

Principios de Estadística

Leonardo Collado Torres y María Gutiérrez Arcelus

Licenciatura en Ciencias Genómicas, UNAM

www.lcg.unam.mx/~lcollado/index.php

www.lcg.unam.mx/~mgutierr/index.php

Cuernavaca, México
Febrero - Junio, 2009

Histogramas

Principios de
Estadística

Histogramas

Comparar
distribuciones

Comparar
cuantiles

1 Histogramas

2 Comparar distribuciones

3 Comparar cuantiles

Objetivos

Principios de
Estadística

Histogramas

Comparar
distribuciones

Comparar
cuantiles

- Vamos a aprender aplastar los datos de las anteriores gráficas, osea, hacer histogramas.
- Además, revisaremos unas gráficas útiles para comparar distribuciones visualmente.

Data!!

Principios de
Estadística

Histogramas

Comparar
distribuciones

Comparar
cuantiles

- Primero que nada, obtengamos unos datos aleatorios.

```
> set1 <- rnorm(1000)  
> set2 <- runif(1000)
```
- Si se acuerdan, en la clase pasada usamos `plot` para visualizar nuestros datos. El problema es que queríamos aplastarlos hacia el eje Y , ya que el eje X representaba la posición del número azaroso. Osea, iba desde 1 hasta 1000 en este caso.

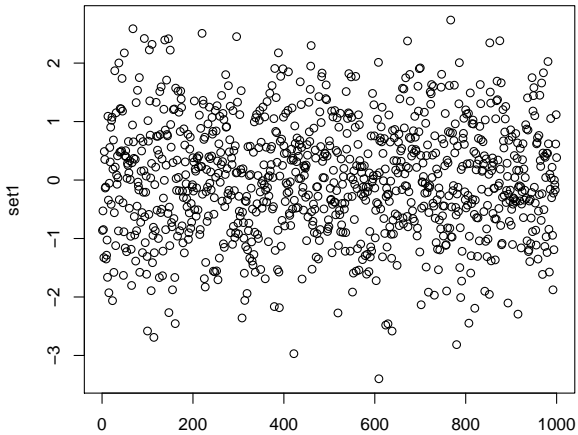
plot clase pasada

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



Index

- La forma de hacer este *aplastamiento* es con una línea representando la densidad o con un histograma usando la función `hist`. Si tienen curiosidad, hay funciones alternativas cuando su set de datos es pequeño.
 - 1 `stripchart`
 - 2 `dotchart`
- Tengan cuidado, ya que un histograma es sensible al tamaño de cada barra que uses. En R el tamaño lo escoge automáticamente, aunque pueden utilizar el argumento `breaks` para hacerlo de forma manual.
- ```
> `?`(hist)
> hist(set1)
```

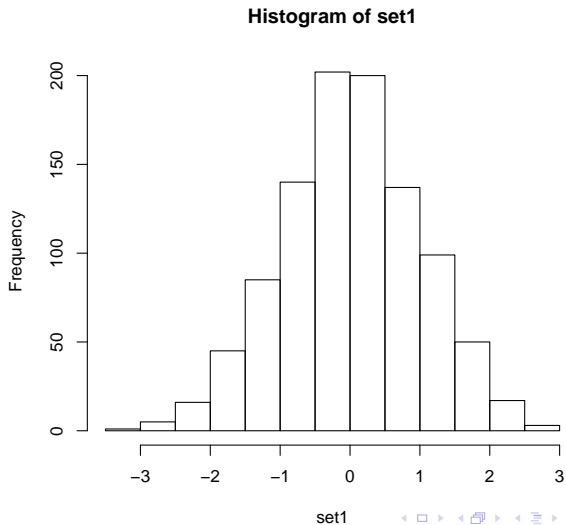
# Ejemplo con set1

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



- En el modo *default* obtenemos un histograma de frecuencias. Podemos obtener un histograma de probabilidades usando el argumento `prob=TRUE`
- Al igual que con otras funciones de gráficas pueden cambiar el color, título, etc.
- El siguiente ejemplo es con datos de R y es una distribución bimodal.

```
> hist(faithful$waiting, col = "light blue",
+ main = "Histograma de faithful$waiting",
+ xlab = "Tiempo de espera entre erupciones",
+ ylab = "Probabilidad", prob = T)
```



# Otro ejemplo

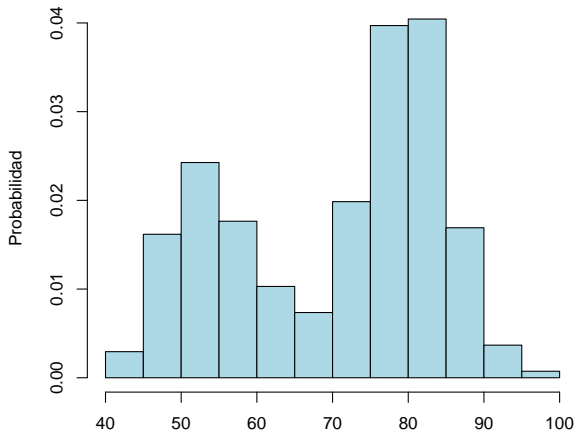
Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

