Resumen para el segundo parcial

Hipótesis nulas y p-valores para los tests

Kolmogorov-Smirnov

Este test pone a prueba si una función desconocida (F_X) tiene la misma distribución que la conocida (F_0). Entonces si el $p - valor > \alpha$ diríamos que no hay evidencia suficiente para rechazar H_0.

$$\begin{cases} H_0 : F_x = F_0 \\ H_1 : F_x \neq F_0 \end{cases}$$

Ejemplo: **Ejercicio 1 del práctico 9.**

Donde la distribución conocida es $Unif(4.7, 5.9)$. Y los datos son: [5.3, 5.1, 4.8, 4.9, 5.3, 5.2, 5.8, 5.5, 5.6, 5.2]. La prueba de hipótesis va a ser de la siguiente forma:

$$\begin{cases} H_0 : F_x = Unif(4.7, 5.9) \\ H_1 : F_x \neq Unif(4.7, 5.9) \end{cases}$$

Primero planteo Kolmogorov-Smirnov:

$$K_n = D \times \sqrt{n} \to \mathcal{K}$$

La región crítica va a tener la siguiente forma:

$$RC = \left\{ \sqrt{n} \times D > t_{1-\alpha} \right\}$$

Para $n > 30$ utilizamos:

$$t_{1-\alpha} \approx \sqrt{-\frac{1}{2} Ln(\frac{\alpha}{2})}$$

Sino hay que usar la tabla y buscar el valor $t_{1-\alpha}$ para un n y un α.

Ahora continuamos hallando el valor de D.

$$D = sup \, |F_X(x) - F_0(x)|$$

$F_X(x)$ es el valor de la distribución desconocida que más se aleja de F_0. En este caso es el quinto valor (en orden ascendente) y es aproximadamente ~0.7 (linea roja). Y el $F_0(x)$ sería el valor al que le corresponde el anterior si fuese de la distribución conocida, en este caso, podemos ver donde se cruzan las rectas azules y es igual a *~0.5*. Entonces, ahora hago todas las cuentas con los números:

$$D = sup \, |0.7 - 0.5| = 0.2$$

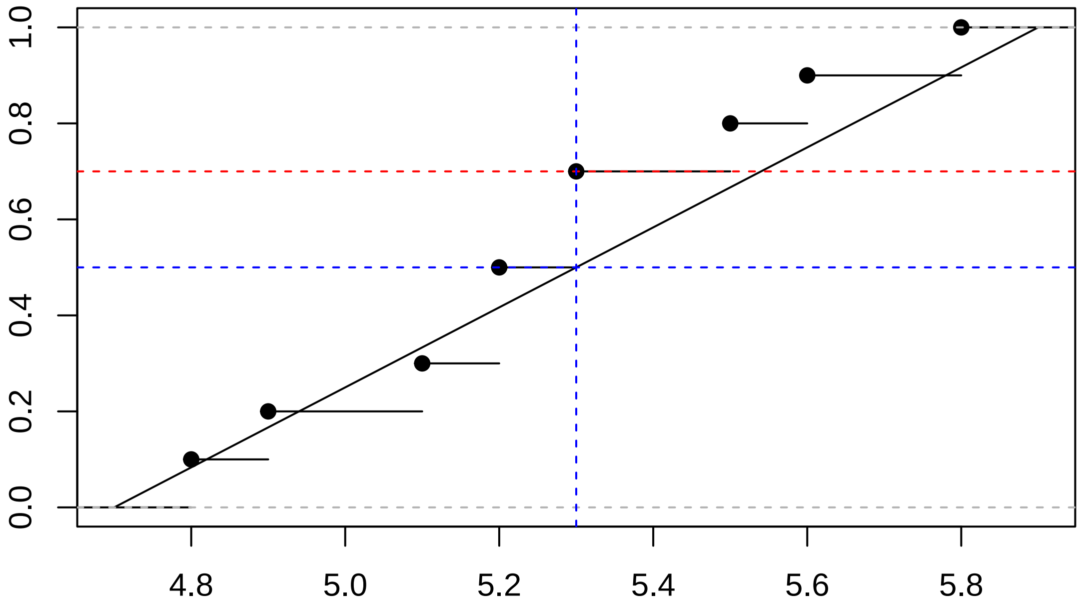
Utilizo la tabla para buscar el valor de $t_{1-\alpha}$ para $n = 10$ y $\alpha = 0.05$ (recordar que es bilateral) y obtengo:

