x$ en vez de $x^0$.

Veremos que log $x$ es el límite cuando $p$ tiende a cero de
$(x^p-1)/p$.

Definiremos un gráfico que nos permitirá encontrar la transformación adecuada.

## Construcción de gráficos de dispersión versus nivel

Nos interesa eliminar la relación entre el nivel y la dispersión de un conjunto de lotes (es decir varios conjuntos de datos, correspondientes a observaciones de una misma variable en diferentes poblaciones).

**Proposición:** Supongamos que la distancia intercuartos de cada conjunto de datos es proporcional a una potencia de la mediana:

$$d_Q = cM^b, \quad (1)$$

ó equivalentemente

$$\log d_Q = k + b \log M. \quad (2)$$

y que existe p tal que para los datos transformados ($X^p$ o log (X) si p=0)



$$\begin{aligned} \text{mediana} &= m & (m > 0) & \\ \text{cuarto superior} &= m + d & (d > 0) & \quad (3) \\ \text{cuarto inferior} &= m - c & (c > 0) & \\ \text{distancia intercuartos} &= d + c & (\text{cte., indep. de } m) & \end{aligned}$$

o sea que la distancia intercuartos de los datos transformados no depende de $m$. Luego (3) se satisface aproximadamente cuando b = 1 - p.

**Corolario:** La potencia p puede estimarse mediante la pendiente del gráfico de dispersión de los valores de log $d_Q$ contra los valores de log $M$.

Denominamos al gráfico de log $d_Q$ *vs* log $M$:
**gráfico de dispersión nivel**

El gráfico de dispersión versus nivel consiste en graficar los valores de log $d_Q$ contra los valores de log $M$ para todos los lotes y luego ajustar una recta al diagrama de dispersión obtenido.

Veremos que si $b$ es la pendiente estimada entonces

$$\boldsymbol{p = 1 - b}$$

es un valor aproximado del exponente de una transformación de potencias de $x$ para estabilizar la dispersión. Cuando $p$ = 0 se utiliza el logaritmo.

**Tabla 15: Logs de medianas - 5 (base 10) y distancias intercuartos para las mayores ciudades de 16 países.**

| País | log $M$ | log $d_Q$ |
|---|---|---|
| Sweden | -.06 | .23 |
| Netherlands | .24 | .66 |
| Canada | .43 | .13 |
| France | .44 | .48 |
| Mexico | .50 | .77 |
| Argentina | .51 | .56 |
| Spain | .54 | .38 |
| England | .70 | .59 |
| Italy | .84 | .87 |
| West Germany | .85 | .69 |
| Brazil | .91 | .67 |
| Soviet Union | 1.04 | .48 |
| Japan | 1.10 | 1.04 |
| United States | 1.12 | 1.20 |
| India | 1.22 | 1.13 |
| China | 1.36 | 1.31 |

**En R**

```
>  spobl16  <-  sapply(pobl16,sort)#  ordeno
cada lote
>  (trunc((10+1)/2)+  1)/2  #  Profundidad  de
los cuartos
[1] 3
> Qinf <-  spobl16[3,]
>  Qsup <-  spobl16[8,]
>  med <- (spobl16[5,]+spobl16[6,])/2
>   dQ <-Qsup-Qinf
>  logdQ <- log10(Qsup-Qinf)
>  logm <- log10(med)
>  cbind(logm,logdQ)
                   logm     logdQ
Sweden        -0.05551733 0.2329961
Netherlands    0.23929948 0.6627578
Canada         0.43216727 0.1303338
France         0.43775056 0.4800069
Mexico         0.50310944 0.7701153
```