Histograma de faithful\$waiting



# Usando par y mfrow

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

- Usando histogramas, una forma sencilla de comparar distribuciones es hacer 2 gráficas pegadas usando `par`. Claro, tengan cuidado porque diferentes distribuciones se pueden parecer mucho bajo ciertos parámetros.

```
> par(mfrow = c(1, 2))
> hist(set1, prob = T)
> hist(set2, prob = T)
> par(mfrow = c(1, 1))
```



# Density con Hist

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

- Lamentablemente no se pueden poner 2 histogramas en una sola gráfica, aunque en teoría podemos utilizar las funciones `lines` junto con `density` para esquivar este problema.
- Primero les muestro una donde la línea es de la misma distribución y luego otra donde no

```
> hist(faithful$waiting, prob = TRUE,
+ ylab = "Prob", col = "light blue")
> lines(density(faithful$waiting),
+ col = "red")
> set3 <- runif(100, 40, 70)
> hist(faithful$waiting, prob = TRUE,
+ ylab = "Prob", col = "light blue")
> lines(density(set3), col = "red")
```

# Iguales

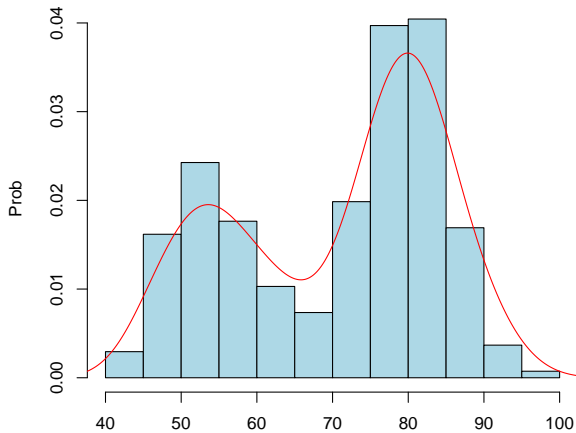
Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

Histogram of faithful\$waiting



faithful\$waiting

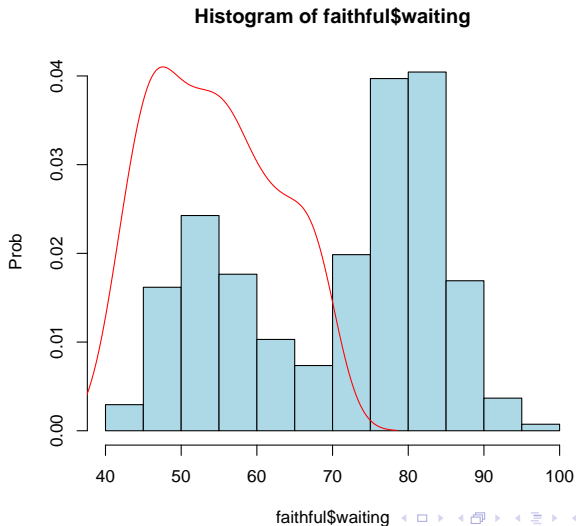
# faithful vs set3

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



# Con 2 líneas

Principios de  
Estadística

Histogramas

Comparar  
distribuciones

Comparar  
cuantiles

- Como se han de imaginar, otra forma es usar un espacio en blanco con dos líneas.

```
> plot(0, 0, ylim = c(0, 1), xlim = c(min(set1),
+ max(set1)))
> lines(density(set1), col = "red")
> lines(density(set2), col = "blue")
```

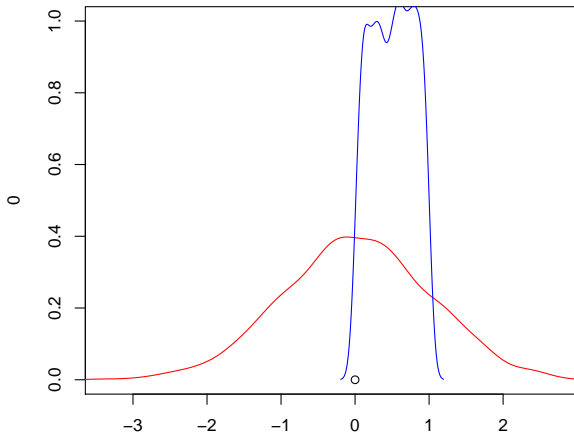
# Set1 vs Set2

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles





# Q-Q plot

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

- La forma *pro* usando gráficas para comparar dos distribuciones es con la llamada "Q-Q plot". Este tipo de gráficas usa como información los cuantiles de las dos distribuciones que vas a comparar. Puede que estes comparando una muestra contra una distribución teórica, dos muestras o dos distribuciones teóricas. Además, tiene la ventaja de que el tamaño de tus 2 poblaciones no importa.
- En R la función que hace este tipo de gráficas es `qqplot`.
- Veamos como se ve `set1` vs `set1` y `set1` vs `set2`.  

```
> qqplot(set1, set1)
> qqplot(set1, set2)
```

