Nombre:	**CI:**	**Nro 2do informe**

SEGUNDA PRUEBA PARCIAL

28 de Noviembre de 2020

Ejercicio 1, 10 puntos

Un investigador recolecta información sobre los patrones de actividad física (A.F.) de los niños de quinto grado de una escuela. Define tres categorías: alta, media, baja. También indaga sobre el consumo de bebidas azucaradas y define 2 categorías: 1 si consume, 2 no consume. Los datos obtenidos son los siguientes

A.F	Consume	No consume
baja	32	12
media	14	22
alta	6	9

Se desea saber si existe o no dependencia entre los patrones de AF y el consumo de bebidas azucaradas.

Para realizar una prueba de hipótesis se tiene la siguiente salida de R

```
datos=matrix(c(32,14,6,12,22,9),ncol=2)
chisq.test(datos)

##
##  Pearson's Chi-squared test
##
## data:  datos
## X-squared = 10.712, df = 2, p-value = 0.004719
```

6 Plantear la prueba de hipótesis que se desea testear (**2 puntos**), el estadístico usado (**2 puntos**), y hallar la región crítica a nivel $\alpha = 0{,}05$ (**2 puntos**). [1]

4 Realizar la prueba de hipótesis a nivel $\alpha = 0{,}01$. Justificar. **4 puntos si hace todo bien, 0 si hace y no justifica nada**

Sea X la variable que indica la actividad física, que toma valores en el conjunto $\{baja, media, alta\}$ y sea Y la variable que indica si el individuo consume o no bebidas azucaradas. Los datos que tenemos son entonces pares $(X_1, Y_1), \dots, (X_{94}, Y_{94})$. Queremos ver si X e Y son independientes a partir de la muetra anterior, que se resume en la tabla. La prueba de hipótesis a plantear es $H_0 : X$ e Y son v.a. independientes contra H_1 no lo son. Para eso se construye una cuadro de contingencia y se calcula el estadístico de Pearson (sección 5.3 de las notas) que se denota T_n. La región crítica es $RC = \{T_n > t_n(\alpha)\}$ donde $t_n(\alpha)$ es tal que $P_{H_0}(RC) = \alpha$, y se usa que nT_n se parece para valores grandes de n a una variable con distribución χ^2 con dos grados de libertad (ya que en nuestro caso $(F-1)(C-1) = 2$). Según los datos que se proporcionan en la letra de parcial, el valor nT_n es 10.712, este valor es más grande que 6 que es el cuantil 0,95 de la χ^2 con 2 grados de libertad. Por lo tanto se rechaza H_0, es decir que son independientes, a nivel $\alpha = 0{,}05$. El otro dato que nos da la salida de R es el p-valor, que en este caso es menor que 0.01, por lo tanto también se rechaza H_0 a nivel $\alpha = 0{,}01$.

[1] recordar que qchisq(0.95,2)=6

Ejercicio 2, 10 puntos

Queremos saber si la siguiente muestra de datos es iid o no.

```
datos = c(-4.3, -2.1, -0.42, -0.39, -0.2, -0.88, 0.15, 0.3, 0.6, 0.7, 2.0, 3.1, 3.5)
```

3 ¿Qué test de hipótesis se puede usar para verificar eso? Hallar el valor del estadístico. **3 si hace todo bien, 1 si se equivoca en algo**

Se realiza una prueba de Spearman de 1 muestra, en la cual se testea H_0 : la muestra es iid, contr a H_1 la muestra no es iid. El valor del estadístico se calcula por medio del vector de rangos. Dicho vector es $(1, 2, 4, 5, 6, 3, 7, 8, 9, 10, 11, 12)$. Por lo tanto el valor del estadístico es