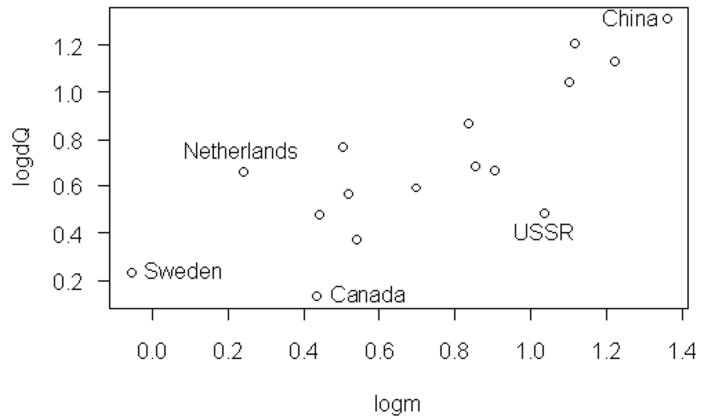


```
Argentina      0.51521130 0.5634811
Spain          0.53844805 0.3747483
England        0.69766516 0.5932861
Italy          0.83695674 0.8680564
West_Germany   0.85430604 0.6893089
Brazil         0.90606554 0.6683859
Soviet_Union   1.03622954 0.4842998
Japan          1.10243371 1.0409977
United_States  1.11544408 1.2049335
India          1.22219605 1.1306553
China          1.35764903 1.3100557
```

**Observación**: 1ro se calcula la distancia intercuartos y la mediana, 2do. se toma logaritmo

EJEMPLO: **GRÁFICO DE DISPERSIÓN VERSUS NIVEL** PARA LAS MAYORES CIUDADES

**Figura 15: Gráfico de dispersión versus nivel, log $d_Q$ contra log $M$**




**En R**

```
plot(logm,logdQ,las=1)
```

```
leg1<- c("Sweden","Netherlands",
"Canada", "USSR","China" )
text(locator(1),leg1[1])
text(locator(1),leg1[2])
text(locator(1),leg1[3])
text(locator(1),leg1[4])
text(locator(1),leg1[5])
```

Ajustamos una recta a ojo a los puntos de la figura 15. Aunque dos personas no llegarán a la misma pendiente por este método, casi seguro trazarán una recta con pendiente entre ½ y 1, y probablemente más cerca de 1 (la recta de regresión ajustada por cuadrados mínimos tiene pendiente 0.69).

Para *b* = 1, *p* es cero y resulta la transformación logaritmo. Análogamente, si *b* = ½ lleva a *p* = ½ , la transformación raíz cuadrada.
A pesar que una potencia entre 0 y ½ puede ser mejor que alguno de estos dos para estabilizar la dispersión, por razones de simplicidad y de interpretabilidad, consideraremos las transformaciones logaritmo y raíz cuadrada.

Aplicamos cada una de las transformaciones.



**Tabla 16. Medianas y distancias intercuartos para los datos de las 10 ciudades mayores de 16 países transformados por logaritmo base 10 y raíz cuadrada.**

| País | Logaritmo de la población | | Raíz cuadrada de la población | |
|---|---|---|---|---|
| | Mediana (más 5) | $d_Q$ | Mediana (por $10^2$) | $d_Q$ |
| Sweden | -.06 | .51 | 2.97 | 2.19 |
| Netherlands | .24 | .62 | 4.17 | 3.98 |
| Canada | .43 | .18 | 5.21 | 1.20 |
| France | .44 | .36 | 5.24 | 2.49 |
| Mexico | .50 | .57 | 5.64 | 4.32 |
| Argentina | .51 | .37 | 5.72 | 2.78 |
| Spain | .54 | .28 | 5.88 | 2.12 |
| England | .70 | .34 | 7.07 | 2.75 |
| Italy | .84 | .42 | 8.29 | 4.21 |
| West Germany | .85 | .24 | 8.46 | 2.61 |
| Brazil | .91 | .29 | 8.98 | 2.75 |
| Soviet Union | 1.04 | .11 | 10.43 | 1.41 |
| Japan | 1.10 | .39 | 11.25 | 4.86 |
| United States | 1.12 | .45 | 11.42 | 6.38 |
| India | 1.22 | .38 | 12.92 | 5.42 |
| China | 1.36 | .35 | 15.09 | 6.37 |

