Facultad de Ciencias | **Bioestadística**
Universidad de la Repúblicas | **Curso 2019**

Nombre:	**CI:**	**Año en que cursó**

Solución Examen - 21 de Febrero de 2020

Problema 1:
Sea $f : \mathbb{R} \to \mathbb{R}$ la función dada por

$$f(x) = \begin{cases} 1 - x/2 & \text{si } x \in [0, a] \\ 0 & \text{en otro caso} \end{cases}$$

1. Hallar a para que f sea una densidad (de una variable que llamaremos X)
   Para que f sea de densidad debemos plantear:
   $$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$
   Resolviendo dicha ecuación se obtiene que $a = 2$.

2. Hallar la función de distribución de X
   La función de distribución corresponde a:
   $$F_X(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x - \frac{x^2}{4} & si \quad 0 \leq x \leq 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

3. Calcular la esperanza, varianza y mediana de X
   - **Esperanza:** $E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^2 x - \frac{x^2}{2}\,dx = \frac{2}{3}$
   - **Varianza:** $Var(X) = \int_{-\infty}^{+\infty} \left(x - \frac{2}{3}\right)^2 f(x)\,dx = \frac{2}{9}$
   - **Mediana:** $Q_2 = \inf\{x \in \mathbb{R} : F_X(x) \geq \frac{1}{2}\} = 2 - \sqrt{2}$

4. Sea $Y = 3X + 1$, hallar su esperanza.
   $E(Y) = 3E(X) + E(1) = 3$

Problema 2:
Dependiendo de cada tipo, los coronavirus pueden causar diversas afecciones, desde el resfriado común hasta enfermedades más graves, como bronquitis, bronquiolitis, neumonía, el síndrome respiratorio de Oriente Medio (MERS-CoV), síndrome respiratorio agudo grave (SRAS-CoV), entre otros. Con el fin de comparar si el factor género influye en la presencia o no del virus se cuenta con una muestra aleatoria de 120 personas de los cuales 18 tienen el virus. Los resultados fueron los siguientes:

	Tiene el virus	No tiene
hombre	8	52
mujer	10	50

1. Plantear una prueba de hipótesis a nivel $\alpha = 0.05$ para decidir si la presencia del virus es independiente o no del género.
   Sea X la variable aleatoria que representa al género (Hombre o Mujer) e Y la variable aleatoria indica si una persona posee o no la enfermedad. El test a realizar corresponde a:

$$\begin{cases} H_0: & \text{El género no influye en la presencia del virus.} \\ H_1: & \text{El género si influye en la presencia del virus.} \end{cases}$$

2. Realizar la prueba de la parte anterior[1]
   Planteamos la prueba χ^2 de independencia. Para esto calculemos primero es estadístico de la prueba:

$$T_n = \frac{\left(\frac{8}{120} - \frac{60}{120}\frac{18}{120}\right)^2}{\frac{60}{120}\frac{18}{120}} + \frac{\left(\frac{52}{120} - \frac{60}{120}\frac{102}{120}\right)^2}{\frac{60}{120}\frac{102}{120}} + \frac{\left(\frac{10}{120} - \frac{60}{120}\frac{18}{120}\right)^2}{\frac{60}{120}\frac{18}{120}} + \frac{\left(\frac{50}{120} - \frac{60}{120}\frac{102}{120}\right)^2}{\frac{60}{120}\frac{102}{120}} = 0.00218$$

   Por otro lado debemos resolver determinar t tal que $P(\chi_1^2 > t) = 0.05$ entonces $t = qchisq(0.95, 1) = 3.84$ y por lo tanto $t_n = \frac{t}{n} = \frac{3.84}{120} = 0.032$.
   Por último dado que $T_n < t_n$ no se rechaza H_0 a nivel 0.05.

Problema 3: Se considera la base de datos *iris* que contiene la información sobre las medidas de los sépalos y pétalos de tres especies de iris. Se realizan las siguientes pruebas para averiguar si hay alguna dependencia lineal entre el ancho del sépalo y el largo del pétalo, así como también entre el ancho del pétalo y su largo.

```
lin1<-lm(Sepal.Width~Petal.Length,iris)
summary(lin1)

##
## Call:
## lm(formula = Sepal.Width ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08463 -0.21537  0.02116  0.21587  1.10380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.45487    0.07610  45.402  < 2e-16 ***
## Petal.Length -0.10579    0.01834  -5.768 4.51e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3952 on 148 degrees of freedom
## Multiple R-squared:  0.1836,Adjusted R-squared:  0.178
## F-statistic: 33.28 on 1 and 148 DF,  p-value: 4.513e-08

lin2<-lm(Petal.Width~Petal.Length,iris)
summary(lin2)

##
## Call:
## lm(formula = Petal.Width ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56515 -0.12358 -0.01898  0.13288  0.64272
##
```

[1] qchisq(0.95,1)=3.84

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.363076   0.039762  -9.131  4.7e-16 ***
## Petal.Length  0.415755   0.009582  43.387  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2065 on 148 degrees of freedom
## Multiple R-squared:  0.9271,Adjusted R-squared:  0.9266
## F-statistic:  1882 on 1 and 148 DF,  p-value: < 2.2e-16
```

1. Plantear los modelos teóricos que se quieren estudiar.
   El modelo teórico correspondiente a la primera prueba es $S = aP + b + \epsilon$, siendo S la variable que representa el ancho del sépalo, P el largo del pétalo y ϵ el error con distribución $N(0, \sigma^2)$.
   Para la segunda prueba en análogo.

2. Plantear y realizar una prueba de hipótesis (a nivel 0.05) y decidir si hay o no una dependencia lineal en ambos casos.
   La prueba de hipótesis a realizar para el primer caso es:

$$\begin{cases} H_0: & a = 0 \\ H_1: & a \neq 0 \end{cases}$$

   Dado que el p-valor $4.51e^{-08} < 0.05$ se rechaza H_0, es decir hay una relación lineal entre la variable ancho del pétalo y la variable largo del pétalo.
   Para la segunda prueba en análogo y tambien se concluye que hay dependencia lineal entre las variables.

3. Elegir y justificar en cual de los casos tenemos más información para poder considerar el modelo lineal encontrado.
   El coeficiente de determinación de la segunda prueba es mucho más grande que el de la primera, lo cual indica que el segundo caso se ajusta mejor.