Práctica 4

TEOREMA CENTRAL DEL LÍMITE

Objetivos:

En esta práctica utilizaremos el paquete SPSS para ilustrar el Teorema Central del Límite. Además calcularemos el área de la cola de algunas funciones de distribución y las representaremos gráficamente.

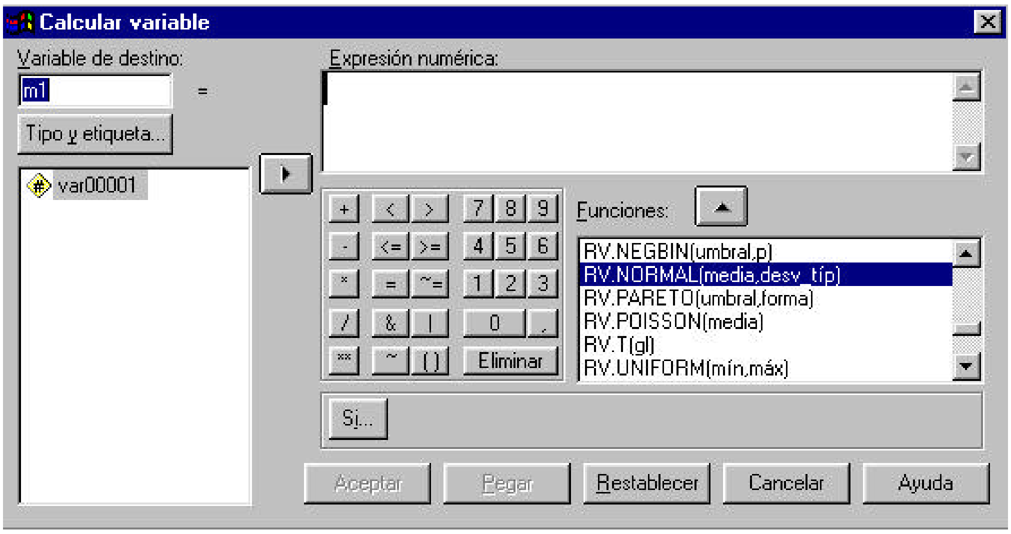
Índice:

1. Generación de muestras aleatorias

En este apartado consideraremos la generación de una muestra de una población con distribución conocida pero de la que no disponemos de datos. Por ejemplo, de una población con distribución Normal de media 20 y desviación típica 5 vamos a generar 100 datos.

En primer lugar, necesitamos crear un nuevo banco de datos (`Archivo/Nuevo/Datos`), al que podemos llamar *Normal*. SPSS genera los datos de una determinada distribución y los sitúa en una columna con la longitud que tenga el banco de datos; es decir, el número de filas con datos que aparezcan en el *Editor de datos*. Dado que acabamos de crear el fichero, no contendrá ningún dato, por lo que la primera operación a realizar será el rellenar la primera columna con datos. Para ello basta con que nos situemos en la casilla correspondiente a la columna 1, fila 100 (`Datos/Ir a Caso`) e insertemos un número cualquiera.

Seleccionamos ***`Transformar/Calcular`*** y nos aparece la ventana de calcular variables. En el campo ***`Variable de destino`***, escribimos **m1** como nombre de la variable que vamos a crear. De la lista de funciones que nos ofrece el SPSS, seleccionamos ***`RV.NORMAL(media,desv_típ)`*** y , con el puntero la situamos en el campo *Expresión numérica*.



Vemos, entonces que en lugar de **media** y **desv_típ** aparecen unos interrogantes, que sustituiremos por los valores de la media y la desviación típica de la distribución Normal de la que pretendemos simular una muestra, en este caso, m=20, s=5:

Se activa, entonces, el botón **Aceptar**, y al pulsarlo, el SPSS genera una muestra que añade en la primera columna libre del editor de datos.

La siguiente tabla muestra algunas de las funciones que proporcionan un valor aleatorio de una distribución determinada. Los argumentos son los parámetros de la distribución. Podéis consultar la ayuda del SPSS para obtener una lista exhaustiva de todas las funciones disponibles.