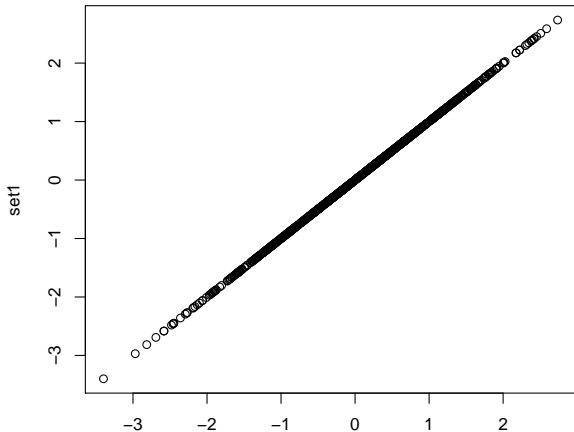
# Set1 vs Set1

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



set1

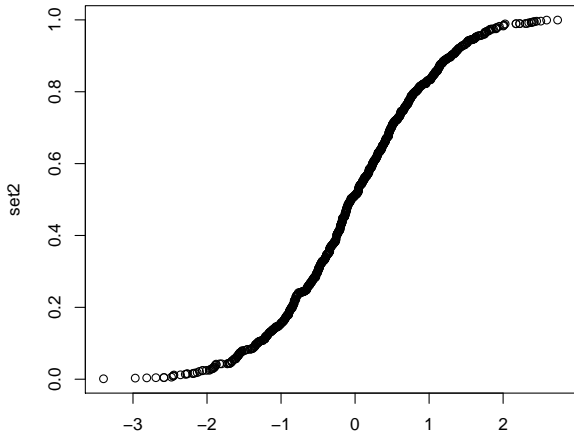
# Set1 vs Set2

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



set1

# Q-Q Norm

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

- Como ya vieron, si obtenemos una diagonal esta nos indica que las dos distribuciones son parecidas o iguales si nos da una diagonal perfecta.
- Ahora, digamos que tienen un set de datos y quieren saber si se distribuyen como normal.
- Para este caso existe la función `qqnorm` la cual compara los cuantiles de tu muestra contra los cuantiles de la normal teórica.
- Chequemos los siguientes ejemplos:
  - > `set4 <- rchisq(100, 3)`
  - > `qqnorm(set4)`
  - > `qqnorm(set1)`

# QQnorm set4

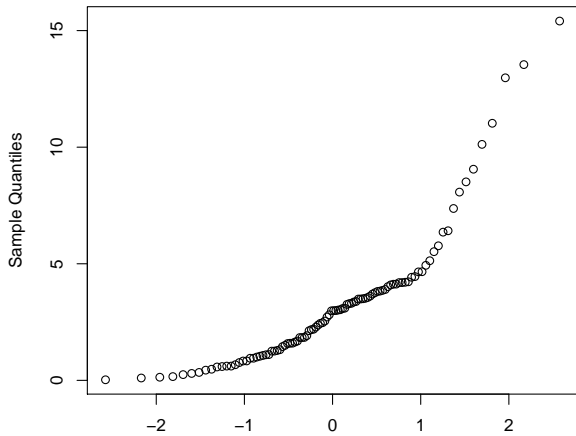
Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles

Normal Q-Q Plot



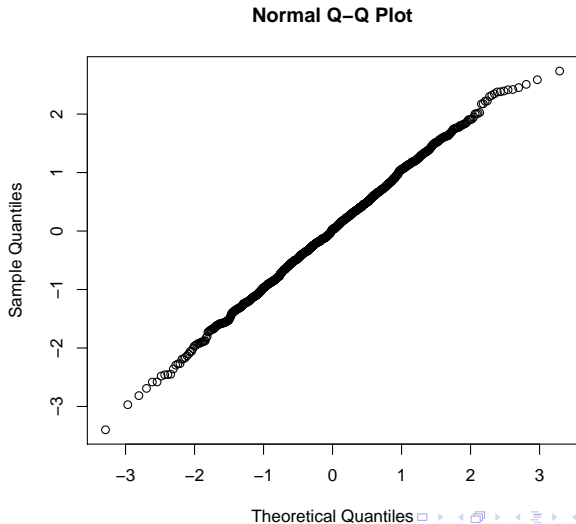
# QQnorm set1

Principios de Estadística

Histogramas

Comparar distribuciones

Comparar cuantiles



# Terminando

Principios de  
Estadística

Histogramas

Comparar  
distribuciones

Comparar  
cuantiles

- Recuerden que nuestro set1 son datos que obtuvimos con `rnorm` mas no obtenemos una diagonal perfecta con `qqnorm`.
- Si se fijan, en el centro si existe nuestra diagonal pero en los bordes se curva. Esto es por que tenemos valores extremos y es muy difícil que estos correspondan a los valores extremos teóricos :P

# Tarea

Principios de  
Estadística

Histogramas

Comparar  
distribuciones

Comparar  
cuantiles

- Asistir al evento de mañana jueves :)
- Hacer un reporte estadístico donde comparen la cantidad de "líquido" ingerido por la población de la LCG en este evento vs un evento promedio.
- Suerte!!!