

## 6 Gráficos cuantil- cuantil (Q-Q plots)

Utilizaremos este tipo de gráficos en el análisis de los datos de microarreglos.

Los gráficos cuantil-cuantil permiten comparar la distribución de un conjunto de datos con:

- la de otro conjunto de datos.
- una distribución ideal teórica.

Si los dos conjuntos de datos tienen la misma distribución el gráfico cuantil-cuantil es lineal.

### 6.1 Función `quantile`

La función `quantile` de una muestra  $x$  es la inversa de la función de distribución empírica (FDE), esto es

$$\text{quantile}(p) = \min \{ z \mid \text{proporción } p \text{ de datos } \leq z \}$$

### 6.2 Función `qqplot`

La función `qqplot(x, y, plot=T)` grafica las funciones `quantile` de una muestra vs. la de la otra.

Las siguientes son las instrucciones que deben ejecutarse en R para obtener un gráfico cuantil-cuantil de

- dos conjuntos de datos almacenados en los vectores  $x$  e  $y$   
> `qqplot(x,y)`
- un conjunto de datos almacenados en el vector  $x$ , comparado con la distribución Normal

```
> qqnorm(x)
> qqline(x)
```

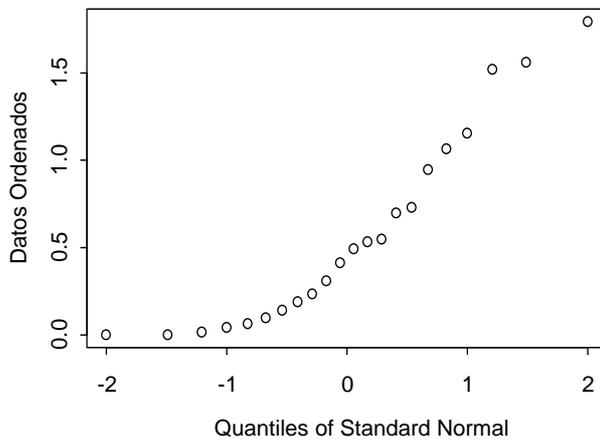
### 6.3 Función qqnorm

La función qqnorm(x) grafica las funciones quantile de una muestra, es decir los cuantiles muestrales vs. los cuantiles teóricos de la distribución Normal.

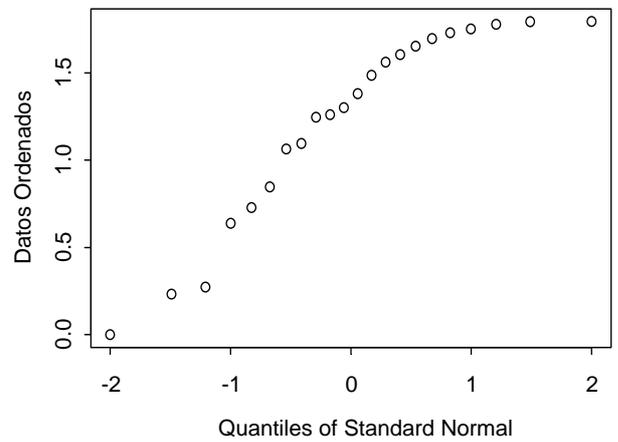
Si el conjunto de datos es “razonablemente” Normal los puntos del gráfico anterior estarán sobre una recta.