Expresión	Descripción
`RV.BINOM(n,prob)`	Devuelve un valor aleatorio de la distribución binomial, con el número de intentos y el parámetro de probabilidad especificados.
`RV.EXP(forma)`	Devuelve un valor aleatorio de una distribución exponencial, con el parámetro de forma especificado
`RV.GEOM(prob)`	Devuelve un valor aleatorio de una distribución geométrica, con el parámetro de probabilidad especificado.
`RV.LNORMAL(a,b)`	Devuelve un valor aleatorio de la distribución log-normal, con los parámetros especificados.
`RV.NORMAL(media,d_típ)`	Devuelve un valor aleatorio de la distribución normal, con la media y la desviación típica especificadas.
`RV.T(gl)`	Devuelve un valor aleatorio de la distribución t de Student, con los grados de libertad gl especificados.
`RV.UNIFORM(mín,máx)`	Devuelve un valor aleatorio de la distribución uniforme, con el mínimo y el máximo especificados

Ejercicio 1:

- Generar muestras de tamaño 100 para las distribuciones siguientes:
  Normal(0,1), Normal(2,4), t-Student(10), Binomial(8,0.3) y Binomial(10,0.1)

2. Área de la Cola

En este apartado vamos a ver cómo calcular el área de la cola de una distribución conocida. En concreto consideramos la expresión *P(x £ a)* siendo *x* una variable aleatoria y *a* un valor determinado.

Utilizamos el mismo procedimiento que en apartado anterior. Seleccionamos `Transformar/Calcular` y nos aparece la ventana de calcular variables. En el campo *Variable de destino*, escribimos el nombre de la variable que vamos a crear . De la lista de funciones que nos ofrece el SPSS, consideramos las que comienzan por `cdf`, escogemos la que queramos y, con el puntero la situamos en el campo *`Expresión numérica`*. En la tabla siguiente se muestran algunas de las funciones que podemos utilizar:

Expresión	Descripción
`CDF.BINOM(cant,n,prob)`	Devuelve la probabilidad acumulada de que el número de éxitos en los n intentos, con una probabilidad de éxito `prob` para cada uno, sea menor o igual que la cantidad `cant`

CDF.GEOM(cant,prob)	Devuelve la probabilidad acumulada de que el número de intentos para obtener un éxito, cuando la probabilidad de éxito es la dada por `prob`, sea menor o igual que la cantidad `cant`
CDFNORM(valorz)	Devuelve la probabilidad de que una variable aleatoria normal, con media 0 y desviación típica 1, sea menor que el valorz, el cual debe ser un valor numérico
CDF.NORMAL(cant,med, dev_típ)	Devuelve la probabilidad acumulada de que un valor de la distribución normal, con la media y la desviación típica especificadas, sea menor que la cantidad cant.
CDF.T(cant,gl)	Devuelve la probabilidad acumulada de que un valor de la distribución t de Student, con los grados de libertad gl especificados, sea menor que la cantidad cant.
CDF.UNIFORM(cant,mín, máx)	Devuelve la probabilidad acumulada de que un valor de la distribución uniforme, con el mínimo y máximo especificados, sea menor que la cantidad cant.

Por ejemplo para obtener el valor de *P(x £ 3)* para una variable *x* con distribución Normal(1,2), seleccionaremos la función ***CDF.NORMAL(cant,media,desv_típ)*** y reemplazaremos el primer campo por 3, el segundo por 1 y el tercero por 2.

Es importante notar que el valor de *P(x £ a)* coincide con el de *P(x < a)* si la función de distribución es continua como en el caso de la Normal o la t de Student, pero no coincide si es discreta como en el caso de la Binomial.

De manera similar podemos realizar el cálculo inverso al planteado; esto es, dado el valor del área de la cola *p*, obtener el valor de *a* tal que:

P(x £ a)=p

Para calcular el percentil 95 de una distribución Normal(1,2), seleccionamos *Transformar/Calcular* y en el campo ***Expresión numérica*** introducimos la expresión ***IDF.NORMAL(0.95,1,2)***. Notar que la función utilizada es similar a la anteriormente descrita salvo que comienza por ***IDF*** en lugar de ***CDF*** y que el primer argumento es *p* en lugar de *a*. Para obtener una lista de las funciones inversas disponibles en SPSS consultar la Ayuda en la ventana de *Calcular Variable.*

Ejercicio 2:

- Obtener el valor de *P(x £ 3), P(x £ 5)* y *P(x = 9)* para una distribución Binomial de parámetros *n* =20 y *p* =0.3.
- Obtener el valor de P(x ≤ -1), P(x ≤ 1), P(x≥2) y P(x >2) para una distribución Normal de parámetros *media* =2 y *desviación típica* =3.
- Obtener los percentiles 90, 5 y 50 de una distribución Normal(0,3).

3. Función de Densidad

SPSS no permite obtener directamente una función de densidad. En este apartado aplicaremos el cálculo visto en el apartado anterior para obtener una representación gráfica de una función de densidad.

Si, por ejemplo, deseamos obtener la representación de la función de densidad de una Normal(2,3) en el intervalo [-10, 10], generaremos primero una serie de valores en dicho intervalo. Veamos como generar 200 valores comenzando por el –9.9 y con incrementos de 0.1.

