**EXAMEN DE ESTADISTICA MATEMATICA.** 1/9/01
2º DIPLOMATURA EN ESTADISTICA

1.- Sea $X_1, X_2,\ldots,X_n$ m.a.s. de una población, se consideran los siguientes estimadores de $\theta$. Estudiar todas las propiedades que conozcas de estos estimadores y compararlos:

a) $T_1 = X_{(1)} \quad T_2 = \dfrac{2(X_1 + X_2 + \ldots + X_n)}{n}$ para una población uniforme en $(\theta, 0)$

b) $T_1 = \dfrac{(X_1 + X_2 + \ldots + X_n)}{n} \quad T_2 = \dfrac{2(X_1 + X_2 + \ldots + X_n)}{n+1}$ para una población exponencial de media $\theta$.

( 1.5 puntos)

2.- Sean $X_1,\ldots,X_n$ v.a.i.i.d. con media $\mu$ y varianza $\sigma^2$ y tales que existe el momento de orden cuatro $\mu_4=E((X_1-\mu)^4)$, siendo $\mu_4>\sigma^4$. Dar una distribución aproximada para la varianza muestral $S_n^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{n}$ donde n se considera suficientemente grande. Para ello se aconseja comenzar utilizando la siguiente expresión $S_n^2 = \dfrac{\sum_{i=1}^{n} Y_i}{n} - (\overline{X}_n - \mu)^2$ donde $Y_i = (X_i - \mu)^2$.

Utilizando dicha aproximación calcular la probabilidad de que la varianza muestral de una m.a.s. de tamaño 10 de una uniforme en (0,1) sea mayor que 0.1.

( 1.5 puntos)

3.- En cierta factoría se trabaja con dos tipos de máquinas A y B. El consumo energético diario de una máquina de tipo A es una v. a. con ley $N(\mu_A, \sigma^2)$ y el de una máquina de tipo B $N(\mu_B, 3\sigma^2)$. Se observan los consumos energéticos diarios de 10 máquinas de cada tipo elegidas al azar y se anotan las medias y varianzas muestrales siguientes:

$$\overline{X}_A = 50.19 \quad \overline{X}_B = 102.085 \quad S_A^2 = 0.8037 \quad S_B^2 = 2.628$$

En el proceso de producción que se desarrolla en dicha factoría, intervienen 2 máquinas de tipo A y una de tipo B. Responder a las siguientes cuestiones:

a) Estudiar la ley de probabilidad del estadístico $\dfrac{n_A S_A^2 + \frac{n_B}{3} S_B^2}{\sigma^2}$.

b) Obtener un intervalo de confianza del 95% para el consumo energético medio diario del proceso de producción. Para ello utilizar el estadístico del apartado anterior para deducir la ley del estadístico pivot que permite construir el intervalo pedido.

c) ¿Soportan los datos la hipótesis de que una máquina de tipo A consume menos de la mitad que una de tipo B?.

d) ¿Soportan los datos la hipótesis de que una máquina de tipo A consume la mitad que una de tipo B?. Indicar para qué niveles es posible rechazar esta hipótesis.

e) En el contraste del apartado c), suponiendo $\sigma^2$=0.64 conocido, especificar para qué niveles es posible rechazar la hipótesis nula y obtener la potencia con que se detectaría, para un nivel $\alpha$=0.05, una alternativa $\mu_A= (\mu_B/2) - 1$.

( 2 puntos)

4.- Hacer un esquema de las técnicas de bondad de ajuste que se conozcan para el caso de hipótesis compuesta, indicando en qué casos son más adecuadas y si hay problemas para obtener la distribución del estadístico test cuando proceda.
(1 punto)

5.- Una revista científica publicó datos sobre el número de meses que transcurrían entre la recepción de un artículo para su publicación y la primera respuesta a los autores por parte de la revista, para los años 1979 y 1983. Los datos se resumen como sigue:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Meses transcurridos* | 0 | 1 | 2 | 3 | 4 | 5 | >6 |
| *Nº de autores 1979* | 26 | 28 | 34 | 48 | 21 | 22 | 34 |
| *Nº de autores 1983* | 28 | 27 | 42 | 44 | 17 | 6 | 16 |