**En R**

```
> #los datos vienen en unidades de 100000
(10^5)
> logpobl <- log(spobl16,10)

> #Raiz VER MULTIPLICACIÓN POR 10
> raizpobl <- sqrt(spobl16*10)


> mediana.log <-  apply(logpobl,2,median)
> mediana.raiz <- apply(raizpobl,2,median)
```

```
> Qinf <-  logpobl[3,]
> Qsup <-  logpobl[8,]
> dQ.log <-Qsup-Qinf; dQ.log
       Sweden   Netherlands
    0.5041047     0.6273081
       Canada        France
    0.1881319     0.3609979
       Mexico     Argentina
    0.5698753     0.3730214
        Spain       England
    0.2782338     0.3400233
        Italy  West_Germany
    0.4252345     0.2427529
       Brazil  Soviet_Union
    0.2851716     0.1129338
        Japan United_States
    0.3789250     0.4517724
        India         China
    0.3850000     0.3497777

> Qinf <-  raizpobl[3,]
> Qsup <-  raizpobl[8,]

> dQ.raiz <-Qsup-Qinf; dQ.raiz
       Sweden   Netherlands
     2.197142      3.990577
       Canada        France
     1.206783      2.487362
       Mexico     Argentina
     4.319424      2.782168
        Spain       England
     1.940042      2.752496
        Italy  West_Germany
     4.208650      2.605598
       Brazil  Soviet_Union
     2.753503      1.407130
        Japan United_States
```

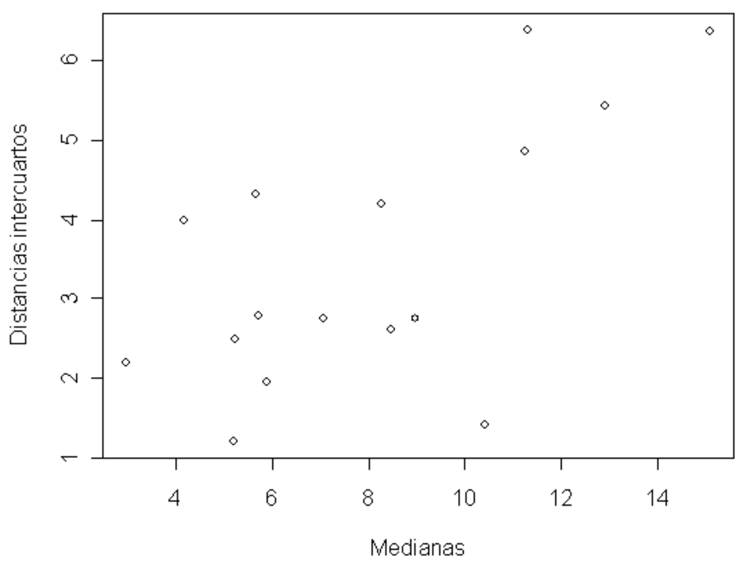


```
     4.857884      6.385353
        India         China
     5.427763      6.369345
```

**Observación:** Aquí la distancia intercuartos se calcula después de transformar los datos.

Los gráficos dados en las figuras 16 y 17 permiten evaluar cuál de las dos transformaciones puede ser mejor para estabilizar las dispersiones.

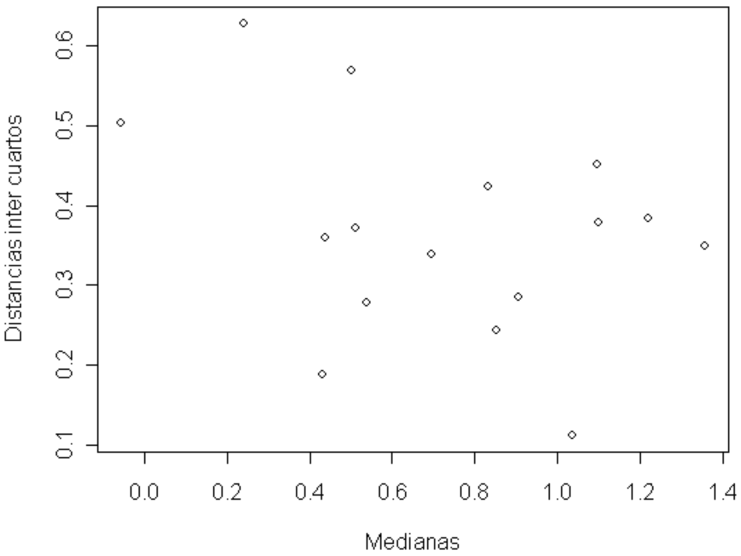
**Figura 16. Distancia intercuartos versus mediana: datos transformados por la raíz cuadrada.**




Las magnitudes relativas de las **pendientes** de los gráficos de las figuras 16 y 17 nos permiten elegir entre las dos transformaciones.

