GUÍA Nro. 6

Ejercicio 1: Método de **medianas repetidas** para la estimación de la pendiente y ordenada al origen de una recta en un diagrama de dispersión (Ureda pag160).

a1) Genere un conjunto de datos y obtenga su diagrama de dispersión
Por ejemplo

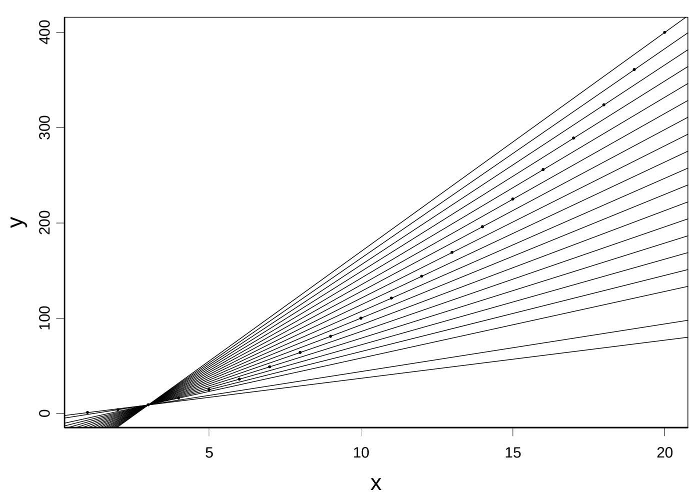
```
x<- seq(1,20)
y <- x*x
plot(x,y))
```

a2) Para esos puntos, halle la **pendiente** (*primero*) de la recta que pasa por dos puntos (por ejemplo, x[3],y[3]) ; (x[18],y[18]) y *luego* la **ordenda al origen** de la recta que tiene *esa* pendiente y pasa por uno de los puntos anteriores.
Puede hacerlo con las siguientes instrucciones (no cierre el gráfico anterior)

```
pendientes<- vector(length=length(y))
for(i in (1:length(x))){
   pendientes[i]<-(y[3]-y[i])/(x[3]-x[i])
}

vectora<- y[3]-pendientes*x[3]
for(i in (1:length(x))[-3]){
abline(vectora[i],pendientes[i])
}
```

Se obtiene el siguiente gráfico:



b) Analice detalladamente que obtiene con las siguientes expresiones vectoriales

b1)

```
> (y[1]-y)/(x[1] - x)
> (y[2]-y)/(x[2] - x)
```

```
 --------------------
> (y[20]-y)/(x[20] - x)
```

Cada una de las instrucciones anteriores podría obtenerse mediante in ciclo. ¿Cuál?

b2)Idem b1)

```
> median((y[1]-y)/(x[1] - x),na.rm=T)
> median((y[2]-y)/(x[2] - x),na.rm=T)
-------------------------------------
```

c) Para cada punto `(x[i], y[i]) i:1,...,20` y otro fijo que usted elija halle la pendiente de la recta que los une, halle la ordenada al origen de la recta que tiene esa pendiente.

d) Agregue un outlier y repita a), b) y c)
Por ejemplo:

```
x <-c(x,20)
y <- c(y,1)
```

e) Analice detalladamente que realiza la siguiente función

```
repmedians <- function(x,y) {
     med.pendientes<- vector(length=length(y))
     #mediana de las pendientes para cada punto fijo i
     for(i in 1:length(x)){
         med.pendientes[i]<-median((y[i]-y)/(x[i] - x),na.rm=T)
     }
     bRM =median(med.pendientes)
     vectora<- y - bRM*x
     aRM <- median(vectora)
     list(ord.origen=aRM,pendiente=bRM)
}
```

f) Si fun es una función de 2 argumentos, la función `outer(v1,v2, FUN)` aplica la función FUN a todas las combinaciones de pares posibles formados con un valor del vector v1 y otro valor del vector v2. Mire el help y genere sus propios ejemplos.

g) Obtenga un vector con las medianas de las pendientes de las rectas que pasan por cada punto

```
med.pendientes=rep(0,length(x))
  for(i in 1:length(x)){
         med.pendientes[i]<-median((y[i]-y)/(x[i] - x),na.rm=T)
     }
```

y compare con lo obtenido con la siguiente instrucción

```
apply( outer(y,-y,"+")/outer(x,-x,"+") ,1,mediana )
```

donde mediana <- function(x)median(x,na.rm=T)

h) Analice la siguiente función vectorial

```
   repmedianas <- function(x,y){
   bRM <- median(apply( outer(y,-y,"+")/outer(x,-x,"+"),1,mediana
   ))
      vectora<- y - bRM*x
      aRM <- median(vectora)
      list(ord.origen=aRM,pendiente=bRM)
  }
```

i) Compare:

```
    repmedians(x,y)#con ciclos semivectorial
    repmedianas(x,y)#cálculo matricial
```

j) Con los datos originales y con los datos con un outlier obtenga las rectas de medianas repetidas, la recta de cuadrados mínimos y la recta resistente explicada en clase.

```
    plot(x,y)
    abline(unlist(repmedianas(x,y)),col=5)
    abline(lm(y~x),col=3)
    abline(ltsreg(x,y),col=6)
```

Ejercicio2: Recordemos que las transformaciones de potencia tienen la forma

$$T_p(x) = \begin{cases} ax^p + b (p \neq 0) \\ c\log\ x + d (p = 0) \end{cases} \tag{1}$$

donde a, b, c, d y p son números reales, con a > 0 para p > 0 y a < 0 para p < 0.

