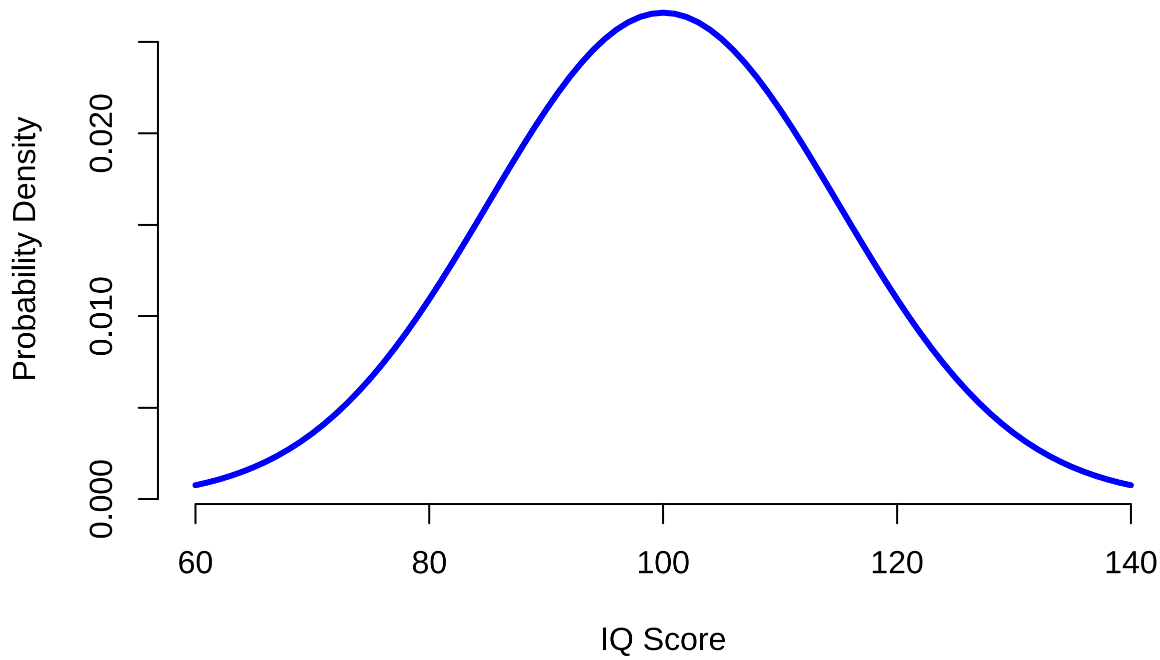
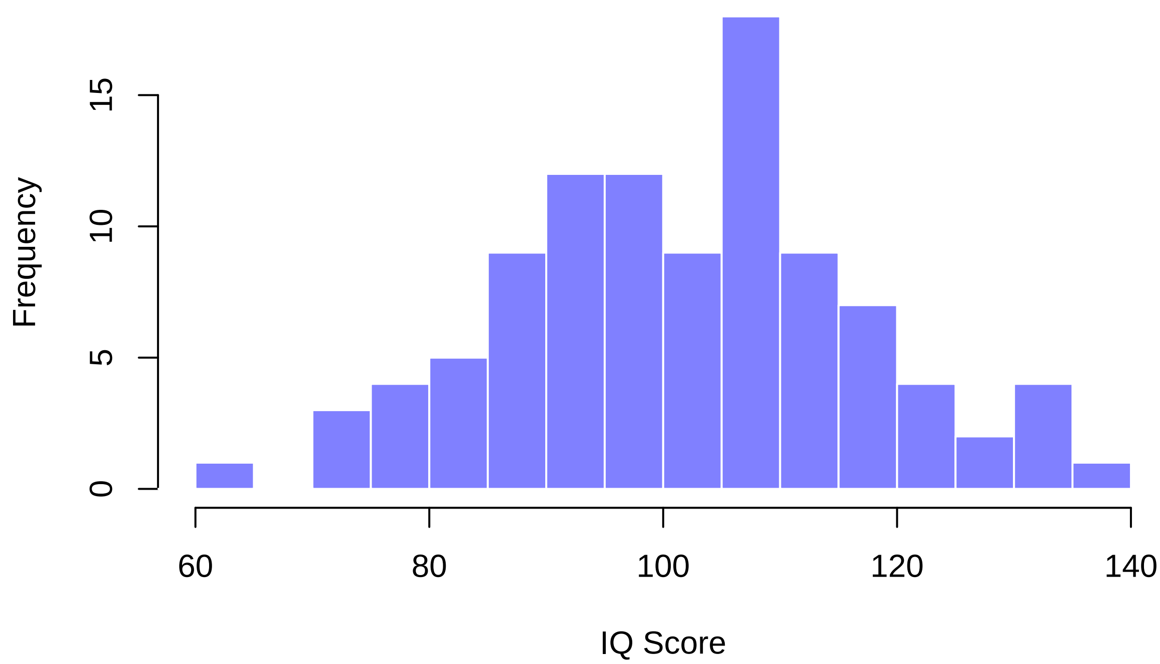
Prepa3

