Grados en Ingeniería Informática

Estadística

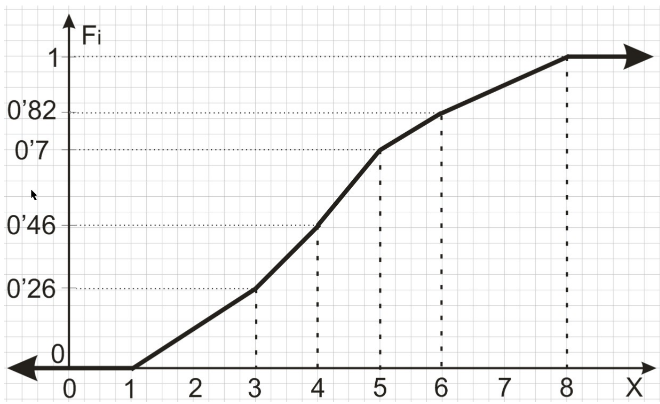
Examen **Convocatoria Diciembre 2015**

- A resolver en **2 horas y 15 minutos.**
- Dejar DNI encima de la mesa.
- **Apagar y guardar el MÓVIL**.

APELLIDOS, NOMBRE:
DNI: Especialidad: Grupo:

---

1. A partir del siguiente diagrama de frecuencias relativas acumuladas y conociendo que el tamaño de la población es 400, hallar la media, varianza y los cuartiles 2 y 3 de la población estudiada.



(1.25 Puntos)

2. Dados los puntos: (1,2), (2,2), (3,4), (4,5) y (5,9), se pide:
   Predecir el valor de y para $x = 3.25$, mediante el modelo: $y = 1 + a \cdot b^x$.

(1 Punto)

3. Dada la variable aleatoria ξ de función de densidad $f(x) = \begin{cases} Kx & x \in [0,2] \\ 0.3 & x \in [2,3] \\ 0 & x \notin [0,3] \end{cases}$, la variable ψ que sigue una distribución normal N(5,4) y la φ cuya función de densidad es $g(x) = \begin{cases} 0 & \psi \leq 3 \\ 1 & \psi > 3 \end{cases}$

   Hallar:

   (a) Media y varianza de ξ.
   (b) El cuartil 1 de ψ.
   (c) El valor "a" tal que $F_\xi(a) = P(\xi \leq a) = P(\varphi = 1)$.
   (d) Esparanza de la variable aleatoria $\Phi = \xi + 2\psi + 3\varphi$

(0.75+0.5+0.75+0.5=2.5 Puntos)

4. Se intenta determinar si existe interacción entre la roya y el moteado del peral (dos enfermedades por hongos). Para ello, se seleccionan 500 parcelas y se confecciona la siguiente tabla:

Roya \ Moteado	*Nada*	*Indicios*	*Atacado*
Nada	198	28	62
Indicios	39	6	12
Atacado	105	15	35

Estudiar al 5% de significación si puede admitirse que el ataque por una y otra enfermedad son independientes entre sí, o por el contrario, el ataque por alguna de ellas se ve favorecido, o repelido, por la presencia de la otra.

(0.75 Puntos)

5. Una organización de defensa del consumidor está analizando el contenido en el principio activo (CPA) de cierta píldora que debe seguir una normal de media 17.4 mg. y varianza 3 mg. Quiere conocer al 90% si el valor de la varianza es fiable y, de acuerdo con el resultado que obtengamos, ver si el peso medio ha disminuido. Para ello se toma una muestra de 25 pastillas que produce los siguientes resultados: $\sum_{i=1}^{25} x_i = 425$ y $\sum_{i=1}^{25} x_i^2 = 7279$.

   (a) Contrastar si la varianza ha variado.

   (b) Contrastar si la media ha disminuido supuesta que la varianza es 3.

   (c) Contrastar si la media ha disminuido supuesta la varianza desconocida.

   (d) Indicar de los contrastes realizados en los apartados (b) y (c), ¿cuál de ellos es más adecuado en este caso? Razone la respuesta.

