LEC/LADE/LECD/LADED

HOJA DE PROBLEMAS 4
CONTRASTES DE HIPÓTESIS Y DIAGNOSIS DEL MODELO

1.- El tiempo en minutos que dura un viaje en tren entre dos ciudades A y B, es una variable aleatoria normal con varianza 16. Hemos extraído una muestra aleatoria simple de 10 viajes, obteniendo los siguientes tiempos:

$$50 \quad 70 \quad 55 \quad 60 \quad 62 \quad 69 \quad 60 \quad 72 \quad 58 \quad 62$$

(a) Deducir la expresión de la desviación típica de la media muestral, y calcularla para esta muestra.

(b) ¿Entre qué valores se encontrará la verdadera media, μ, del tiempo que dura el viaje en tren de A a B, a un nivel de confianza al 95%?

(c) Se desea realizar el contraste

$$H_0 : \mu = 60$$
$$H_1 : \mu \neq 60$$

Utilizando el resultado obtenido en el apartado (b), ¿se rechaza la hipótesis nula con $\alpha = 0,05$? Razona la respuesta.

(d) Sin realizar ningún cálculo, ¿podríamos contestar al apartado (c) pero ahora para $\alpha = 0,01$? Razona la respuesta.

2.-Se toma una muestra del crecimiento de la oferta hotelera (número de camas), en términos porcentuales, de ocho países mediterráneos en el período 2000-2004. Se obtienen los siguientes datos:

$$4,6 \quad 0,1 \quad 3,4 \quad 9 \quad 7,3 \quad 0,8 \quad 1,6 \quad 4,6$$

(a) Especificando las hipótesis necesarias, determinar un intervalo de confianza al 90% para el valor medio y la desviación típica.

(b) Un informe anterior de Eurostat (Oficina de Estadística de la Unión Europea), preveía que el crecimiento medio en la zona no superaría el 3,5% (tómese como H_0). Razonar brevemente si esta previsión fue acertada al 95% de nivel de confianza.

(c) Para el mismo periodo se tomó una muestra del crecimiento de la oferta hotelera en países del Caribe; los datos obtenidos fueron los siguientes:

$$3,1 \quad 2,8 \quad 3,9 \quad 7,1 \quad 6,2 \quad 4,3 \quad 1,3$$

Especificando las hipótesis necesarias, determinar si es aceptable la afirmación de que el crecimiento medio de la oferta hotelera es igual en ambas zonas turísticas al 99% de nivel de confianza, suponiendo que la varianza poblacional es la misma en ambas zonas.

(d) Acotar el p-valor del contraste realizado en (c).

3.-Un grupo de científicos está estudiando los diferentes hábitos de alimentación entre dos colectivos de Madrid y Tenerife . Sospechan que estas diferencias pueden influir de forma significativa en el peso de los recién nacidos. Por ello se han tomado dos muestras de pesos de recién nacidos, una en Madrid y otra en Tenerife, obteniéndose los siguientes resultados:

	Tenerife	Madrid
Count	28	22
Average	3,13714	3,38364
Variance	0,157917	0,280043
Minimum	2,4	2,08
Maximum	4,03	4,7
Range	1,63	2,62
Stnd. skewness	0,0881308	-0,341812
Stnd. kurtosis	-0,254173	1,89641

a) Teniendo en cuenta estas dos muestras, y suponiendo normalidad en las poblaciones, determinar si existe suficiente evidencia al 5% de significación para afirmar que las varianzas poblacionales son distintas. Responder en base a la siguiente información que ofrece Statgraphics:

```
Standard deviation  0,397388           0,529191
Variance            0,157917           0,280043
Df                  27                 21

     Ratio of Variances = 0,563904

95,0% Confidence Intervals
     Standard deviation of Tenerife: [0,314183;0,5409]
     Standard deviation of Madrid: [0,407134;0,756248]
     Ratio of Variances: [0,241516;1,26131]

F-test to Compare Standard Deviations

   Null hypothesis: sigma1 = sigma2
   Alt. hypothesis: sigma1 NE sigma2
   F = 0,563904   P-value = 0,160884
```

b) ¿Hay evidencia estadística con un 95% de confianza, de que la media poblacional de los pesos de recién nacidos madrileños es mayor que la media poblacional de los pesos de los recién nacidos tinerfeños? Acotar el p-valor del contraste.