Alimi Garmendia

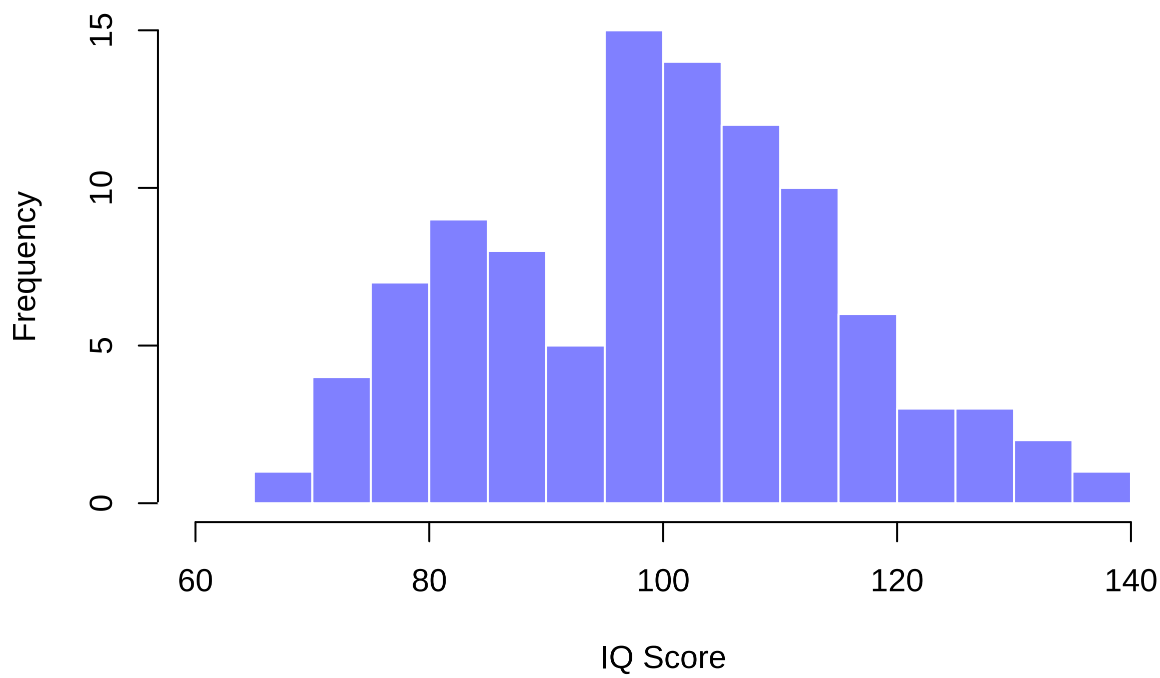
6/4/2021

Muestreo, realmente importa?

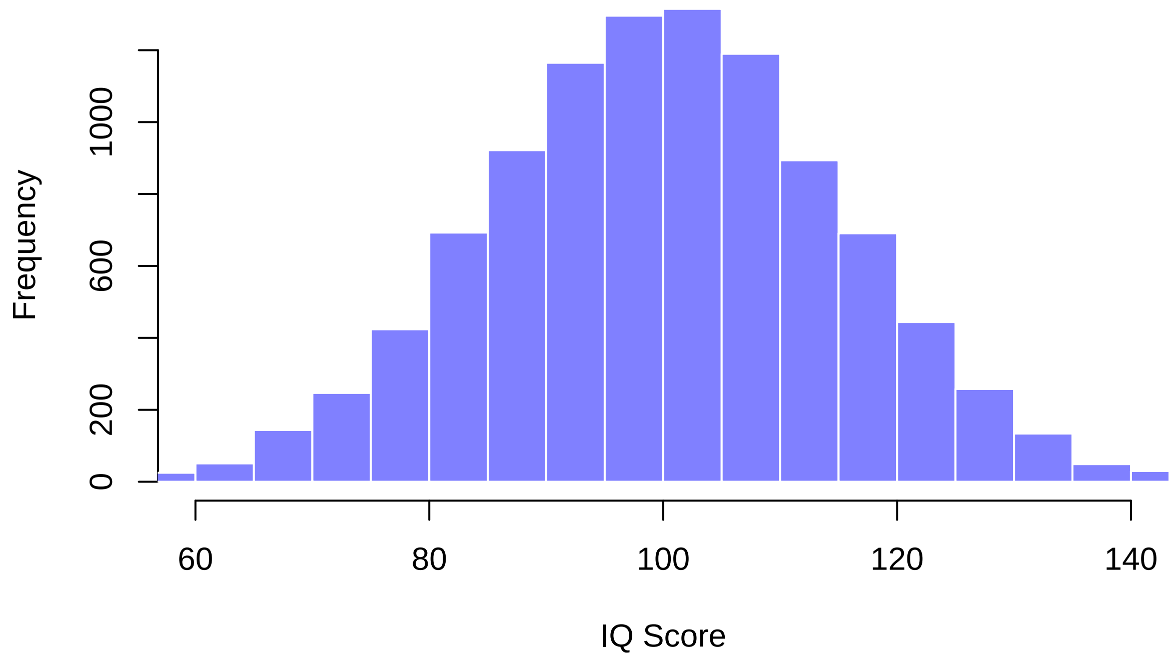




```
## [1] "n= 100 mean= 101.538883761345 sd= 15.3367186620186"
```




```
## [1] "n= 100 mean= 99.5875485938297 sd= 15.2909112094141"
```




```
## [1] "n= 10000 mean= 100.098111303822 sd= 15.0106257608165"
```

