1. (1.5 puntos) Lee atentamente el siguiente código de R y el correspondiente resultado y contesta a las preguntas:

```
> x <- rnorm(10, mean = 0.5)
> ks.test(x, pnorm)

One-sample Kolmogorov-Smirnov test

data:  x
D = 0.27724, p-value = 0.3577
alternative hypothesis: two-sided
```

(a) Escribe cuál es exactamente la hipótesis nula que se está contrastando. ¿Qué decisión se toma a partir de estos resultados a nivel $\alpha = 0{,}05$? ¿Es la decisión correcta?
(b) ¿Qué mide exactamente el valor `D` de la salida anterior?

2. (1 punto) Sea $\hat{f}$ el estimador del núcleo (núcleo gaussiano y parámetro de suavizado h) calculado a partir de la muestra de tamaño 3 formada por los valores $x_1 = 1$, $x_2 = 2$ y $x_3 = 9$. Si X^* es una variable aleatoria con función de densidad $\hat{f}$, explica brevemente cómo se podrían simular realizaciones de X^*.

3. (1.5 puntos) En un grupo de 435 pacientes que habían sufrido quemaduras de tercer grado se midió el área de la zona afectada por las quemaduras [la variable `x` corresponde a $\log(\text{área} + 1)$]. Algunos de los pacientes sobrevivieron (`y=1`) y otros fallecieron (y=0). Con el fin de estudiar cómo influye el área de las quemaduras en la probabilidad de supervivencia se ajustó un modelo de regresión logística a los datos con los resultados siguientes:

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8518  -0.6998   0.1860   0.5239   2.2089

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   22.708      2.266  10.020   <2e-16
x            -10.662      1.083  -9.848   <2e-16
---

    Null deviance: 524.69  on 433  degrees of freedom
Residual deviance: 335.23  on 432  degrees of freedom
AIC: 339.23
```

(a) ¿Aportan estos datos evidencia (a nivel $\alpha = 0{,}01$) de que el área afectada es significativa para determinar la probabilidad de que el individuo sobreviva?
(b) Determina una regla de clasificación para predecir si un individuo sobrevivirá o no en función del valor de la variable `x`. ¿Cuál es la predicción si `x` vale 2,5?

4. (1.5 puntos) Sea F_3 la función de distribución empírica correspondiente a la muestra de tamaño 3 formada por los valores $x_1 = 1$, $x_2 = 2$ y $x_3 = 9$. Estas observaciones son i.i.d. con distribución desconocida F. Sea X^* una variable aleatoria cuya distribución es F_3.

(a) Calcula la varianza de X^*.

(b) Con los 3 datos anteriores, calcula el estimador bootstrap de $\mathrm{Var}_F(\bar{X})$, donde $\bar{X} = (X_1 + X_2 + X_3)/3$ con X_1, X_2, X_3 v.a.i.i.d. con distribución F.

5. (1.5 puntos) Explica línea a línea para qué sirve el siguiente código de R, así como el resultado obtenido. ¿Te parece el resultado satisfactorio?

```
> x1 <- rnorm(10)
> x2 <- rnorm(10)
> x3 <- rnorm(10)
> y <- x2 + rnorm(10, sd=0.5)
> datos <- data.frame(y, x1, x2, x3)
>
> resultado <- leaps::regsubsets(y ~ ., data = datos, method = 'forward')
> summary(resultado)

Subset selection object
Call: regsubsets.formula(y ~ ., data = datos, method = "forward")
3 Variables  (and intercept)
   Forced in Forced out
x1     FALSE      FALSE
x2     FALSE      FALSE
x3     FALSE      FALSE
1 subsets of each size up to 3
Selection Algorithm: forward
         x1  x2  x3
1  ( 1 ) " " "*" " "
2  ( 1 ) " " "*" "*"
3  ( 1 ) "*" "*" "*"

> summary(resultado)$bic

[1] -15.46186 -14.20034 -12.32209
```

6. (1 punto) Escribe la definición del error cuadrático medio integrado de un estimador de la función de densidad. Enumera las condiciones que debe cumplir el parámetro de suavizado de un estimador del núcleo de la función de densidad para que su error cuadrático medio integrado tienda a cero, si $n \to \infty$.

7. (1 punto) Explica brevemente las diferencias y semejanzas entre el estimador ridge y el estimador lasso de regresión.

8. (1 punto) La matriz de covarianzas muestral de un conjunto de datos bidimensional X es

$$S = \begin{pmatrix} 10 & 0 \\ 0 & 20 \end{pmatrix}.$$

Escribe cuál es la primera componente principal de estos datos y qué porcentaje de varianza explica.