$$1 - 6 * 12/(13 * (13^2 - 1)) = 0{,}967033.$$

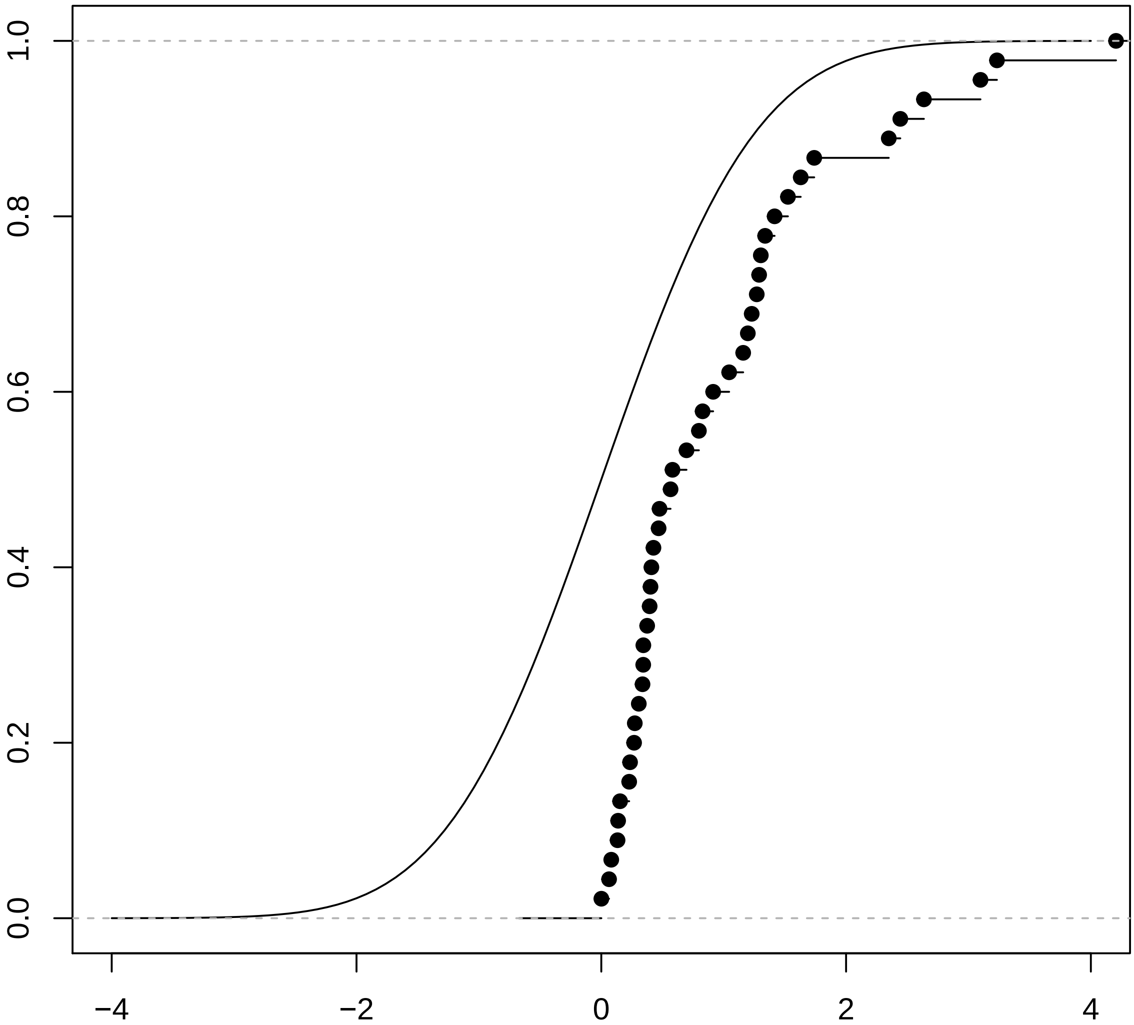
4 ¿Qué se puede concluir de la misma si tomamos el nivel $\alpha = 0{,}05$?.

3 Dar un valor aproximado del p-valor.

El valor del estadístico está muy próximo a 1, con lo cual el $p - valor$ va a ser casi 0. Esto se puede ver tambien en la tabla de Spearman, donde si vamos a la fila $n = 13$ de la misma vemos que el p-valor tiene que ser mucho menor que 0,002 (que en particular es menor que 0,05), (ya que $P_{H_0}(\hat{\rho}_{13} > 0{,}797)) = 0{,}002$, y el valor que obtuvimos es bastante mas grande que 0.797. Por lo tanto claramente se rechaza H_0 a nivel $\alpha = 0{,}05$.

Ejercicio 3, 15 puntos

Se desea testear si una muestra de 45 datos se ajustan a una distribución normal con media 0 y varianza 1. Para eso se cuenta con el siguiente gráfico donde la curva es la función de distribución de dicha normal, y la función escalonada es la distribución empírica de los datos.



3 Plantear la prueba de hipótesis que se desea testear.

La prueba que se desea contrastar es $H_0 : F_X = N(0,1)$ contra $H_1 : F_X \neq N(0,1)$ donde F_X es la función de distribución de los datos.

5 Hallar la región crítica (aproximada) para $\alpha = 0{,}05$.[2]

En este caso usamos que $P_{H_0}(\sqrt{n}D_n)$ se parece, para valores grandes de n a una variable aleatoria K, cuyo cuantil $1-\alpha$ se aproxima por $\sqrt{-(1/2)\log(\alpha)}$. Por lo tanto la región crítica (usando el dato que nos dan), es $\{\sqrt{n}D_n > \sqrt{3/2} = 1{,}2247\}$.

5 Usando el gráfico anterior, dar un valor aproximado para el estadístico D_n de Kolmogorov-Smirnov.

2 Realizar la prueba con el valor dado en la parte anterior.

Es claro del gráfico que el valor de D_n es mayor que 0.4, si calculamos, $\sqrt{45} * 0{,}4 = 2{,}68$, que es mayor que 1.2247, por lo tanto se rechaza H_0.

[2] Puede ser de utilidad que $\log(0{,}05)) \approx -3$