c) Sin hacer ningún cálculo, ¿cómo cambiaría la respuesta del apartado anterior si hubiéramos utilizado un nivel de confianza de un 90%? ¿Y si hubiera sido de un 99%?

d) Explicar el significado estadístico de nivel de confianza $(1-\alpha)100\%$.

4.- Sea X la variable "rentabilidad de cierto tipo de fondos de inversión tras una apreciación fuerte del marco con respecto al dolar". Se considera que la media de esta variable es 15. Un economista afirma que dicha rentabilidad media ha variado, por lo que lleva a cabo un estudio en las condiciones reseñadas anteriormente sobre una muestra de 9 fondos cuya media muestral resulta ser de 15.308 y cuya varianza muestral corregida (cuasivarianza) es 0.193.

a) Especificando las hipótesis necesarias, contrastar la afirmación del economista al 5%.

b) A partir del resultado de **a)**, razonar si el intervalo de confianza para la media (centrado en $\overline{x}$) al 95% contendrá o no al valor 15.

c) Acotar el p-valor. Si el contraste se hubiera realizado al 10%, ¿aceptaríamos la hipótesis de que la media de la rentabilidad es 15 tras una apreciación fuerte del marco con respecto al dolar?. Razonar brevemente la respuesta.

5.- La entrada en vigor de la Ley 28/2005 por la que se prohíbe fumar en lugares públicos ha suscitado opiniones muy encontradas. Con la aplicación de dicha Ley, un porcentaje de fumadores ha decidido dejar el tabaco y cierto organismo establece en el 20% este porcentaje.

(a) Deducir la distribución de la proporción muestral de fumadores cuya intención es dejar el tabaco (en caso de muestras grandes).

(b) ¿Es insesgado el estimador proporción muestral de fumadores con intención de dejar el tabaco? Razonar la respuesta.

(c) El Gobierno asegura que el porcentaje de fumadores que van a decidir dejar de fumar con la aplicación de esta Ley, va a ser al menos el 28%. Para contrastar dicha hipótesis, se encuesta a 800 fumadores de los que 150 manifiestan su intención de dejar de fumar desde la entrada en vigor de la Ley. ¿Hay evidencia estadística suficiente para rechazar la hipótesis del Gobierno?

6.-Un fabricante de pasta indica en el envoltorio que la cantidad de pasta que contiene cada paquete es de 500 gr. Se cree que la desviación típica del peso de los paquetes es de 3 gr. Se toma una muestra de 20 paquetes y se obtiene un peso medio de 498 gr con una desviación típica corregida de 4 gr. Especificando las hipótesis necesarias, responder:

(a) ¿Hay evidencia de que por término medio los paquetes contienen menos pasta de lo indicado? Según tu conclusión, ¿qué tipo de error podrías estar cometiendo y con qué probabilidad?

(b) ¿Es la desviación típica mayor de lo que se cree? Utiliza un nivel de significación del 5%.

(c) Calcular la potencia del contraste de la media del apartado (a) para un nivel de significación del 5% y un valor alternativo de la media de 497.

7.- En un área metropolitana se selecciona una muestra de 250 personas al azar y se les pregunta cuántas veces compraron un cierto producto durante el mes anterior.

Los resultados son los de la tabla adjunta.

Veces que compran	N^o de personas
0	85
1	80
2	45
3	20
4	15
5	5

Contrastar con un 95% de confianza la hipótesis de que las observaciones constituyen una muestra de una población que se distribuye como una Poisson.