#### Principales alejamientos respecto de la distribución Normal que pueden visualizarse en un gráfico cuantil- cuantil

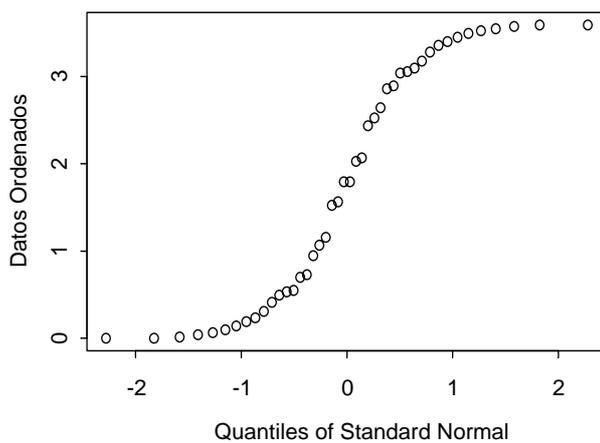
Colas Pesadas a Derecha



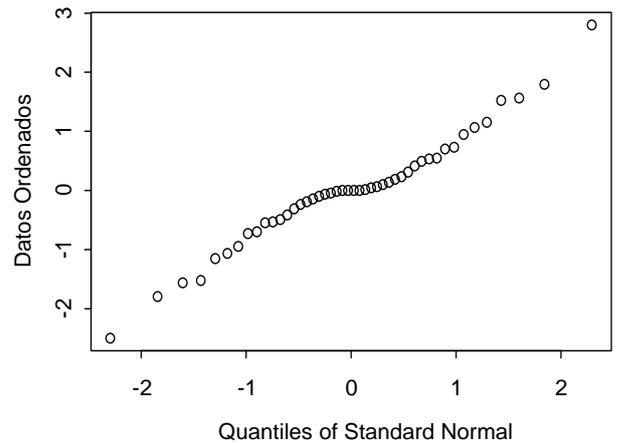
Colas Pesadas a Izquierda



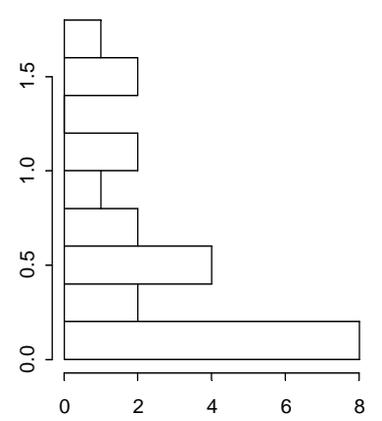
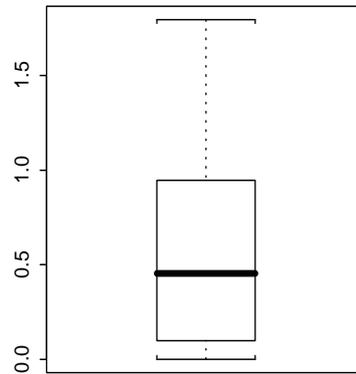
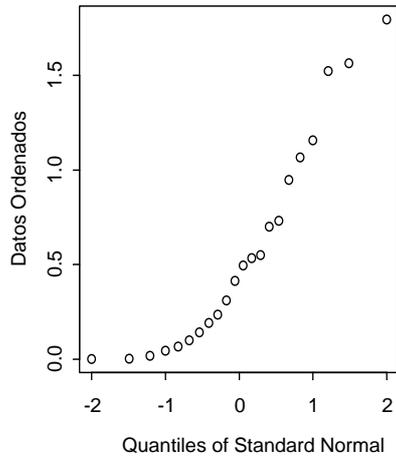
Colas Cortas



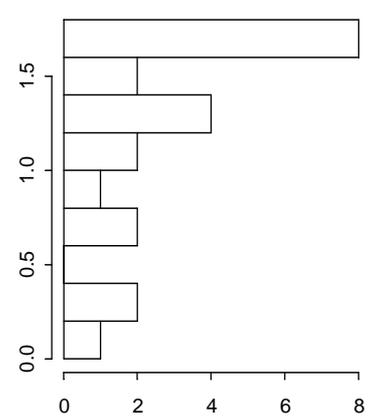
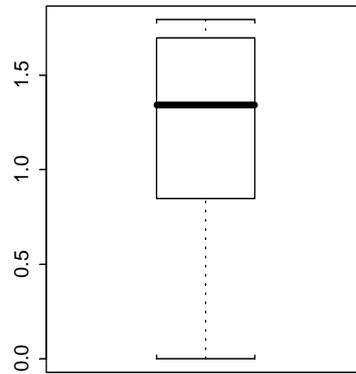
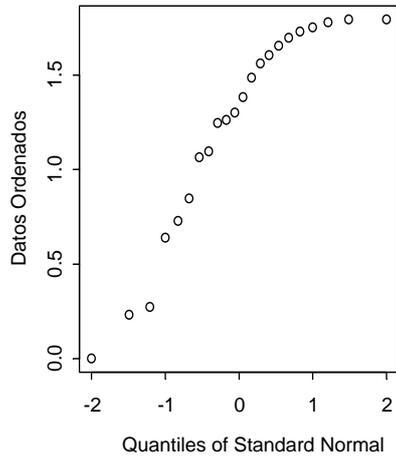
Colas Largas