(0.7+0.7+0.7+0.4=2.5 Puntos)

6. **PRÁCTICAS (Sólo instrucciones)**

   Dados los puntos: (1,2), (2,2), (3,4), (4,5) y (5,9), se pide:

   (a) Ajustar una recta $y = c + dx$.

   (b) Ajustar una función de la forma $y = 1 + a \cdot b^x$.

   (c) Predecir el valor dado por cada modelo para $x = 3.25$.

   (d) Indicar el modelo más adecuado mediante el cálculo de los parámetros que procedan.

(0.2+0.3+0.2+0.3=1 Punto)

7. **PRÁCTICAS (Sólo instrucciones)**

   El contenido en zinc puede determinar la procedencia de una muestra de barro. Un arqueólogo pretende determinar la procedencia de un tipo de vasijas encontradas en un yacimiento arqueológico. Toma muestras de cada vasija obteniendo los la proporción de zinc en ppm. (partes por millón).

   x={8.9699, 9.6203, 9.6129, 8.2467, 9.2151, 9.0587, 9.9781, 10.5182, 10.5368, 8.2482, 9.3397, 7.8416, 7.9583, 9.2421, 11.0279, 8.3572, 9.6808, 8.9883, 10.5461, 7.9867, 9.2878, 9.8909, 10.5267, 11.0413, 9.3497, 7.5198, 8.3889, 8.0186, 11.9765, 8.5359, 10.1178, 9.0268, 10.2808, 8.3628, 7.6234, 7.6000, 9.8163, 9.0442, 9.0226, 10.8964, 9.5882}

   (a) Contrastar que la media de estos cocientes es 8.5 al nivel de significación $\alpha = 0.01$.

   (b) Dar un intervalo de confianza al 99% para el contenido en zinc.

(0.5+0.5=1 Punto)

Soluciones:

Problema 1:

L_i	x_i	F_i	f_i	n_i	N_i	$n_i x_i$	$n_i x_i^2$
$1-3$	2.0	0.26	0.26	104	104	208	416
$3-4$	3.5	0.46	0.20	80	184	280	980
$4-5$	4.5	0.70	0.24	96	280	432	1944
$5-6$	5.5	0.82	0.12	48	328	264	1452
$6-8$	7.0	1.00	0.16	72	400	504	3528
				400		1688	8320

Media: $\bar{x} = \frac{1688}{400} = 4.22$, **Var** $= \frac{8320}{400} - 4.22^2 = 2.9916$

Cuartil 2: $Q_2 = \text{Mediana} = 4 + \frac{200-184}{96} 1 = 4.1667$, **Cuartil 3:** $Q_3 = 5 + \frac{300-280}{48} = 5.4167$

Problema 2:

$y = 1 + a \cdot b^x \Rightarrow y - 1 = a \cdot b^x \Rightarrow Y = Ln(y-1) = Ln(a) + xLn(b) = A + Bx$ con $A = Ln(a)$ y $B = Ln(b)$

x_i	y_i	$Y_i = Ln(y_i - 1)$	x_i^2	$x_i Y_i$
1	2	$Ln(1) = 0$	1	0
2	2	$Ln(1) = 0$	4	0
3	4	$Ln(3) = 1.0986$	9	3.2958
4	5	$Ln(4) = 1.3863$	16	5.5452
5	9	$Ln(8) = 2.0794$	25	10.3972
15		4.5643	55	19.2382

$$\Rightarrow \left. \begin{array}{r} 4.5643 = 5A + 15B \\ 19.2382 = 15A + 55B \end{array} \right\} \Rightarrow \begin{array}{r} A = -0.7507, \ a = e^{-0.7507} = 0.4720 \\ B = 0.5545, \ \ b = e^{0.5545} = 1.7411 \end{array}$$

Luego el ajuste realizado da $y = 1 + 0.4720\,(1.7411)^x$, que para x=3.25 da $y = 1 + 0.4720\,(1.7411)^{3.5} = 4.2875$

Problema 3:

a: Lo primero será hallar K:

$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 Kxdx + \int_2^3 0.3dx = 1, \Rightarrow \left[K\frac{x^2}{2}\right]_0^2 + 0.3(3-2) = 1 \Rightarrow 2K + 0.3 = 1 \Rightarrow K = 0.35$

$\mu = E(\xi) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^2 0.35x^2dx + \int_2^3 0.3xdx = \left[0.35\frac{x^3}{3}\right]_0^2 + \left[0.3\frac{x^2}{2}\right]_2^3 = 0.35\frac{8}{3} + 0.3\frac{5}{2} = 1.6833$

$m_2 == \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^2 0.35x^3dx + \int_2^3 0.3x^2dx = \left[0.35\frac{x^4}{4}\right]_0^2 + \left[0.3\frac{x^3}{3}\right]_2^3 = 0.35\frac{16}{4} + 0.3\frac{8}{3} = 3.3$

$V(\xi) = \sigma^2 = 3.3 - 1.6833^2 \approx 4.4664$

b: Buscamos $a = Q_1$ que debe cumplir: $F_\psi(a) = 0.25 = P(\psi \leq a) = P\left(\frac{\psi-5}{4} \leq \frac{a-5}{4}\right) = P\left(z \leq \frac{a-5}{4}\right) = 0.25$

$P\left(z \geq -\frac{a-5}{4}\right) = 0.25 \Rightarrow -\frac{a-5}{4} = 0.6745 \Rightarrow a = Q_1 = 5 - 0.6745(4) = 2.3020$

c: La función de densidad de φ es: $g(x) = \begin{cases} 0 & \psi \leq 3 \Rightarrow P(\psi \leq 3) = P(z \leq \frac{3-5}{4}) = 0.3085 \\ 1 & \psi > 3 \Rightarrow P(\psi > 3) = 1 - P(\psi \leq 3) = 0.6915 \end{cases}$

Es decir φ sigue una distribución de Bernouilli con $P(\varphi = 0) = 0.3085$ y $P(\varphi = 1) = 0.6915$.

Nos piden, por tanto, hallar a para que $P(\xi \leq a) = 0.6915$ que sabemos que está en el primer trozo [0,2] de la distribución de ξ pues es menor de 0.7 (el segundo trozo contiene el 30% de la población), así:

$\int_0^a 0.35xdx = 0.6915 \Rightarrow \left[0.35\frac{x^2}{2}\right]_0^a = 0.35\frac{a^2}{2} = 0.6915 \Rightarrow a = 1.9878$

d: $E(\Phi) = E(\xi) + 2E(\psi) + 3E(\varphi) = 1.6833 + 2(5) + 3(0.6915) \approx 13.7578$

Problema 4:

Nos han dado la tabla de frecuencias absolutas observadas $n_{i,j} = O_{i,j}$ y debemos construir la de frecuencias esperadas $e_{i,j}$ en el caso de independencia (Hipótesis nula H_0). Se construye mediante $e_{i,j} = \frac{n_{i,\cdot} n_{\cdot,j}}{N}$:

$O_{i,j}$	N	I	A	
N	198	28	62	288
I	39	6	12	57
A	105	15	35	155
	342	49	109	500

$e_{i,j}$	N	I	A
N	196.992	28.224	62.784
I	38.988	5.586	12.426
A	106.020	15.190	33.790

Calculando el estadístico experimental

$$E = \sum_i \sum_j \frac{O_{i,j}^2}{e_{i,j}} - N = \frac{198^2}{196.992} + \frac{28^2}{28.224} + \frac{62^2}{62.784} + \frac{39^2}{38.988} + \frac{6^2}{5.586} + \frac{12^2}{12.426} + \frac{105^2}{106.02} + \frac{15^2}{15.19} + \frac{35^2}{33.79} - 500 \approx 0.1175$$

Y lo comparamos con el teórico (en las tablas): $\chi^2_{0.05,4} = 9.488$, donde hemos entrado con 4 grados de libertad por ser una tabla 3x3 (gdl=(3-1)(3-1)=4).

Resulta que $E = 0.1175 < 9.488 = \chi^2_{0.05,4}$ luego aceptamos la hipótesis nula de independencia, es decir, concluimos que **no existe interacción**.