Las distancias intercuartos de los logaritmos decrecen levemente con el nivel mientras las de la raíz cuadrada aumentan con el nivel.

**Figura 17. Distancia intercuartos versus nivel: datos transformados por logaritmo.**




```
plot(mediana.log,dQ.log, xlab="Medianas",
     ylab="Distancias inter cuartos", main="
Transformación logaritmo")

plot(mediana.raiz,dQ.raiz, xlab="Medianas",
     ylab="Distancias  intercuartos",  main="
Transformación raíz cuadrada")
```

## Tomando la decisión



¿Cómo podemos tomar la decisión entre la transformación logaritmo, raíz cuadrada o alguna otra transformación de potencia con *p* entre 0 y 1?

Idealmente, una transformación no sólo iguala dispersiones sino también tiene una explicación temática.

Por ejemplo, en demografía, un modelo muy utilizado supone que las poblaciones tienden a crecer exponencialmente. Si esto es así, el logaritmo de la población crecerá aproximadamente de manera lineal.

Las ventajas del crecimiento lineal, tales como la facilidad de detectar apartamientos del ajuste y la conveniencia en la interpolación sugieren al logaritmo como una transformación adecuada para poblaciones humanas.

También la raíz cúbica, *p* = 1/3 es una transformación que algunas veces tiene significado físico.

**Tabla 17. Transformaciones de potencia más usadas**[a]

| Transformación | Potencia | Pendiente del gráfico dispersión - nivel |
|---|---|---|
| Cúbica | 3 | -2 |
| Cuadrada | 2 | -1 |
| No cambio | 1 | 0 |
| Raíz cuadrada | 1/2 | 1/2 |
| Logaritmo | 0 | 1 |
| Inversa de la | -1/2 | 3/2 |
| raíz cuadrada | -1 | 2 |
| Inversa | | |

[a]Corresponden a los miembros principales de la "escalera de potencias" de Tukey.

Para las ciudades más grandes elegimos la transformación logaritmo por que se trata de datos poblacionales. Este modelo teórico simple que favorece al logaritmo, más que una fuerte evidencia en los datos, es la base para tomar esa decisión.