Muestre que todas las transformaciones de potencia especificadas en la definición anterior preservan el orden de cualquier lote de números positivos.

Construya un ejemplo en el cual el orden no se preservaría si c fuera negativo cuando p = 0.

Ejercicio 3: Muestre un conjunto de datos de tamaño $n = 6$ que ilustre que la mediana es, aproximadamente pero no exactamente, preservada por una transformación raíz cuadrada.

Ejercicio 4:Sea $T(x)$ una transformación de potencias. La mediana del lote transformado es igual (salvo diferencias por interpolación o redondeo) a la transformada de la mediana de los datos crudos. Una relación similar no es satisfecha por la media muestral ni por la esperanza poblacional. Este ejercicio muestra que el apareamiento puede reducir esa diferencia.

a) Muestre que la correspondiente transformación apareada a $T(x)$ en x_0 está dada por

$$z = a + bT(x)$$

con

$a = x_0 - \dfrac{T(x_0)}{T'(x_0)}$ y $b = \dfrac{1}{T'(x_0)}$

b) Sea

$Z = a + bT(X)$

la variable aleatoria definida por la transformación apareada dada en el ítem a).

Utilice un desarrollo de Taylor de $T(X)$, alrededor de x_0 hasta el término cuadrático para obtener la siguiente aproximación para la esperanza de Z

$$E(Z) \approx x_0 - \frac{T(x_0)}{T'(x_0)} + \frac{1}{T'(x_0)}\Big[T(x_0) + E(X - x_0)T'(x_0)$$

$$+ \frac{1}{2}E(X - x_0)^2 T''(x_0) \Big]$$

y por lo tanto

$$E(Z) \approx E(X) + \frac{1}{2}E(X - x_0)^2 \frac{T''(x_0)}{T'(x_0)}.$$

c) Muestre que

$E(X - x_0)^2 = \mathrm{Var}(X) + (E(X) - x_0)^2$

y si x_0 está cerca de $E(x)$ entonces

$$E(Z) \approx E(X) + \frac{1}{2}\mathrm{Var}(X)\frac{T''(x_0)}{T'(x_0)}.$$

Podemos interpretar la ecuación anterior como sigue:

Si Z es una transformación apareada de X en un punto cercano a la media de X, la media de Z es aproximadamente igual a la media de X más un término que depende de la dispersión de X y de la curvatura de la transformación.

Ejercicio 5: Las tablas 1 y 2 muestran transformaciones no apareadas

a) Halle tres transformaciones lineales para aparear los datos transformados con los datos originales.

b) Halle los valores letra y reconstruya las tablas 1 y 2 para los datos apareados

c) Explique porqué la comparación de los cuatro lotes se facilita mediante el apareamiento.

d) Observe como cada una de las transformaciones acerca los valores altos hacia el centro y aleja los valores bajos hacia afuera, inclusive más allá de la cota natural el cero.

Tabla 1. Despliegue de los valores letra y los resúmenes medios para una muestra de ingresos familiares; en la escala original y en la escala logarítmica (

		Escala original Ingresos familiares en dólares			Escala logarítmica (base 10) Ingresos familiares		
#	994						
M	497.5		3480.0			3.54	
F	249	2412.0	3678.0	4944.0	3.38	3.54	3.69
E	125	1788.0	4115.5	6443.0	3.25	3.53	3.81
D	63	1517.0	4400.5	7284.0	3.18	3.52	3.86
C	32	1248.0	4799.0	8350.0	3.10	3.51	3.92
B	16.5	963.5	4978.8	8994.0	2.98	3.47	3.95
A	8.5	727.5	5241.0	9754.5	2.86	3.43	3.99
Z	4.5	579.0	5394.5	10210.0	2.76	3.39	4.01
Y	2.5	345.0	5510.3	10675.5	2.54	3.28	4.03
	1	114.0	5494.0	10874.0	2.06	3.05	4.04

Tabla 2. Despliegue de valores letra junto con los resumen medios para las escalas de raíz cuadrada y raíz cuarta en el ejemplo de los ingresos familiares.

		Escala raíz cuadrada Ingresos familiares			Escala raíz cuarta Ingresos familiares		
#	994						
M	497.5		58.1			7.68	
F	249	49.1	59.7	70.3	7.01	7.70	8.39
E	125	42.3	61.3	80.3	6.50	7.73	8.96
D	63	38.9	62.1	85.3	6.24	7.74	9.24
C	32	35.3	63.4	91.4	5.94	7.75	9.56
B	16.5	31.0	62.9	94.8	5.57	7.65	9.74
A	8.5	27.0	62.9	98.8	5.19	7.57	9.94
Z	4.5	24.1	62.6	101.0	4.91	7.48	10.1
Y	2.5	18.6	60.9	103.3	4.31	7.24	10.2
	1	10.7	57.5	104.3	3.27	6.74	10.2