$$t_{1-\alpha} = 0.40925$$

Finalmente:

$$RC = \left\{ D \times \sqrt{n} > 0.40925 \right\}$$

Ya que $D \times \sqrt{n} = 0.63$ es más grande que $t_{1-\alpha}$ se puede decir que rechazamos H_0.




```
test_ks <- ks.test(datos,"punif", 4.7, 5.9)
test_ks
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  datos
## D = 0.2, p-value = 0.8186
## alternative hypothesis: two-sided
```

Al realizar el ejercicio en R obtenemos un p-valor de 0.819, ya que es mayor a *0.05* podemos decir que no hay evidencias significativas como para rechazar H_0. Y esto es completamente lo contrario a lo que nos da al calcular *"a mano"*, donde rechazabamos H_0

Kolmogorov-Smirnov 2 muestras:

Esta prueba testea si dos distribuciones son iguales. La prueba de hipótesis va a ser de la siguiente forma:

$$\begin{cases} H_0 : F_X = F_Y \\ H_1 : F_X \neq F_Y \end{cases}$$

El estadístico va a ser:

$$K_{m,n} := \sqrt{\frac{nm}{n+m}} \sup |F_n(x) - F_m(x)| \xrightarrow{d} \mathcal{K}$$

La región crítica será:

$$RC = \{D_{n,m} > t_{n,m}\}$$

Ahora hay que hallar un $t_{1-\alpha}$ tal que $P(\mathcal{K} > t_{1-\alpha}) = \alpha$. Y este valor se obtiene desde una tabla o usando la fórmula aproximada si $n > 40$ igual que en la parte de K-S de una muestra.

Rechazaremos H_0 si:

$$\sqrt{\frac{nm}{n+m}}D_{n,m} > t_{1-\alpha}$$

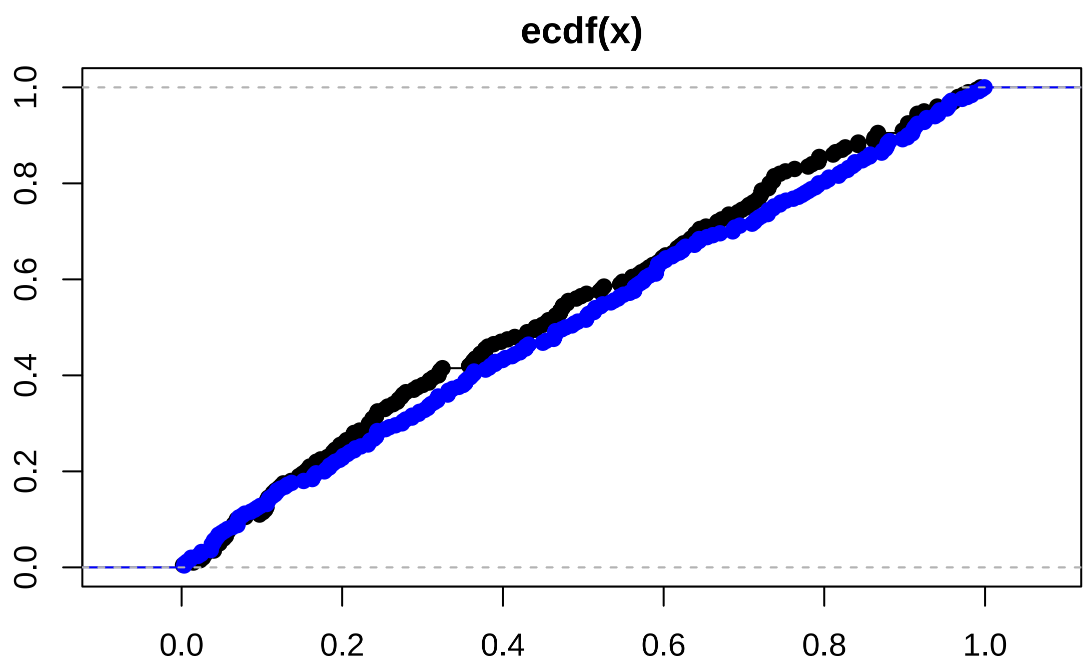
El D se obtiene de igual forma que el test para una mustra.

