Pràctica 9. La distribució normal: ajust de dades i càlculs

L'objectiu d'aquesta pràctica és familiaritzar-nos amb la llei de distribució normal. Per això analitzarem la seva simetria i forma a la primera part de la pràctica mitjançant la comparació amb mostres de dades. Després, suposant que certes variables es distribueixen segons aquesta llei, veurem com fer càlculs de probabilitats acumulades i d'intervals fent servir l'SPSS.

▶ Recordeu activar les opcions:

- "Mostrar comandos en anotaciones" a la pestanya de "Visor".
- "Nombre y etiquetas" per a les variables i "Valores y etiquetas" per als valors a l'apartat de "Etiquetado de tablas pivot" de la pestanya de "Etiquetas de resultados".

▶ En aquesta pràctica treballarem amb els arxius *maternitat.sav*, *poblacio.sav* i *prac9mostres.sav*.

1. Exemples de dades que ajusten o no al model normal

Obriu el fitxer *maternitat.sav* i mireu les variables que conté. Quantes variables numèriques conté aquest fitxer? Què signifiquen els valors que apareixen a les columnes de les variables Pesmare i Pesnado?

▶ Feu una anàlisi exploratòria de les variables Pesmare i Pesnado.

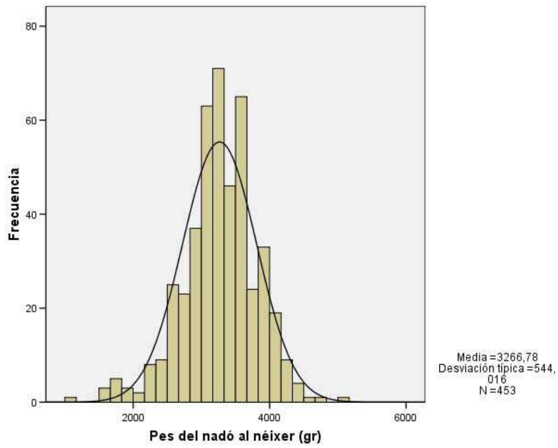
- Realitzeu un histograma per cadascuna d'elles.
- Estudieu la simetria d'aquestes variables.

L'objectiu d'aquesta primera secció és explorar la forma i simetria d'una variable per a determinar si les dades s'ajusten a una llei de distribució normal. Aquesta distribució té una forma característica coneguda com "campana de Gauss". És simètrica centrada a la mitjana de la variable. De fet, la seva forma ve completament determinada pel coneixement de la mitjana i la desviació típica. Haureu observat doncs que els histogrames anteriors tenen una forma semblant a la d'una distribució normal.

Per poder observar si les dades s'ajusten o no al model normal (s'aproximen per una llei de distribució normal), podem demanar a l'SPSS que ens representi en una mateixa gràfica tant l'histograma com la forma de la distribució normal. Per això, cal activar l'opció Mostrar curva normal, a la finestra de l'histograma. És a dir, feu

Gráficos → Histograma,

activeu **Mostrar curva normal** i seleccioneu la variable que voleu representar. Anem a fer-ho per a la variable pesnado.

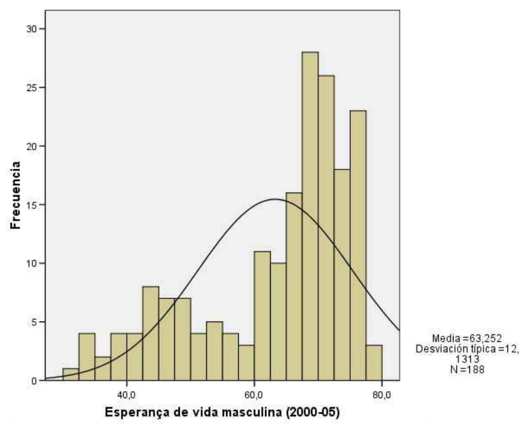


Quina conclusió traieu de la figura anterior? Dirieu que s'ajusten a un model normal?

▶ Repetiu el que acabem de fer per a la variable pesmare.

Tot seguit analitzarem la variable espvidam que es troba al fitxer *poblacio.sav*. Comenceu obrint aquest fitxer i repetiu el procés d'anàlisi que acabem de realitzar però ara amb la variable espvidam.

