

Examen Final de Estadística I, 23 de junio de 2010.
Grados ADE, ADE-DER, ADE-INF, ECO, ECO-DER.

- NORMAS:** 1) Entregar cada problema en un cuadernillo distinto, aunque esté en blanco.
 2) Realizar los cálculos con al menos dos cifras decimales significativas.
 3) No se podrá abandonar el examen hasta transcurridos 30 minutos después de haber empezado.
 4) Una vez empezado el examen, no estará permitido salir del aula.

Problema 1 (10 puntos) La siguiente tabla contiene información sobre la edad y la antigüedad, ambas medidas en años, de una plantilla de 75 trabajadores de una empresa (atención: se han dejado en blanco las casillas que contienen un cero):

Edad	Antigüedad										
	1	2	3	4	5	7	9	10	12	15	20
[20,30)	3	6		4							
[30,40)		1	4	2	3						
[40,50)				1	4	13	2	3	1		
[50,60)							2	8	7	4	
[60,70]										2	5

- a) (1 punto) Indica el tipo de las variables **Edad** y **Antigüedad**.
- b) (1.5 puntos) Especifica la distribución marginal de la variable **Antigüedad**. ¿Cual es la antigüedad media para los trabajadores de esta empresa?
- c) (2 puntos) Compara la dispersión de las variables **Edad** y **Antigüedad**, ayudándote de la información siguiente:

```
numSummary(Datos[,c(Edad)], statistics=c("mean","sd","quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd    0%   25%   50%   75%   100%   n
Edad 44.891892 11.730498  20   38   46  53.0   67   74
```

- d) (1.5 puntos) Especifica la distribución de la variable **Antigüedad** para los trabajadores con edades comprendidas entre 40 y 50 años. ¿Qué puedes decir acerca de la asimetría de esta distribución?
- e) (2 puntos) A partir de la información siguiente, ¿qué puedes decir sobre la relación entre ambas variables? Explica en qué te basas.

```
Call:lm(formula=ANTIG\UEDAD ~ EDAD, data=Datos)

Residuals:
    Min       1Q   Median       3Q      Max
-4.8303  -1.5464  -0.3001   1.4345   6.0651

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.45427    1.12128   -7.54  1.10e-10 ***
edad          0.37315    0.02418  15.44  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.423 on 72 degrees of freedom
Multiple R-squared:  0.8679, Adjusted R-squared:  0.7647
F-statistic: 238.2 on 1 and 72 DF, p-value: < 2.2e-16
```

- f) (2 puntos) Utilizando la información del apartado e), encuentra el modelo lineal que relaciona la variable antigüedad en función de la edad. Interpreta los coeficientes y usa este modelo para predecir cuanto tiempo durará en la empresa una persona que tiene 35 años. Justifica si crees razonable realizar dicha predicción.

Problema 2 (10 puntos) En un departamento de una empresa aseguradora se procesan comunicaciones de clientes de seguros de hogar. Dichas comunicaciones están relacionadas con incidencias de dos tipos (robos, tipo R y daños, tipo D). En promedio en un mes se reciben 86 comunicaciones de los dos tipos, de las que 24 incluyen incidencias de tipo R y 68 incidencias de tipo D . Una vez estudiadas, se aprueba compensar (C) un 65% de las reclamaciones de tipo R y un 80% de las reclamaciones de tipo D .

Partiendo de los datos anteriores, se pide que contestes las siguientes preguntas:

- a) (3 puntos) ¿Cuál es la probabilidad de que una reclamación incluya incidencias de los dos tipos, R y D ?
- b) (3 puntos) ¿Cuál es la probabilidad de que se apruebe compensar una comunicación cualquiera?
- c) (4 puntos) Si te envían para revisar una comunicación para la que se ha aprobado compensación (C), ¿cuál es la probabilidad de que sea de tipo D ?

Problema 3 (10 puntos) Según un estudio realizado en cierta entidad bancaria durante el primer trimestre de 2010, en cada una de sus sucursales el número de cheques sin fondo que se reciben sigue una distribución de Poisson, con una media de 10 cheques sin fondo al día. Se elige una muestra de 200 sucursales mediante muestreo aleatorio simple y se registra el número de cheques sin fondo recibidos en cada una de ellas en un determinado día.

- a) (4 puntos) ¿Cuál es la probabilidad de que el promedio de cheques sin fondo recibidos en las 200 sucursales sea mayor que 12?
- b) (3 puntos) ¿Cuál es la probabilidad de que en una de las sucursales el número de cheques sin fondo recibidos un determinado día sea menor que 3?
- c) (3 puntos) Suponiendo ahora que desconocemos el número medio de cheques sin fondo que llegan al día a una sucursal, es decir, sólo sabemos que la distribución del número de cheques sin fondo por día y por sucursal es Poisson de parámetro λ , propón un estimador insesgado para λ , basándote en una muestra aleatoria simple de tamaño 200 y razonando la respuesta.