8.-El gerente de un hospital quiere reducir el tiempo de espera de los pacientes en el servicio de urgencias. Para ello, realiza un estudio del tipo y número de pacientes que llega a este servicio cada hora. Concluye que los pacientes se pueden clasificar en dos grandes grupos:

- Grupo L: pacientes con dolencias leves o moderadas.
- Grupo G: pacientes con dolencias graves.

También concluye que el número, X, de pacientes del grupo L que llega en una hora sigue una distribución $Poisson(\lambda_L)$ y que el número, Y, de pacientes del grupo G que llega en una hora sigue una distribución $Poisson(\lambda_G)$.

Posteriormente, toma 100 observaciones de X e Y en horas seleccionadas en diferentes días. Los estadísticos resumen de estas muestras son:

```
Summary Statistics

                    X                   Y
----------------------------------------------------
Count               100                 100
Average             10.27               5.0
Variance            10.6637             4.60606
Standard deviation  3.26554             2.14617
Minimum             5.0                 1.0
Maximum             21.0                10.0
Range               16.0                9.0
Stnd. skewness      2.8468              0.561879
Stnd. kurtosis      1.56713             -1.15913
----------------------------------------------------
```

(a) Calcular un intervalo de confianza al 95% para el parámetro λ_L.

   (i) ¿Bajo qué hipótesis es válido el intervalo calculado?.

   (ii) ¿Puede aceptarse la hipótesis de que la media del número de pacientes de tipo L es igual a 11 con un nivel de significación $\alpha = 0.1$?. Justificar la respuesta.

(b) Contrastar, con un nivel 0.05, la hipótesis (nula) de que el número medio de pacientes del grupo G es menor o igual que 4. Escribir claramente las hipótesis del contraste, el estadístico de contraste y la región de rechazo del mismo.

(c) El gerente cree que la suma de los pacientes de tipo L y de tipo G que llegan en una hora, $Z = X+Y$, sigue una distribución $Poisson(\lambda_Z)$. Completar la siguiente salida de `Statgraphics` para contrastar la hipótesis $H_0 : Z \sim Poisson$ a un nivel, $\alpha = 0.05$. En el caso del p-valor es suficiente con acotar su valor.

```
Goodness-of-Fit Tests for Z

                              Chi-Square Test
---------------------------------------------------------------
               Lower         Upper      Observed      Expected
               Limit         Limit     Frequency     Frequency
---------------------------------------------------------------
         at or below           9.5             4          6.15
                 9.5          11.5            17         10.60
                11.5          13.5            18         17.04
                13.5          15.5            13         20.26
                15.5          17.5            21         18.51
                17.5          19.5             9         13.40
                19.5          21.5            14          7.88
above           21.5                           4          6.16
---------------------------------------------------------------
Chi-Square = ? with ? d.f.   P-Value = ?
```

9.- Se recogen medidas de polución atmosférica en 10 sitios de una ciudad:

$$3.3;\ 1.7;\ 3.7;\ 4.6;\ 2.3;\ 3.9;\ 4.3;\ 1.4;\ 1.6;\ 3.6$$

(a) Hallar un intervalo de confianza al 95% para la varianza poblacional, mencionando las hipótesis estadísticas que hay que asumir.

(b) La polución atmosférica media de estos 10 sitios medida un año antes fue 2.27. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha variado. Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo.

(c) Una organización ecologista sospecha que la polución atmosférica media está aumentando cada año. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha aumentado (tómese como H_0). Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo. Comentar el resultado de los contrastes realizados en (b) y (c).

10.- La llegada de aviones a cierto aeropuerto sigue una distribución Poisson (λ). Se realiza un estudio para determinar si es conveniente ampliar el aeropuerto. Para ello se analiza la distribución de la variable X: “ número de llegadas al aeropuerto cada hora ”. Tras observar una muestra aleatoria simple de 60 períodos de una hora, estos son los resultados obtenidos:

Valor de X	0	1	2	3	4	5	6	7	8	9
Frecuencia	1	2	13	7	10	16	8	2	0	1

Se pide:

a) Sabiendo que $V[X] = E[X] = \lambda$, determinar el estimador de máxima verosimilitud para la media de la distribución. Hallar una estimación puntual del parámetro basándonos en las observaciones obtenidas.

