Prácticas de Probabilidad y Estadística

Métodos de Ayuda a la Decisión Médica

1 Teorema de Bayes.

Un médico aplica un test a 10 alumnos de un colegio para detectar una enfermedad cuya incidencia sobre una población de niños es del 10%. La sensibilidad del test es del 80% y la especificidad del 75%.

- ¿Cual es la probabilidad de que exactamente a cuatro personas le de un resultado positivo?
- Si en la muestra hay cuatro personas a las que el test le da positivo, ¿cuál es la probabilidad de que entre estas, exactamente dos estén sanas?
- Calcular la probabilidad de que el test suministre un resultado incorrecto para dos personas.
- Calcular la probabilidad de que el resultado sea correcto para más de 7 personas.

2 Representaciones estadísticas en MATLAB.

Como introducción al lenguaje de programación de MATLAB generaremos datos siguiendo una distribución normal $N(\mu, \sigma^2)$ de media *mu* y desviación estándar *sigma*. Calcularemos los estadísticos más habituales, representaremos la distribución de probabilidad y generaremos gráficos de probabilidad para comprobar el tipo de distribución [1].

```
clear all
close all
%inicialización
rand('seed',sum(100*clock));
randn('seed',sum(100*clock));
%generación de datos
mu=20;
sigma=1;
N=300;
y=normrnd(mu,sigma,N,1);
%estadísticos básico
mN=mean(y)
sN=std(y)
```

[1] Utilizar el comando help para ayuda sobre una funcion, ej. help normrnd. La mayoría de funciones que se necesitan están en la librería de estadística, help stats

```
skwN=skewness(y)
kurN=kurtosis(y)
pN = prctile(y,[2.5 25 50 75 97.5])
%representaciones
figure(1)
subplot(321)
stem(y,'.')
xlabel('muestras')
ylabel('valor')
subplot(322)
divisiones=20;
hist(y,divisiones)
title('histograma')
subplot(323)
boxplot(y)
subplot(324)
cdfplot(y)
subplot(325)
qqplot(y)
subplot(326)
probplot('normal',y)
```

3 Distribuciones de probabilidad.

En la sala de espera de pediatría de un hospital la probabilidad p de que un niño tenga gripe es del 5%. Comprueba experimentalmente que el número de niños con gripe ng sigue aproximadamente una distribución normal $ng \rightsquigarrow N(\mu = np, \sigma^2 = np(1-p))$ siendo $n = 50$ el número de niños que acude cada día a consulta. Calcula el valor medio y la desviación estándar muestral.

```
clear all
close all
%inicialización
rand('seed',sum(100*clock));
randn('seed',sum(100*clock));
% generar repetidos experimentos de una distribucion binomial
n=100;
p=0.05;
N=300;
for i=1:N
    ng(i)=binornd(n,p);
end
% valores estadísticos muestrales
mu_muestral=mean(ng)
sigma_muestral=std(ng)
% distribucion normal teórica
mu=p*n;
sigma=sqrt(n*p*(1-p));
x=linspace(mu-4*sigma,mu+4*sigma,100);
```

```
pdf_normal=normpdf(x,mu,sigma);
% distribucion muestral
ngH=hist(ng,x);
pdf_muestral=hist(ng,x)./sum(ngH);
%representacion
figure(1)
plot(x,pdf_normal);
hold on
bar(x,pdf_muestral);
title('Distribucion muestral');
xlabel('Valor'); ylabel('PDF');
figure(2)
normplot(ng)
```

4 Contraste de Hipótesis.

En unas pruebas de ensayo se quiere determinar si un determinado fármaco X aumenta, en promedio, la concentración plasmática de la hemoglobina Cp. Consultando la bibliografía se encuentra que el valor habitual es de unos 14 gr/dL con una desviación estándar de 1.5 gr/dL. Se plantean las siguientes hipótesis:

1. H0: $Cp = 14gr/dl$
2. H1: $Cp > 14gr/dl$

El ensayo consiste en inyectar el fármaco a 30 personas. Experimentalmente se obtiene que, en promedio, la concentración 14.7 gr/dL. ¿Qué conclusión se obtiene y para qué nivel de confianza?

```
clear all
close all
%inicialización
rand('seed',sum(100*clock));
randn('seed',sum(100*clock));
% hipoteis nula
mu = 14; % Valor medio
sig = 1.5; % Desviación conocida de la distribución
N = 30;  % Número de muestras
alpha = 0.05; % Nivel de significancia
conf = 1-alpha; % Nivel de confianza (probabilidad)
limite = norminv(conf, mu, sig/sqrt(N))
%intervalos de confianza asimétrico para 1-alpha
z=norminv(1-alpha,0,1);
intervalo_superior=mu+z*sig/sqrt(N)
%representación de la distribución de la media muestral
x = [linspace(10,limite), linspace(limite,20)]; % Eje x
y = normpdf(x,mu,sig/sqrt(N));
plot(x,y);
xhi = [limite, x(x>=limite)];
yhi = [0, y(x>=limite)];
```


```
patch(xhi,yhi,'b'); %Sombreamos la zona azul
title('Distribucion muestral');
xlabel('Valor'); ylabel('PDF');
```

5 Análisis de supervivencia.

Calcular la función de supervivencia y la tasa de riesgo acumulada de una población generada con una distribución de Weibull de parámetros $\lambda = 15000$ y $k = 3$ cuando se fija un tiempo de estudio de 14000.

```
% generación de los datos del estudio
rand('state',1);
t_vida = wblrnd(15000,3,90,1)
t_estudio = 14000;
t_fin = sort(min(t_estudio,t_vida));
fallecidos = t_fin(t_fin<t_estudio)
sobreviven = t_fin(t_fin==t_estudio)
censurados = t_fin >= t_estudio
nfallecidos= length(fallecidos)
nsobreviven = length(sobreviven)
ncensurados= length(censurados)
%representaciones
subplot(1,2,1);
[empF,x,empFlo,empFup] = ecdf(t_fin,'censoring',censurados,'function','survivor');
stairs(x,empF);
hold on;
stairs(x,empFlo,':');
stairs(x,empFup,':');
hold off
xlabel('t estudio');
ylabel('Proporcion de supervivencia');
title('Supervivencia P(T>=t)')

subplot(1,2,2);
[empF,x,empFlo,empFup] = ecdf(t_fin,'censoring',censurados,'function','cumulative hazard')
stairs(x,empF);
ylabel('Proporcion de fallos');
title('Riesgo acumulado')
xlabel('t estudio');
```