**Reanálisis en la escala logarítmica.**

**Tabla 18: medidas resumen del logaritmo de las poblaciones de las 10 ciudades mayores de 16 países**

| | | | |
|---|---|---|---|
| SWEDEN | NETHERLANDS | CANADA | FRANCE |
| Min. :-0.1612 | Min. :0.1106 | Min. :0.2279 | Min. :0.2989 |
| 1st Qu.:-0.1038 | 1st Qu.:0.1590 | 1st Qu.:0.4030 | 1st Qu.:0.3733 |
| Median :-0.0555 | Median :0.2393 | Median :0.4321 | Median :0.4366 |
| Mean : 0.1270 | Mean :0.4041 | Mean :0.5081 | Mean :0.5827 |
| 3rd Qu.: 0.2904 | 3rd Qu.:0.6901 | 3rd Qu.:0.5504 | 3rd Qu.:0.6759 |
| Max. : 0.8960 | Max. :0.9385 | Max. :1.0760 | Max. :1.4490 |
| MEXICO | ARGENTINA | SPAIN | ENGLAND |
| Min. :0.2695 | Min. :0.3874 | Min. :0.2279 | Min. :0.4757 |
| 1st Qu.:0.3606 | 1st Qu.:0.4302 | 1st Qu.:0.4397 | 1st Qu.:0.5473 |
| Median :0.5015 | Median :0.5096 | Median :0.5382 | Median :0.6976 |
| Mean :0.6340 | Mean :0.6436 | Mean :0.6570 | Mean :0.8122 |
| 3rd Qu.:0.8244 | 3rd Qu.:0.7553 | 3rd Qu.:0.6938 | 3rd Qu.:0.8451 |
| Max. :1.4940 | Max. :1.4720 | Max. :1.4150 | Max. :1.9020 |
| ITALY | WGERMANY | BRAZIL | URSS |
| Min. :0.5263 | Min. :0.7528 | Min. :0.4440 | Min. :0.9624 |
| 1st Qu.:0.6498 | 1st Qu.:0.8215 | 1st Qu.:0.7366 | 1st Qu.:1.0160 |
| Median :0.8326 | Median :0.8542 | Median :0.9061 | Median :1.0360 |
| Mean :0.8744 | Mean :0.9469 | Mean :0.9744 | Mean :1.1600 |
| 3rd Qu.:1.0660 | 3rd Qu.:1.0230 | 3rd Qu.:0.9841 | 3rd Qu.:1.1070 |
| Max. :1.3730 | Max. :1.3410 | Max. :1.6970 | Max. :1.8020 |
| JAPAN | USA | INDIA | CHINA |
| Min. :0.8476 | Min. :0.8751 | Min. :0.8579 | Min. :1.046 |
| 1st Qu.:0.9302 | 1st Qu.:0.9499 | 1st Qu.:0.9973 | 1st Qu.:1.245 |
| Median :1.1020 | Median :1.0980 | Median :1.2220 | Median :1.357 |
| Mean :1.1900 | Mean :1.2000 | Mean :1.2110 | Mean :1.400 |
| 3rd Qu.:1.2610 | 3rd Qu.:1.3710 | 3rd Qu.:1.3500 | 3rd Qu.:1.552 |
| Max. :2.0420 | Max. :1.8910 | Max. :1.6570 | Max. :1.839 |



Las transformaciones de potencia son monótonas para valores positivos, luego:

**los estadísticos de orden de los datos transformados serán los estadísticos de orden originales transformados**

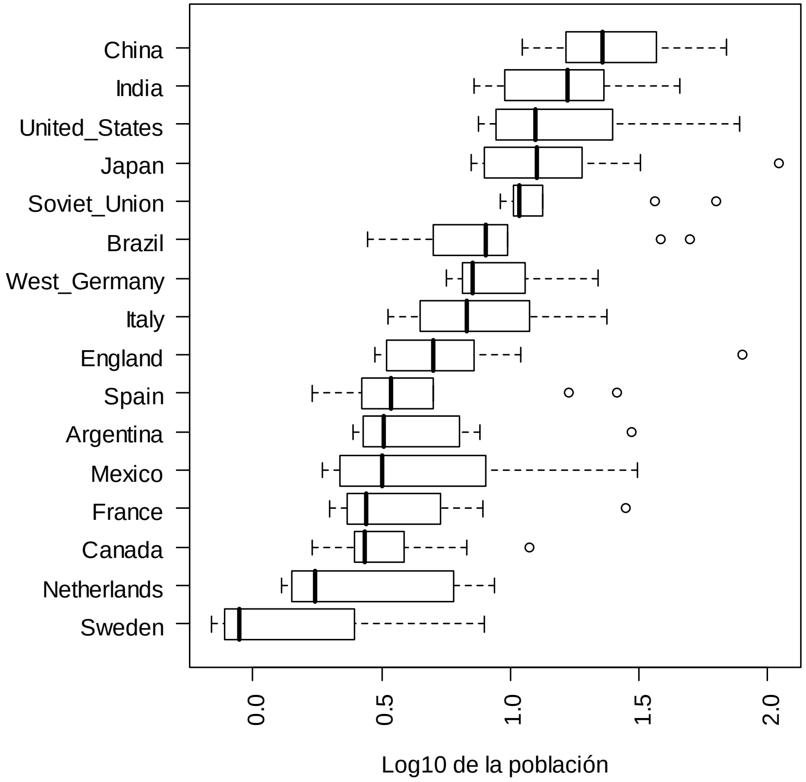
(salvo por los efectos del redondeo e interpolación).

Para obtener los boxplots es necesario recalcular la distancia intercuartos y los puntos de corte para los outliers.

Puede ocurrir que algunos datos que originalmente eran outliers del lado alto dejen de serlo y aparezcan algunos del lado bajo. Esto último es poco probable cuando tomamos los 10 valores mayores.