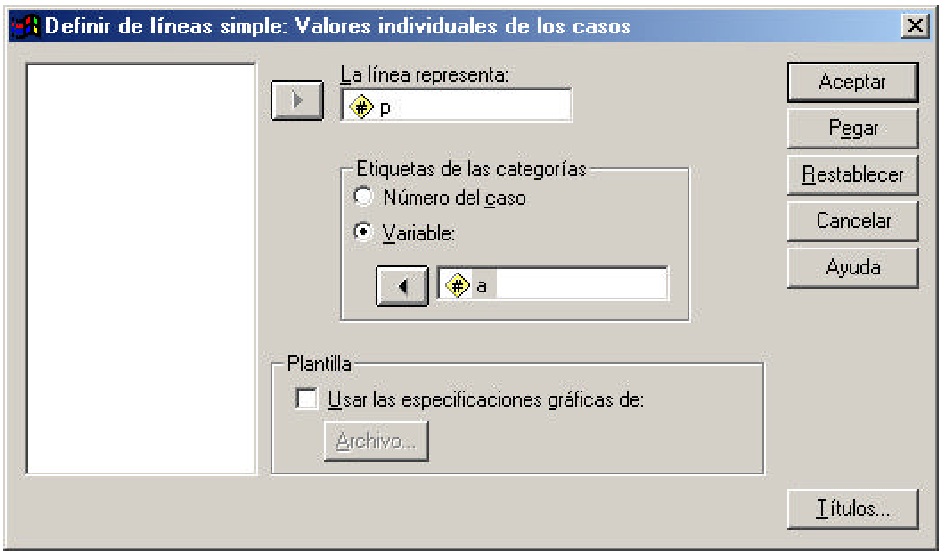
Vamos a trabajar con 200 casos, así que introduciremos un número en la fila 200 para dimensionar el editor de datos. Seleccionamos `Transformar/Calcular` y aparece la ventana de calcular variables. En el campo `Expresión numérica` vamos a introducir una expresión para obtener los valores de *a* para los que queremos calcular la función de densidad. La variable `$casenum` proporciona para cada caso el número correspondiente; esto es, en la fila primera el 1.00 en la segunda el 2.00 y así sucesivamente. Para obtener los valores deseados en el rango [-10, 10] introduciremos la expresión:

$$\frac{\$casenum}{10} - 10$$

Obtenemos una variable `a` con los valores –9.90, -9.80, ..., 9.90, 10. Entramos de nuevo en `Transformar/Calcular` e introducimos la siguiente expresión, donde `lag(a)` proporciona el valor de `a` en la fila anterior y el cociente sobre 0.1 nos da el promedio sobre la longitud del intervalo:

```
(CDF.NORMAL(a,2,3) - CDF.NORMAL(lag(a),2,3)) / 0.1
```

Almacenamos el resultado en la variable *p*. Finalmente, para obtener la representación gráfica seleccionamos `Gráficos/Líneas` y la opción `Valores individuales de los casos`. Introducimos las variables *a* y *p* tal y como muestra la figura siguiente.



Ejercicio 3:

- Obtener la representación gráfica de la función de densidad de una distribución Normal(1,2).
- Obtener la representación gráfica de una distribución Binomial de parámetros *n*=20 y *p* =0.3 realizando los siguientes pasos:
  1. Crear la variable a con los números enteros del 0 al 20 mediante
     `$casenum – 1`
  2. Crear la variable p con las probabilidades de x = a mediante
     `CDF.BIN(a,20,0.3) - CDF.BIN(lag(a),20,0.3)`
  3. Obtener la representación gráfica mediante `Gráficos/Barras`

4. Teorema Central del Límite

El objetivo de este apartado es ilustrar el teorema central del límite y la aproximación normal a la Binomial. Para ello seguiremos los siguientes pasos:

1. Generaremos 500 muestras de tamaño *n'* = 10.
2. Calcularemos la media de cada muestra y consideraremos la distribución muestral de las 500 medias obtenidas.
3. Comprobaremos que la distribución de las medias se asemeja a una normal cuya media es la media de la población original y cuya desviación típica es la desviación típica de la población original dividida por la raíz cuadrada del tamaño de la muestra.

4.1. Generación de las Muestras

Vamos a considerar una distribución Binomial con parámetros n = 20 y p = 0.4. En lugar de generar las 500 muestras de tamaño 10, vamos a generar 10 muestras de tamaño 500 y luego consideraremos las filas (en lugar de las columnas) obtenidas.