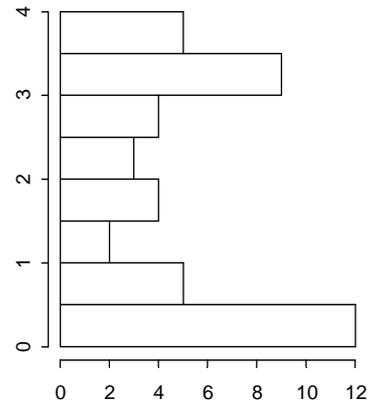
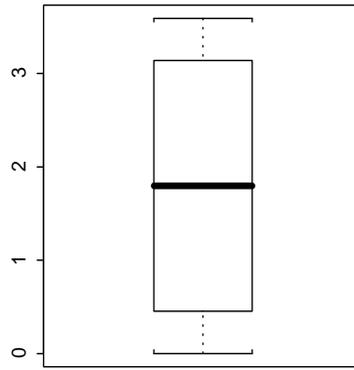
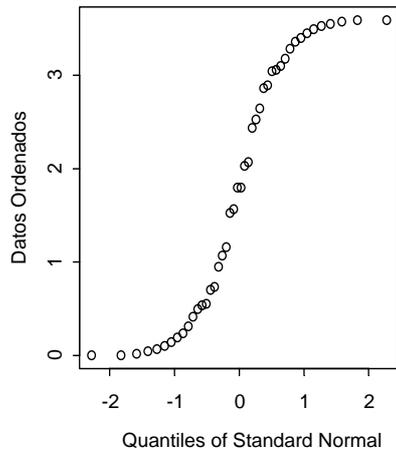
### Colas Pesadas a Derecha



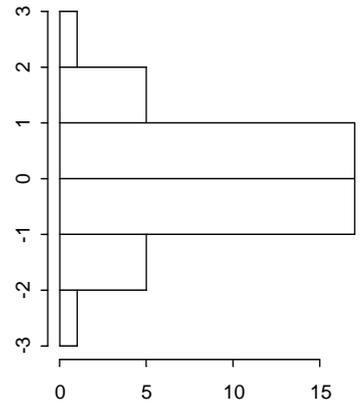
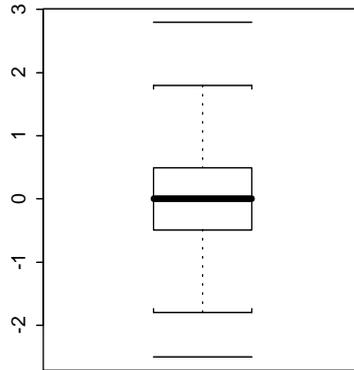
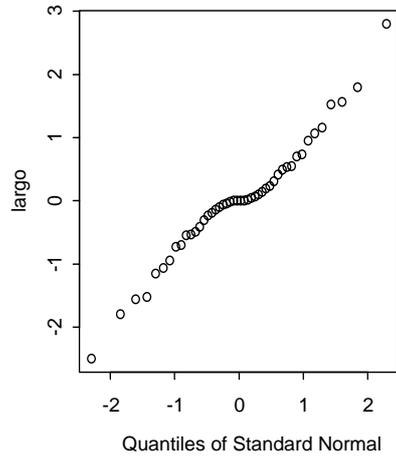
### Colas Pesadas a Izquierda