Ejemplo:

```
x=runif(200)
y=runif(250)
ks.test(x, y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.068, p-value = 0.6831
## alternative hypothesis: two-sided
```

```
# Gráfica
par(mar = c(2,2,2,2))
plot(ecdf(x))
plot(ecdf(y), add=TRUE, col="blue")
```



Si el $p-valor > \alpha$ entonces diríamos que no hay evidencia suficiente para rechazar H_0.

Lilliefors

No realizamos este test "*a mano*", solo se hizo en R

Pone a prueba lo mismo que Kolmogorov-Smirnov, pero se utiliza cuando no se conoce la media, ni la varianza, ni ningún parámetro que necesite la distribución. H_0 será de la forma:

$$\begin{cases} H_0 :\in \{N(\mu,\sigma) : \sigma > 0, \mu > 0\} \\ H_1 : \text{"no pertenece a la familia"} \end{cases}$$

Donde todo lo que esta entre {} es el conjunto de todas las variables (en este caso) con alguna media μ y algún desvío σ.

Ejercicio 3 del práctico 9

```r
# No son los mismos que se ven en el práctico porque estos ya estan estandarizados.
datosLilliefors <- c(-0.085 , -2.46 , -0.335 , 0.235 , 0.645 , -0.075 , 0.965 , -1.38 , -2.115 ,
↪  -0.055 , -0.48 , 0.615 , 1.315 , 0.11 , -0.445 , 0.61)
# Realizo el test
testLilliefors <- LcKS(datosLilliefors,"pnorm")
```

Este test devuelve los siguientes datos:

1. Un "D observado" 0.2006556.
2. Un vector muy grande que se llama "D.sim", no se qué significa.
3. El p-valor: 0.0836.

Como el $p - valor > \alpha$ no hay evidencias suficientes para rechazar H_0 a nivel *5%*, es decir que la muestra proviene de una $\mathcal{N}(\mu, \sigma)$

Prueba de $\mathcal{X}^2$ de Pearson

Este test sirve únicamente para **distribuciones discretas**, no continuas como los anteriores.

Prueba de Spearman de una muestra

Testea si una muestra tiene un comportamiento monóntono, si es creciente o no, si es $i.i.d$ no tiene un comportamiento monótono.

La pruba de hipótesis sera:

$$\begin{cases} H_0 : X_1, ..., X_n \text{ son i.i.d} \\ H_1 : X_1, ..., X_n \text{ no son i.i.d} \end{cases}$$

Y el estadístico tiene la siguiente forma:

$$\rho_s = 1 - \frac{6S}{n(n^2 - 1)}$$

Con

$$S = \sum_{i=1}^{n} (R_i - i)^2$$

Ejercicio 1 del práctico 10

i	valores	Ri
1	74.3	12
2	74.1	11
3	75.4	14
4	67.4	1
5	69.3	3
6	70.5	8
7	70.1	5
8	69.9	4
9	68.7	2
10	70.3	7
11	70.7	9
12	71.1	10
13	74.4	13
14	70.2	6

Entonces calculamos S y nos queda:

$$S = (12-1)^2 + (11-2)^2 + (14-3)^2 + ... + (6-14)^2 =$$

Prueba de Spearman de dos muestras