

Práctica 5. Modelo Normal

Una cooperativa decide invertir en una calibradora automática de naranjas. Existen en el mercado diversos tipos de modelos que la cooperativa clasifica según el rango de diámetro D de naranjas que pueden recoger :

- Tipo 1 : $6\text{cm} \leq D \leq 8\text{cm}$
- Tipo 2 : $6\text{cm} \leq D \leq 10\text{cm}$
- Tipo 3 : $4\text{cm} \leq D \leq 10\text{cm}$

El precio de los modelos es mayor si el rango de diámetros posibles es mayor. La cooperativa quiere ser asesorada sobre qué tipo de calibradora comprar. El primer paso consiste en describir la situación de acuerdo con las nociones teóricas que hemos visto en clase: podemos introducir el experimento aleatorio "se escoge al azar una naranja en la producción de las fincas de los miembros", el espacio muestral es por lo tanto toda la producción de las fincas. Se define la variable aleatoria continua D como el diámetro de la naranja escogida. Lo que quiere saber la cooperativa es qué proporción de naranjas de la producción tiene un diámetro comprendido entre 6 y 8cm, entre 6 y 10cm, y entre 4 y 10cm. Por lo tanto lo que necesita saber son las siguientes probabilidades: $\Pr(6 \leq D \leq 8)$; $\Pr(6 \leq D \leq 10)$ y $\Pr(4 \leq D \leq 10)$.

La cooperativa realiza un estudio preliminar y coge una muestra de 1000 naranjas en el momento de la recogida en las fincas de los miembros. Eso se puede traducir de la manera siguiente : puesto que es imposible para la cooperativa el medir el diámetro de todas las naranjas de la producción, se contenta con realizar el experimento aleatorio 1000 veces y dispone de los valores de D para esas 1000 realizaciones del experimento. Su esperanza es que las conclusiones que se puedan sacar a partir de la muestra sigan válidas para toda la producción.

El objetivo de esa práctica es doble :

1. Realizar el estudio descriptivo de la muestra recogida.
2. Proponer un modelo para la distribución de la v.a. D en toda la producción, e investigar la validez de ese modelo a partir de los datos de la muestra.

Finalmente deberemos concluir la práctica asesorando a la cooperativa.

Resolución

1. Estudio descriptivo de la muestra recogida

a.- En primer lugar debemos cargar en nuestro hoja de cálculo de Statistix los datos de la muestra. Estos se encuentran en un fichero ASCII llamado:

diam.dat

ubicado en nuestro directorio de prácticas */prácticas/estadística/datos/*.

Antes de importar los datos desde el fichero **diam.dat** puede ser conveniente visualizar el fichero para asegurarse de que los datos están bien presentados. Para ellos seleccionamos:

File -> View Ascii File

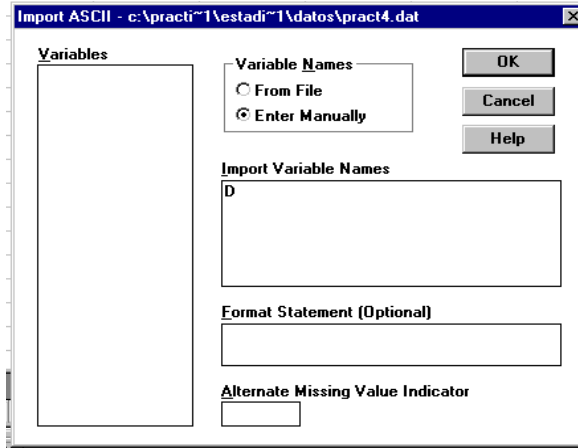
y en el cuadro de diálogo que aparece, nos movemos en las carpetas hasta llegar a */prácticas/estadística/datos/*, seleccionamos **diam.dat** y aceptamos. Al visualizar el fichero **diam.dat**, podemos comprobar en particular que sólo contiene datos.

Podemos ahora pasar a importar los datos desde el fichero **diam.dat**, para ello seleccionamos

File -> Import

Aparece un cuadro de diálogo y recorriendo la estructura de carpetas del disco duro, llegamos hasta la carpeta */prácticas/estadística/datos/*. Después de haber seleccionado *diam.dat*, pinchamos "Aceptar"

En el cuadro de diálogo, al contener el fichero *diam.dat* sólo los datos y no el nombre de la variable, debemos activar la opción "Enter manually" y especificar en el cuadro "Import variable names" el nombre de la variable a la que queremos asignar los datos importados, por ejemplo *D*.



b.- Establecer la tabla de frecuencias para la variable *D*, utilizando como límite inferior de las clases 3cm, como límite superior 11cm y como amplitud de clase 0.5cm. **Representar el histograma correspondiente.**



c.- ¿Cuál es el porcentaje de naranjas en la muestra cuyo diámetro está comprendido entre 6cm y 8cm?

