Universidad Nacional Autónoma de México
Estadística II

Las pruebas deben tener: planteamiento de hipótesis, procedimiento, estadístico de prueba, cuantil de contraste y conclusión estadística.
Utilice R, para la conclusión utilice la prueba considerando cuantiles. El estadístico de contraste puede obtenerlo de tablas o por simulación y tome $\alpha$ =0.05

1. Se tomo una m.a. que arrojo los siguientes datos:

$$1, 1, 1, 0, 1, 0, 0, 1, 1, 1$$

   Realice una prueba de bondad de ajuste $\chi^2$ para $H_0 : X \sim Blli(\hat{\theta})$

2. Considere la siguiente m.a.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 10.13 | 3.58 | -1.10 | 5.91 | 4.31 | 5.03 | 6.93 | 3.15 | 8.22 | 3.90 |

   Realice PBA adecuada a la situación considerando

$$H_0 : X \sim N(\hat{\mu}, \hat{\sigma^2})$$

   - Utilizando prueba Kolmogorov-Smirnov
   - Utilizando prueba Anderson-Darling

   *Hint: Recuerde que para el cálculo de probabilidades bajo $H_0$ se tiene una familia que tiene distribución estandar. Haga uso de esa propiedad para usar la tabla*

3. Considere la siguiente m.a.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1,14 | 2,36 | 1,28 | 0,44 | 0,98 | 0,32 | 1,02 | 0,23 | 2,75 | 0,34 |

   Realice PBA adecuada a la situación considerando

$$H_0 : X \sim Gamma(1, 2/3)$$

   Donde la función de densidad de la v.a. $Gamma(n, \lambda)$ con $n > 0$ y $\lambda > 0$ viene dada por:

$$f_X(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \text{ para } x > 0$$

   - Kolmogorov-Smirnov
   - Anderson-Darling
   - Cramer-Von-Misses

4. Explique el uso de la función nativa de R *ad.test()* que proviene del paquete *nortest*. ¿A qué PBA se refiere? De un ejemplo de aplicación a nivel código. Y explique la salida que produce R y cómo se interpreta.

5. Para la prueba K-S Lilliefors para todos los posibles casos indique si sufre modificación la distribución asociada al estadístico de prueba habitual de K-S. Si su respuesta es afirmativa explique cómo se modifica la simulación de la distribución del estadístico de prueba asociado.

6. Para la PBA $\chi^2$ caso discreto indique el uso de la función nativa de R *goodfit()* ¿Cuáles son sus ventajas y desventajas? De un ejemplo de aplicación a nivel código.

7. Escriba un código que implemente la simulación de la verdadera distribución Cramer-Von-Misses y genere la tabla.

8. El siguiente ejercicio se enfoca en probar la potencia de la prueba para detectar normalidad.

   - Simule m=100000 muestras de tamaño $n = 20$ de una distribución $t(25)$
   - Realice la prueba de hipótesis $H_0 : X \sim N(0,1)$ utilizando PBA $\chi^2$ con $k = 4$, K-S y A-D. Registrando el número de veces que se rechazo $H_0$ para cada PBA.
   - Calcule el promedio de rechazo para $H_0$ en cada PBA, observe que esto puede interpretarse como la probabilidad de no equivocación de la prueba o bien la potencia de la prueba. En su opinión cuál considera que es la PBA más potente para detectar normalidad.

9. Recordemos lo siguiente: Si X es v.a. $Unif(0,1)$ y sea $\lambda > 0$, entonces la v.a. $Y = -\frac{1}{\lambda}ln(X)$ sigue una distribución $exp(\lambda)$. Por otra parte, si $X_1, X_2, X_3, ..., X_k$ son m.a. $exp(\lambda)$, entonces $\sum_{i=1}^{k} X_i$ sigue una distibución $Gamma(k, \lambda)$. Con la información anterior simule una muestra de tamaño 60 de una $Gamma(3,5)$ y realice PBA.

   - Utilizando $\chi^2$
   - Utilizando Kolmogorov-Smirnov
   - Utilizando Anderson-Darling
   - Utilizando Cramer-Von-Misses

10. Simular la distribución de las estadísticas de prueba de la Andersson-Darling (caso simple y caso compuesto), cuando $F_0(x;\theta)$ es la distribución normal con parámetros $(\mu, \sigma^2)$ =(2.34, 1.27), para valores de $m \in \{30, 50\}$. Hagan lo siguiente:

    *a*) Generen 5,000 muestras de tamaño m de una distribucion $N(x|2.34, 1.27)$

    *b*) Para cada muestra, calculen la estadística $A_m^2$, haciendo:

    $$u_i = \int_{-inf}^{x_i} \frac{1}{\sqrt{2.54\pi}} \exp\left\{-\frac{1}{2{,}54}(t - 2.34)^2\right\} dt, \tag{1}$$

    ordenen la muestra.

    *c*) Ordenen las 5,000 estadísticas obtenidas (que forman una muestra de la distribución de $A_m^2$ para $m \in 30, 50$, suponiendo $H_0$ verdadera), calculen los cuantiles 0.75, 0.90, 0.95, 0.99.

*d*) Digan si 1.662, 1.624, 1.508, 2.571, 2.032, 1.743, 1.475, 1.329, 1.951, 3.171, 0.350, 1.249, 3.483, 1.974, 2.165, puede ser considerada una muestra de una $N(x|2.34, 1.27)$. ahora vamos a suponer que se desconoce $(\mu, \sigma^2)$ y desea probar la hipótesis la población proviene de una muestra normal, ahora deberá usar la transformación

$$u_i = \int_{-inf}^{x_i} \frac{1}{\sqrt{s_m^2 \pi}} \exp\left\{-\frac{1}{s_m^2}(t - \bar{x}_m)^2\right\} dt,$$

donde

$$m\bar{x} = \sum_{j=1}^{m} x_j \qquad s_m^2 = \sum_{j=1}^{m} (x_j - \bar{x})^2 \tag{2}$$

verifiquen si 6.466, 7.528, 6.55682, 6.3396, 6.38785, 7.60078, 5.72222, 5.16782, 8.21855, 7.60404, 7.98301, 6.90226, 7.20831, 8.2409, 9.50381, puede ser considerada una muestra de una distribución normal.

*e*) Repitan todo lo anterior, pero en el punto (1), en lugar de generar de una densidad normal, háganlo de una distribución Cauchy con los mismos parámetros. Esto significa, que generan muestras de una Cauchy y los transforman como si fueran normales. Usen el cuantil 0.95 que encontraron anteriormente, para decidir si rechazan la hipótesis. Este último punto es para ver la potencia de la prueba.