### Colas Cortas



### Colas Largas



## 6.4 Ejemplo detallado de un gráfico Cuantil-Cuantil

Comparamos los percentiles empíricos de un conjunto de datos, con los percentiles teóricos de una Normal

Datos	Posición i	$p_i$ (i-0.5) / n	$\phi_i^{-1}(p_i)$	Datos	Posición i	$p_i$ (i-0.5) / n	$\phi_i^{-1}(p_i)$
3	1	0.02	-2.00	25	12	0.52	0.06
11	2	0.07	-1.49	26	13	0.57	0.17
16	3	0.11	-1.21	27	14	0.61	0.29
19	4	0.16	-1.00	28	15	0.66	0.41
20	5	0.20	-0.83	29	16	0.70	0.54
21	6	0.25	-0.67	29	17	0.75	0.67
22	7	0.30	-0.54	30	18	0.80	0.83
23	8	0.34	-0.41	30	19	0.84	1.00
24	9	0.39	-0.29	31	20	0.89	1.21
24	10	0.43	-0.17	32	21	0.93	1.49
24	11	0.48	-0.06	34	22	0.98	2.00

El data frame **ejemploQQ** contiene inicialmente una única variable V1 que ordenamos en la misma variable

```
ejemploQQ$V1 <- sort(ejemploQQ$V1)
```

```
n <- length(ejemploQQ$V1)
```

```
a <- 0.5
```

Agregamos las variables, **pi** (vector de probabilidades que asignamos a los estadísticos de orden i) y **phiInv** (cuantiles teóricos Gaussianos de esas **pi**), al data frame

```
ejemploQQ$pi <- (1:n-a)/(n+1-2*a)
```

```
ejemploQQ$phiInv <- qnorm(ejemploQQ$pi)
```

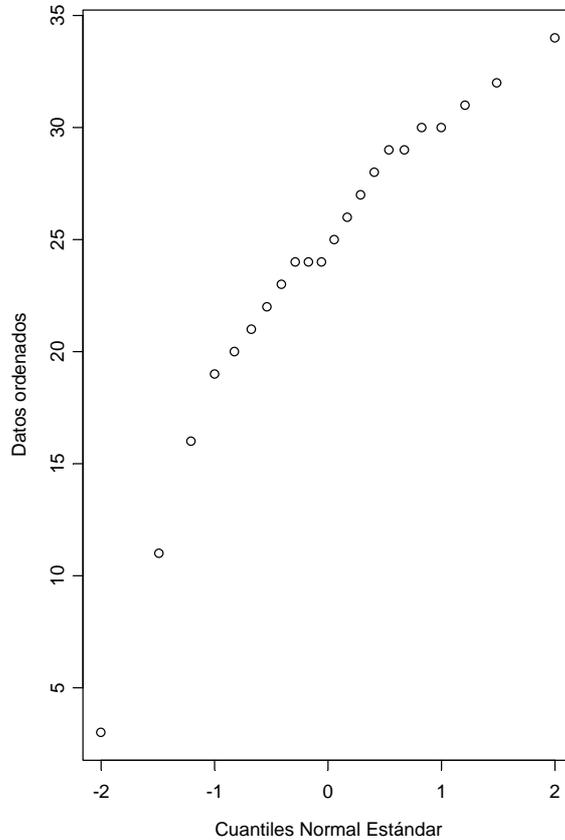
```
par(mfrow=c(1,2))
```

```
plot(ejemploQQ$phiInv,ejemploQQ$V1,  
     xlab="Cuantiles Normal Estándar",  
     ylab="Datos ordenados", main="Ejemplo(a mano)"  
)
```