b) Teniendo en cuenta las anteriores estimaciones, contrastar "el número medio de llegadas a la hora es superior a 5 ", (tómese como hipótesis alternativa). Especificar las hipótesis asumidas.

c) La siguiente tabla nos ofrece los valores esperados que se hubieran obtenido si las observaciones siguiesen realmente una distribución de Poisson de parámetro 4:

Valor de X	Menos de 2	2	3	4	5	6	7 o más
Frecuencia esperada	5,49	8,79	11,72	11,72	9,37	6,25	6,64

Determinar si existe evidencia significativa (al 1%) de que la muestra sigue una distribución de Poisson

11.- En un servicio de urgencias resulta crucial el tiempo de espera hasta que un ciudadano es atendido. Se sospecha que la variable "Tiempo de espera" puede ajustarse mediante una distribución normal. Por ello, se toma una muestra aleatoria simple de 50 datos y se obtiene el siguiente resultado al realizar un test Chi-cuadrado de bondad de ajuste:

```
--------------------------------------------------------------------------
               Límite         Límite    Frecuencia    Frecuencia
             Inferior       Superior    Observada      Esperada
--------------------------------------------------------------------------
   menor o igual de          15,1013           8             7,14
           15,1013          16,717            4             7,14
           16,717           17,96            ___            7,14
           17,96            19,1196           9             7,14
           19,1196          20,3627           8             7,14
           20,3627          21,9783           9             7,14
      mayor que             21,9783           5             7,14
--------------------------------------------------------------------------
Chi-cuadrado = _______ vs Chi-cuadrado con _____ grados de libertad
```

a) Completar los datos que faltan en la tabla anterior.

b) Plantear el contraste de hipótesis y acotar el p-valor. ¿Hay evidencia estadística para rechazar la hipótesis de normalidad?

c) Para reforzar la decisión se realiza también el contraste de Kolmogorov-Smirnov obteniéndose $D_n = 0.0937677$. ¿Qué decisión tomaría con un nivel de significación del 5%? Usar la siguiente tabla:

Tabla Kolmogorov-Smirnov

Tamaño muestral n	Nivel de significación				
	0,20	0,15	0,10	0,05	0,01
>30	$\frac{0,736}{\sqrt{n}}$	$\frac{0,768}{\sqrt{n}}$	$\frac{0,805}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

12.- En cierto proceso industrial se considera la variable aleatoria X = coste unitario de producción (en Euros). Se sabe que esta variable aleatoria tiene una distribución normal de media μ y desviación típica $\sigma = 20$. Se toma una muestra aletoria de 25 observaciones obteniéndose una media muestral $\bar{x} = 120$.

a) A partir de esta información muestral, calcular un intervalo de confianza al 95% para el coste unitario medio de producción.

b) ¿Qué tamaño muestral será necesario tomar para que un intervalo, del 95% de confianza, garantice una precisión en la estimación, dada por una amplitud máxima de 10 Euros?. (Amplitud del intervalo = diferencia entre sus extremos).

c) Ante la sospecha de que la varianza de X está aumentando se calcula la cuasivarianza (o varianza corregida) muestral, s^2, de la muestra anterior y se obtiene un valor igual a 512.12. ¿Existe evidencia muestral al 99% para rechazar la hipótesis de que $\sigma = 20$?

13.- Los defensores de un nuevo molino de viento afirman que la potencia diaria que puede generar es una variable aleatoria (X) cuya media es al menos 800 Kw. Se asume que X es normal con desviación típica 120 Kw. Para realizar un contraste de la hipótesis mantenida por los defensores del nuevo molino, se toma una muestra aleatoria simple de tamaño 100 y se obtiene una media muestral igual a 776 Kw.

(a) Definir y calcular el p-valor del contraste.

(b) Definir y calcular la potencia del contraste tomando como hipótesis alternativa $\mu = 770$ y nivel de significación $\alpha = 0.05$.