

GUÍA TEÓRICO PRÁCTICA 7

TRANSFORMACIONES

Ejercicio 1: En el estudio de las 10 ciudades más grandes de 16 países (clases teóricas) se llegó a la conclusión que la transformación raíz cuadrada era muy débil y logaritmo era un poco fuerte (aunque satisfactoria). Estas observaciones sugieren que una transformación intermedia, como por ejemplo la raíz cuarta, podría también ser efectiva para estabilizar la dispersión de los 16 lotes.

a) ¿Por qué decimos que la raíz cuarta es una transformación intermedia entre la raíz cuadrada y el logaritmo?

b) ¿Qué pendiente en el gráfico de dispersión-nivel corresponde a $p=1/4$?

c) Aplique la transformación raíz cuadrada a los datos y determine el efecto en los lotes graficando las distancias intercuartiles versus las medianas para la nueva escala.

d) Construya boxplots paralelos, ordenados por la mediana, para los datos transformados por logaritmo y por la raíz cuarta. ¿Cuál de las dos transformaciones parece más efectiva para estabilizar la dispersión?

e) Por cuadrados mínimos obtuvimos una recta con pendiente 0.69. Repita la parte d) para la transformación de potencias sugerida por esta pendiente.

Ejercicio 2:

a) Lea los datos que están en el objeto **ships** de la librería **MASS** y vea en el **help** de qué se tratan.

```
> library(MASS)
> v1<-ships$incidents
> v2<-ships$type
> v1
 [1] 0 0 3 4 6 18 0 11 39 29 58 53 12 44 0 18
[17] 1 1 0 1 6 2 0 1 0 0 0 0 2 11 0 4
[33] 0 0 7 7 5 12 0 1
> v2
 [1] A A A A A A A A B B B B B B B C C C C C C C C
[25] D D D D D D D D E E E E E E E E
```

b) Sumo **uno** a todos los valores de la variable incidentes para eliminar los ceros.

```
> v11 <- v1+1
```

PASOS PARA CONSTRUIR EL GRAFICO DE DISPERSION VS. NIVEL

c) Vea qué produce la función **split**. Utilice el **help**.

```
> data.class(v11)
[1] "numeric"

> aaa1 <- split(v11,v2)
> data.class(aaa1)
[1] "list"
```

d) Cálculo de la mediana y cuartiles superior e inferior de la cantidad de incidentes por cada tipo de barco (A, B, C, D, E). ¿Qué devuelve la función **sapply**? ¿Qué obtiene con **> sapply(aaa1, sort)** ?

```
>qinf<-(sapply(aaa1, sort)[2, ]+ sapply(aaa1, sort)[3, ])/2
>qsup<-(sapply(aaa1, sort)[6, ]+ sapply(aaa1, sort)[7, ])/2

> dq <- qsup-qinf
> medianas<- sapply(aaa1, median)

> logdq <-log10(dq)
> logmedianas <- log10(medianas)
> plot(logmedianas, logdq)
```

¿Que transformaciones de la escalera de Tukey le parecen adecuadas para los datos en estudio, de acuerdo con el gráfico de dispersión nivel?

d) Compare los boxplots de los datos originales con los de los transformados por logaritmo y raíz cuadrada:

```
> lv11<-log10(v11)
> sqrtv11<-sqrt(v11)

> aaa<-split(v1,v2)
> laaa<-split(lv11,v2)
> sqraaa<-split(sqrtv11,v2)

> par(mfrow=c(3,1))
```

```
> boxplot(aaa[sort.list(sapply(aaa,median))],  
+ main="accidentes segun tipo de barco",  
+ sub="datos crudos")  
  
> boxplot(laaa[sort.list(sapply(laaa,median))],  
+ sub="datos transformados por log10")  
  
> boxplot(sqraaa[sort.list(sapply(sqraaa,median))],  
+ sub="datos transformados por sqrt")
```

d) Ajuste de una recta por cuadrados mínimos al gráfico de dispersión-nivel

```
> lm(logdq~logmedianas)
```

e) Pruebe con la transformación sugerida por el ajuste de cuadrados mínimos para estabilizar la dispersión de los incidentes en los distintos tipos de barcos y compare.

Ejercicio 3: Un gráfico de dispersión-nivel podría construirse utilizando las medias muestrales para el nivel de los lotes y el desvío estándar muestral para las dispersiones. La relación entre la pendiente y la potencia se mantiene.

a) Calcule la media y los desvíos necesarios para la construcción del gráfico de dispersión-nivel para los datos de las grandes ciudades.

b) Construya el gráfico de dispersión-nivel modificado. ¿Cuál es la potencia sugerida?

c) ¿Cuáles son las ventajas o desventajas de utilizar esta modificación?