En la figura 18 observamos que las cajas son más similares en longitud y que la desigualdad remanente no parece estar muy relacionada con el nivel (aunque quizás se hayan reducido un poco de más las dispersiones de los niveles más altos, recordemos que la pendiente de la recta ajustada era 0.7).

**Figura 18. Boxplots de los logaritmos de las poblaciones de las 10 ciudades mayores en 16 países.**




```
> par(las=2,cex=1,mar=c(5,7,2,2))
> boxplot(log10(pobl16), ylab= "", horizontal=T )
```

En la nueva escala muchos outliers han sido llevados hacia adentro. De los 19 originales, 8 ya no son outliers y los demás se han movido hacia los puntos de corte superior.
Los nuevos boxplots son más fáciles de mirar y los países están desplegados a un nivel de detalle similar.



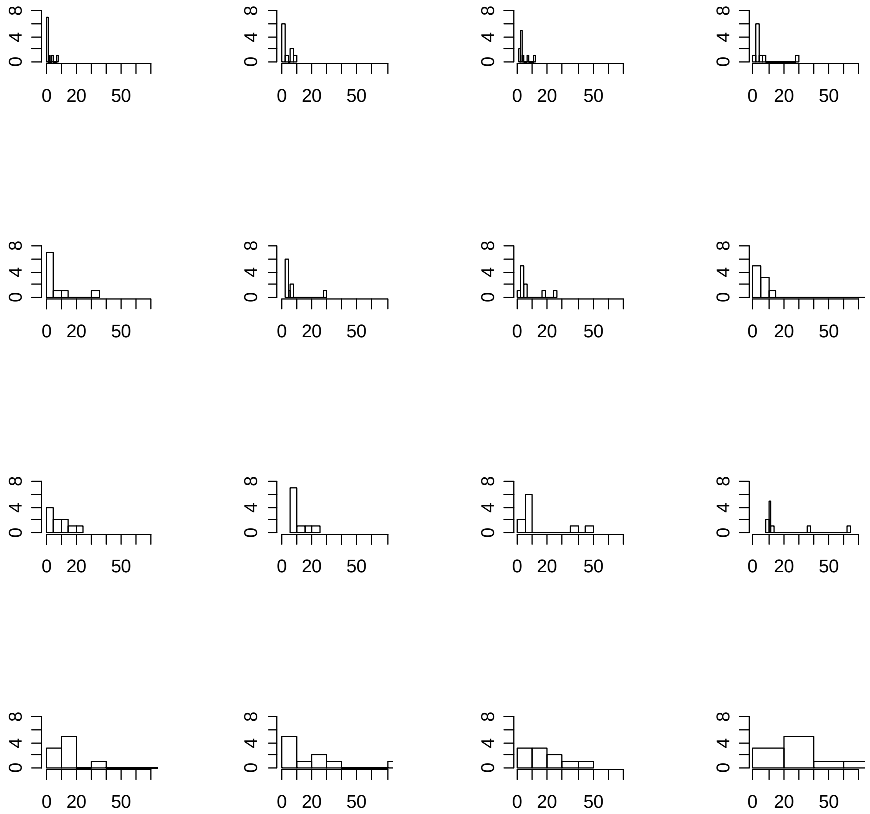
En la escala original los valores de Suecia (Sweden), Holanda (Netherlands) y Canadá son más difíciles de leer del gráfico que los de India y China.

En los gráficos con escala logarítmica los detalles aparecen similarmente bien para todos los países.

**Múltiples histogramas en un gráfico**

Datos sin transformar

```
x11()
  par(mfrow=c(4,4))
apply( pobl16,2,hist , main="", xlab="",
ylab="",
       xlim=c(0,70),
ylim=c(0,8),breaks="FD")
```



```
x11()
par(mfrow=c(4,4))
for (i in c(1:16)){
 nombre <- names(pobl16)[i]
 hist(pobl16[,i],xlab=  nombre,  ylab=  "",
main= " ",
      xlim=c(0,70), ylim=c(0,8), breaks="FD"
)
```



```
}
```