¿entre 6cm y 10cm?

¿entre 4cm y 10cm?

d.- Basándose en los datos de la muestra, ¿qué tipo de modelo de recogedoras aconsejarías a la cooperativa?

2. Modelo para la distribución de la variable D

En este apartado estamos interesados, basándonos en la información recogida en la muestra, en proponer para la distribución de la v.a. D un modelo particular de distribución de los que hemos introducido en clase, es decir, para el comportamiento del diámetro para toda la producción.

a.- Calcular, para la muestra dada, la media y la desviación típica.

b.- Representar el histograma de las frecuencias absolutas, junto con la curva normal¹.



c.- Basándote en la muestra, ¿qué modelo de distribución continua te parece más adecuado para D?

¿Con qué media y qué varianza?

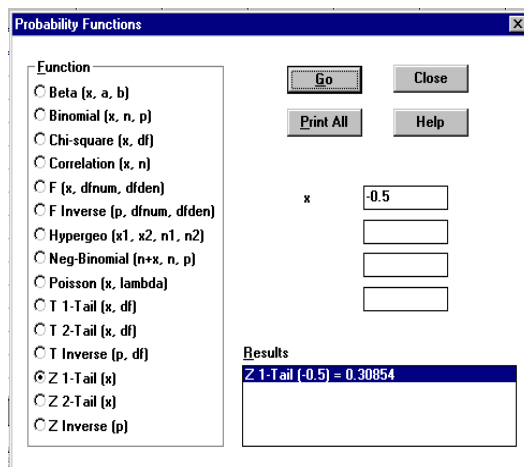
d.- Suponiendo que D sigue una distribución normal $N(8;1)$, vamos a calcular la probabilidad $P(7 \leq D < 7,5)$:

Para ello, empezamos por considerar la variable tipificada:

$$\Pr(7 \leq D < 7,5) = \Pr (\text{----} \leq Z < \text{----}) = \Phi(- 0,5) - \Phi (- 1) =$$

¹ Notar que, al activar la opción "Display Normal Curve", Statistix representa una densidad normal multiplicada por $N * h$, (N: tamaño muestral, h: amplitud de clase), de media y desviación típica calculadas a partir de la muestra

En el menú *Statistics* -> *Probability functions*, la opción *Z1 Tail (x)* nos permite calcular $\Pr(Z \leq x)$ si $x < 0$ y $\Pr(Z > x)$; si $x \geq 0$:



También podéis utilizar la calculadora estadística NCSSCALC.



El resultado es $P(7 \leq D < 7,5) = 0,30854 - 0,15866 = 0,15$.

Completar la tabla siguiente

Clases	muestra	Población	Clases	muestra	Población
<i>diámetro (cm)</i>	<i>Frec.rel.</i>	<i>Pr ($\leq D <$)</i>	<i>diámetro (cm)</i>	<i>Frec.rel.</i>	<i>Pr ($\leq D <$)</i>
3,0 $\leq D < 3,5$		0.0000031	7,5 $\leq D < 8,0$		0.1915
3,5 $\leq D < 4,0$		0.0000283	8,0 $\leq D < 8,5$		
4,0 $\leq D < 4,5$			8,5 $\leq D < 9,0$		
4,5 $\leq D < 5,0$			9,0 $\leq D < 9,5$		
5,0 $\leq D < 5,5$			9,5 $\leq D < 10,0$		
5,5 $\leq D < 6,0$		0.0165	10,0 $\leq D < 10,5$		
6,0 $\leq D < 6,5$		0.0441	10,5 $\leq D < 11,0$		
6,5 $\leq D < 7,0$					
7,0 $\leq D < 7,5$		0.1500			

e.- Utilizando el apartado anterior, ¿te parece que la muestra induce a pensar que la v.a D, diámetro de las naranjas en toda la producción, sigue una distribución $N(8; 1)$?

f.- En la producción entera, ¿qué porcentaje de naranjas tienen un diámetro comprendido entre 6cm y 8cm?

¿y entre 6cm y 10cm?

¿y entre 4cm y 10cm?

g.- ¿Qué tipo de modelo aconsejarías a la cooperativa?