Siguiendo las instrucciones del apartado 1, genera 10 muestras (**m1** a **m10**) de tamaño 500 de la distribución *Bi(20,0.4)*. La media y desviación típica de esta distribución son, respectivamente:

$$np = 8 \text{ y } \sqrt{np(1-p)} = 2.1909.$$

Una vez generadas, tendremos en el editor de datos 10 columnas que contienen un total de 5000 valores aleatorios de la distribución Binomial utilizada. Si los consideramos por columnas, tenemos 10 muestras de tamaño 500. Sin embargo los vamos a considerar por filas, por lo que tenemos 500 muestras de tamaño 10 de esa misma distribución.

4.2 Cálculo de las Medias

Creamos una nueva variable, con nombre **Medias,** que calcule la media aritmética de las 10 variables que hemos creado antes (**m1** a **m10**). Para ello seleccionamos

`Transformar/Calcular`, llamamos Medias a la `Variable destino` e introducimos en `Expresión Numérica` la siguiente expresión:

```
(m1+m2+m3+m4+m5+m6+m7+m8+m9+m10) / 10
```

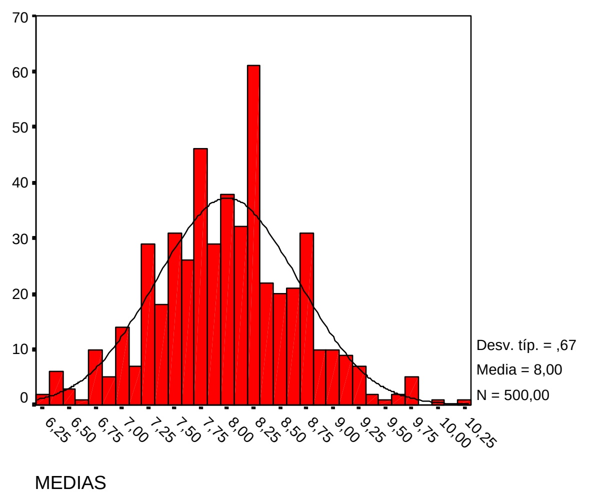
o alternativamente:

```
mean(m1, m2, m3, m4, m5, m6, m7, m8, m9, m10)
```

De esta forma obtenemos una variable Medias que contiene 500 valores, cada uno de los cuales es la media muestral de muestras de tamaño 10 de una población distribuida según la función Bi(20, 0.4).

4.3 Visualización de la Normalidad

Con el objetivo de visualizar la normalidad vamos a representar el histograma de la variable Medias. Para ello seleccionamos `Gráficos/Histograma` con la opción `Mostrar curva Normal`. Obtenemos un gráfico como el siguiente:



Podemos comprobar cómo, aproximadamente, el histograma se ajusta a la curva normal. Podemos ver cómo la media y desviación típica obtenidas son muy similares a las teóricas que indica el teorema central del límite:

$$\text{media} = np = 8 \qquad \text{desviación típica} = \frac{\sqrt{np(1-p)}}{\sqrt{n'}} = 0.7$$

Notar que hemos generado datos aleatorios por lo que los datos que tienes no coinciden exactamente con los de otro compañero y, por lo tanto, las gráficas tampoco serán exactamente iguales. Del mismo modo encontrarás algunas diferencias entre la gráfica que has obtenido y la que aparece en este texto.

5. Ejercicios

1. Repite los pasos de la sección 4 sobre 1000 muestras de tamaño 20 de una distribución Binomial(10, 0.8).
2. Análogamente para una Binomial(20, 0.1).
3. Calcular el área de la cola por debajo de *x* =0.5, 1, 2 y –0.75 de la t-student con 25 grados de libertad. ¿Tiene sentido calcular el área para un valor de *x* negativo?
4. Calcular los percentiles 3, 10 y 30 de una t de Student con 15 grados de libertad.
5. Calcular P(-1 ≤ x ≤ 1.5) para la t de Student con 25 grados de libertad.
6. Calcular P(3 ≤ x ≤ 7) para la Binomial(15, 0.3)
7. Obtener la representación gráfica de una distribución Binomial(100,0.8) y de una función de densidad Normal(10,4).
8. Generar una muestra aleatoria de tamaño 10 de una distribución Normal(0,1). Construir el histograma correspondiente con la curva normal superpuesta. ¿Se ajusta el histograma a la curva? Repetir para una muestra de tamaño 1000 y comentar los resultados.
9. Generar dos muestras de tamaño 10 de una distribución Normal(0,1), calcular y comparar sus medias y desviaciones típicas. Repetir para dos muestras de tamaño 1000 comentando los resultados.