¿ Hay evidencia de una disminución en el tiempo medio de espera de 1983 respecto a 1979?.
(1. 25 puntos)

6.- Se tienen 20 máquinas de características iguales que envasan huevos en cartones. Para cada una de ellas se ha anotado el número de cartones envasados sin defecto antes de envasar el primer defectuoso, obteniéndose los siguientes resultados:

9, 13, 5, 9, 35, 14, 0, 7, 2, 1, 0, 16, 2, 5, 1, 6, 14, 6, 25, 5

a) ¿Se puede considerar que los datos anteriores proceden de una distribución geométrica?. Para contestar a la pregunta se pueden utilizar los cálculos que se indican en la tabla de abajo pero es necesario detallar el procedimiento que se utiliza.
b) Si suponemos que los datos proceden de una distribución geométrica, ¿la proporción de cartones defectuosos envasados por dichas máquinas puede condiderarse del 10%?.
c) Calcula el nivel de significación de los datos para el contraste del apartado anterior por dos procedimientos distintos.

| Límite Inferior | Límite Superior | Frecuencia observada | Frecuencia esperada |
|---|---|---|---|
| - | 0.5 | 2 | 2.05 |
| 0.5 | 2.5 | 4 | 3.49 |
| 2.5 | 4.5 | 0 | 2.81 |
| 4.5 | 6.5 | 5 | 2.27 |
| 6.5 | 9.5 | 3 | 2.60 |
| 9.5 | 13.5 | 1 | 2.38 |
| 13.5 | 19.5 | 3 | 2.10 |
| 19.5 | - | 2 | 2.30 |

(1.5 puntos)

7.- De entre los individuos tratados con analgésicos no esteroideos, se tomaron 85 que habían padecido una hemorragia digestiva y 10295 que no la habían padecido. De entre los primeros 42 tenían menos de 60 años y de los segundos 7496. ¿Es la edad un factor de riesgo?.
Comenta los siguientes cuadros de la salida de SPSS para estos datos y explica cada uno de los resultados que en ella aparecen, destacando e interpretando los más adecuados para contestar a la pregunta planteada.

**Tabla de contingencia EDAD * ENFERMED**

Recuento

| | | ENFERMED | | |
|---|---|---|---|---|
| | | NO | SI | Total |
| EDAD | < 60 | 7496 | 42 | 7538 |
| | > 60 | 2799 | 43 | 2842 |
| Total | | 10295 | 85 | 10380 |

**Estimación de riesgo**

| | | Intervalo de confianza al 95% | |
|---|---|---|---|
| | Valor | Inferior | Superior |
| Razón de las ventajas para EDAD (< 60 / > 60 ) | 2.742 | 1.788 | 4.204 |
| Para la cohorte ENFERMED = NO | 1.010 | 1.005 | 1.015 |
| Para la cohorte ENFERMED = SI | .368 | .241 | .562 |
| N de casos válidos | 10380 | | |

**Estimación de riesgo**

| | | Intervalo de confianza al 95% | |
|---|---|---|---|
| | Valor | Inferior | Superior |
| Razón de las ventajas para ENFERMED (NO / SI ) | 2.742 | 1.788 | 4.204 |
| Para la cohorte EDAD = < 60 | 1.474 | 1.188 | 1.828 |
| Para la cohorte EDAD = > 60 | .537 | .435 | .665 |
| N de casos válidos | 10380 | | |

(1.25 puntos)

# Notas:

$$\mathrm{Var}_{H_0} W_s^* = \frac{mn(N+1)}{12} - \frac{mn\sum_{i=1}^{e}(d_i^3 - d_i)}{12N(N-1)}$$

$$\mathrm{Var}_{H_0} V_s^* = \frac{N'(N'+1)(2N'+1)}{24} - \frac{\sum_{i=1}^{e}(d_i^3 - d_i)}{48}$$