Problema 5:

a: Se trata de un contraste del valor de la varianza: $H_0 : \sigma^2 = 3$ contra $H_a : \sigma^2 \neq 3$

El estadístico de contraste es: $E = \frac{(n-1)s^2}{\sigma^2}$, sabemos que n=25, $\sigma^2 = 3$, $\alpha = 0.10$ necesitamos calcular s^2 a partir de los datos, pero:

$\mu = \frac{425}{25} = 17$, $s^2 = \left(\frac{7279}{25} - 17^2\right)\frac{25}{24} = 2.25$, así que: $E = \frac{(n-1)s^2}{\sigma^2} = \frac{24(2.25)}{3} = 18$, mientras que el intervalo de aceptación: $I = [\chi^2_{1-\frac{\alpha}{2},n-1}, \chi^2_{\frac{\alpha}{2},n-1}] = [\chi^2_{0.95,24}, \chi^2_{0.05,24}] = [13.848, 36.415]$.

Como $E = 18 \in [13.848, 36.415]$, concluimos que la varianza **no ha variado.**

b: Se trata ahora de un contraste unilateral de la media con varianza conocida.

$H_0 :\ \mu \geq 17.4$ contra $H_a :\ \mu < 17.4$

La región crítica es $E = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha = -z_{0.10} = -1.2816$

$E = \frac{17-17.4}{\frac{\sqrt{3}}{\sqrt{25}}} = -1.1547$ que es mayor que -1.2816, luego la media **no ha disminuido.**

c: Se trata ahora de un contraste unilateral de la media con varianza desconocida y muestra pequeña N=25¡30.

$H_0 :\ \mu \geq 17.4$ contra $H_a :\ \mu < 17.4$

La región crítica es $E = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} < -t_{\alpha,n-1} = -t_{0.10,24} = -1.3178$

$s = \sqrt{s^2} = \sqrt{2.25} = 1.5$ $E = \frac{17-17.4}{\frac{1.5}{5}} = -\frac{4}{3} \approx -1.3333$ que es menor que -1.3178, luego la media **ha disminuido.**

d: Los apartados b y c realizan el mismo contraste, pero con varianza conocida (Var=3) y con varianza desconocida dando resultados diferentes.

El correcto es el del apartado b, es decir, con varianza conocida e igual a 3 pues realizado el contraste del apartado a resultó que **la varianza no ha cambiado**, por lo que se admite que sigue valiendo 3.

Problema 6 MATLAB:

```
x=[1 2 3 4 5]; y=[2 2 4 5 9];
disp('a'), r=polyfit(x,y,1), c=r(2),d=r(1)
disp('b'), Y=log(y-1), p=polyfit(x,Y,1), B=p(1),A=p(2),a=exp(A),b=exp(B)
disp('c'), ya=c+d*3.25,yb=1+a*b^3.25
disp('d'), res1=y-(c+d*x);Vres1=var(res1,1), res2=1+a*b.^x; Vres2=var(res2,1)
```

Problema 7 MATLAB

```
x=[8.9699, 9.6203, 9.6129, 8.2467, 9.2151, 9.0587, 9.9781, 10.5182, 10.5368, 8.2482, 9.3397, ...
   7.8416, 7.9583, 9.2421, 11.0279, 8.3572, 9.6808, 8.9883, 10.5461, 7.9867, 9.2878, 9.8909, ...
 10.5267, 11.0413, 9.3497, 7.5198, 8.3889, 8.0186, 11.9765, 8.5359, 10.1178, 9.0268, 10.2808, ...
  8.3628, 7.6234, 7.6000, 9.8163, 9.0442, 9.0226, 10.8964, 9.5882]
alfa=0.01;
disp('a'), [H,P]=ttest(x,8.5,alfa,'both')
disp('b'), N=length(x), m=mean(x), s=sqrt(var(x)), T=tinv(1-alfa/2,N-1)
I=[m-T*s/sqrt(N), m+T*s/sqrt(N)]
disp('Al mismo resultado (intervalo) se llega mediante:')
[H,P,I]=ttest(x,m,alfa)
```