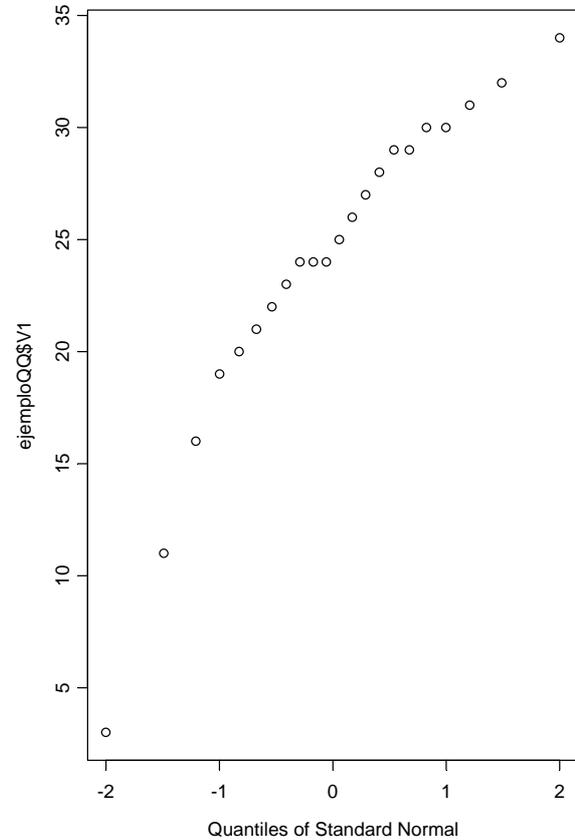
```
qqnorm(ejemploQQ$V1)
```

```
title("Usando la función qqnorm")
```

Ejemplo(a mano)



Usando la función qqnorm

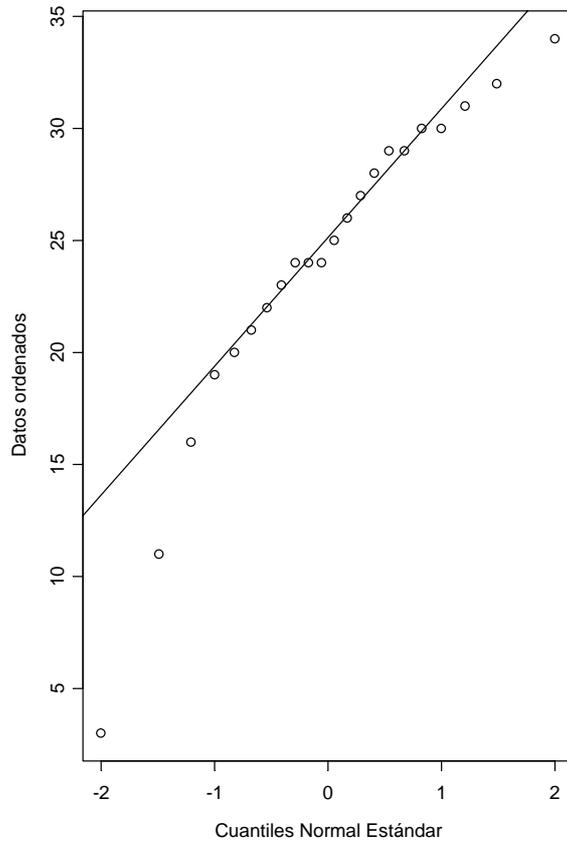


```
plot(ejemploQQ$phiInv,ejemploQQ$V1,  
     xlab="Cuantiles Normal Estándar",  
     ylab="Datos ordenados", main="Ejemplo(a mano)"  
)
```

```
datos.cuartiles <- quantile(ejemploQQ$V1, c(0.25,0.75))  
norm.cuartiles <- qnorm(c(0.25, 0.75))  
b <- (datos.cuartiles[2] - datos.cuartiles[1]) /  
     (norm.cuartiles[2] - norm.cuartiles[1])  
a <- datos.cuartiles[1] - norm.cuartiles[1] * b  
abline(a, b)
```

```
qqnorm(ejemploQQ$V1)  
title("Usando las funciones qqnorm qqline")  
qqline(ejemploQQ$V1)
```

Ejemplo(a mano)



Usando las funciones qqnorm qqline