No siempre vamos a tener acceso a muchas muestras de la poblacion, es por ello que debemos muestrear de la mejor manera posible, que las muestras, en conjunto, representen de mejor manera la poblacion.

Una vez tomada nuestra muestra debemos *estimar* las medidas estadisticas, una vez mas, escogiendo los aquellos estimadores que tengan menor cesgo o *bias*

Estimadores

Para estimar la media de la poblacion, podemos usar el mismo procedimiento que usamos para calcular la media de la muestra, es decir

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Asi $\overline{X}$\$ es un estimador insesgado de la media poblacional \$$\mu$.

simbolo	que es?	Sabemos que es?
$\overline{X}$	Media de la **muestra**	Si, lo calculamos de los datos
μ	Media real de la **poblacion**	Casi nunca sabremos su valor con certeza
$\overline{\mu}$	Estimado de la media de la poblacion	Si, es identico a la media de la muestra

Para estimar la desviacion estandar de la muestra tenemos que realizar un pequeños ajuste de lo que conocemos

Habiamos definido la varianza de una muestra como

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2$$

Sin embargo, no podemos asumir este valor como estimador de la varianza de la poblacion pues este estimador es un estimador cesgado. Se puede demostrar que, para obtener un estimador incesgado de la varianza poblacional podemos usar como estimador

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2$$

Luego podemos estimar la desviacion estandar usando

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2}$$

simbolo	que es?	Sabemos que es?
S^2	Varianza de la **muestra**	Si, la calculamos de los datos
σ^2	Varianza de la **poblacion**	Casi nunca sabremos su valor real con certeza
$\hat{\sigma}^2$	**Estimador** de la varianza de la **Poblacion**	Si, es **casi** igual que la varianza de la muestra pero con un ligero ajuste

Como hemos visto, no es deficil calcular los estimadores de la poblacion, sin embargo, como podemos corroborar, estos estimadores dependen de la muestra que tome. Si tomo muchas muestras y calculo varias veces estos estimadores, veremos que varian de muestra en muestra. Es por eso que debemos buscar la manera de asegurar el rango en que nuestros estimadores se encontraran, con algun grado de certeza.

Intervalos de confianza

Suponiendo que nuestras muestras tienen una distribucion aproximadamente normal. Que la media de la poblacion es μ y la SD σ . Imaginemos que he hecho un estudio que incluye N y que la media del IQ de estos participantes es $\overline{X}$. Recordemos que, en una distribucion normal, hay una prob de 95% que una cantidad con distribucion normal se encuentre dentro de dos desviaciones* estandar de la media. Es decir:

```
qnorm(p = c(0.025,0.975))
```

```
## [1] -1.959964  1.959964
```

Es decir, nuestra media muestral cumple que:

$$\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}$$

Asi, trabajando un poco el algebra, podemos llegar a

$$\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

Asi, este rango de valores tiene 95% de probabilidad de contener la media de la poblacion.

Ejercicios

1. Imaginemos que quieres rentar un apartamento de una habitacion en Caracas. La media de la renta mensual para una muestra aleatoria de 60 apartamentos que viste en el periodico es de $1000. Asuma que la varianza de la poblacion es de $200. Construya un intervalo de confianza del 95%

   ```
   cuantil = qnorm(c(0.025,0.975))
   sqrt_n = sqrt(60)
   mean_muestra = 1000
   varianza = 200

   IC = 1000+cuantil*varianza/sqrt_n
   IC
   ```

   ```
   ## [1]  949.3939 1050.6061
   ```

2. Sobre cual poblacion de los apartamentos de Caracas podemos inferir, dados los resultados del apartado anterior?

   Podemos usar el resultado anterior para estimar la media de los apartamentos de que vismo en el periodico. No podemos asegurar nada de todos los apartamentos de Caracas, pues no podemos asegurar que la muestra del periodico sea representativa

3. Que tan grande debe ser la muestra de los apartamentos de los apartados anteriores si queremos estiumar la media de la poblacion dentro de un margen de $50 con una confianza del 90%?

```
alpha = 1-0.9
cuantil = qnorm(c(1-alpha/2))
cuantil
```

```
## [1] 1.644854
```

$$50 = Z * \frac{\sigma}{\sqrt{n}}$$

```
n = (cuantil*varianza/50)**2
n
```

```
## [1] 43.2887
```

Es decir, para tener un margen de error de +- $50 necesitariamos una muestra aleatorioa de 44 apartamentos