- Analitzeu la seva simetria.
- Podeu treure les mateixes conclusions que abans?



▶ Repetiu el que acabem de fer per a altres variables del fitxer *poblacio.sav*.

2. Les mitjanes s'ajusten (aproximadament) al model normal

En aquest apartat estudiarem un exemple que segueix la llei normal i que serà molt important. Donada una població i una variable, podem considerar les mitjanes de totes les possibles mostres de la mateixa mida. Això ens dóna una variable la distribució de la qual és normal.

Obriu el fitxer *prac9mostres.sav*. Aquest fitxer conté les dades recollides en l'estudi de `Despesa per llar en habitatge, aigua, electricitat i combustible (Catalunya 2000)`. Les dades estan organitzades en 70 mostres de 70 llars cadascuna.

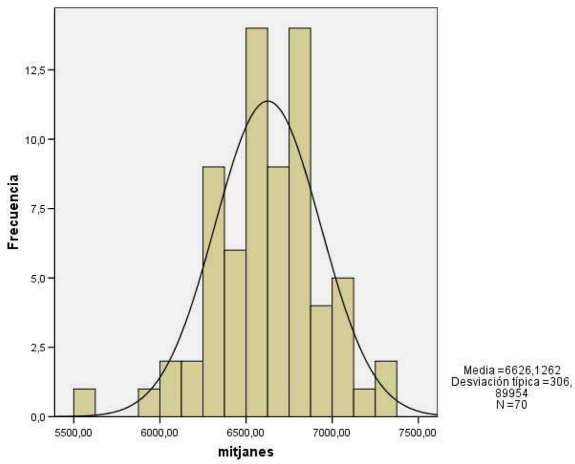
► Calculeu la mitjana de cadascuna d'aquestes 70 mostres amb

Analizar → Estadísticos descriptivos → Descriptivos.

Dins de **Opciones**, podeu desactivar totes les opcions menys la mitjana i la desviació típica.

Observeu que la columna de mitjanes és justament l'última variable mitjanes del fitxer. És a dir, recull les mitjanes de totes les mostres.

► Feu un histograma amb corba normal de la variable mitjanes per veure si s'ajusta a un model normal.



3. Càlculs amb la distribució normal

Recordeu que una distribució normal ve determinada per la seva mitjana μ i desviació típica σ. Si escrivim $X \sim N(\mu, \sigma)$ volem dir que la variable X segueix distribució normal de mitjana μ i desviació típica σ.

Un exemple pot ser el de la variable X= despesa per llar en habitatge, aigua, electricitat i combustibles (euros/any), i suposem que la podem ajustar a una distribució normal de mitjana 6624 i desviació típica 2184:

$$X \sim N(6624, 2184)$$

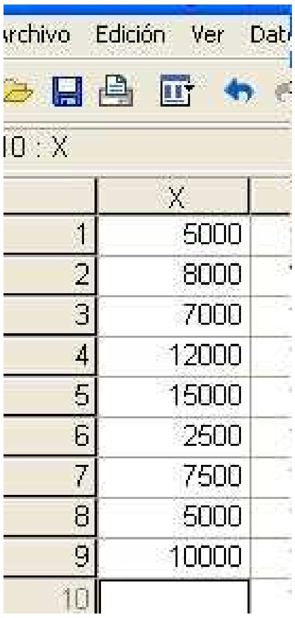
Aleshores, ens podem formular les següents qüestions:

- Quin % de llars tenen una despesa inferior a 5000 euros? (equivalentment, expressat com a proporció: quina és la probabilitat que una llar escollida a l'atzar tingui una despesa inferior a 5000 euros?)
- Quin % de llars tenen una despesa inferior a 8000 euros?

- Quin % de llars tenen una despesa superior a 7000 euros?
- Quin % de llars tenen una despesa superior a 12000 euros?
- Quin % de llars tenen una despesa superior a 15000 euros?
- Quin % de llars tenen una despesa entre 2500 i 7500 euros?
- Quin % de llars tenen una despesa entre 5000 i 10000 euros?

Per respondre aquestes preguntes, usant l'SPSS, haurem de seguir els següents passos:

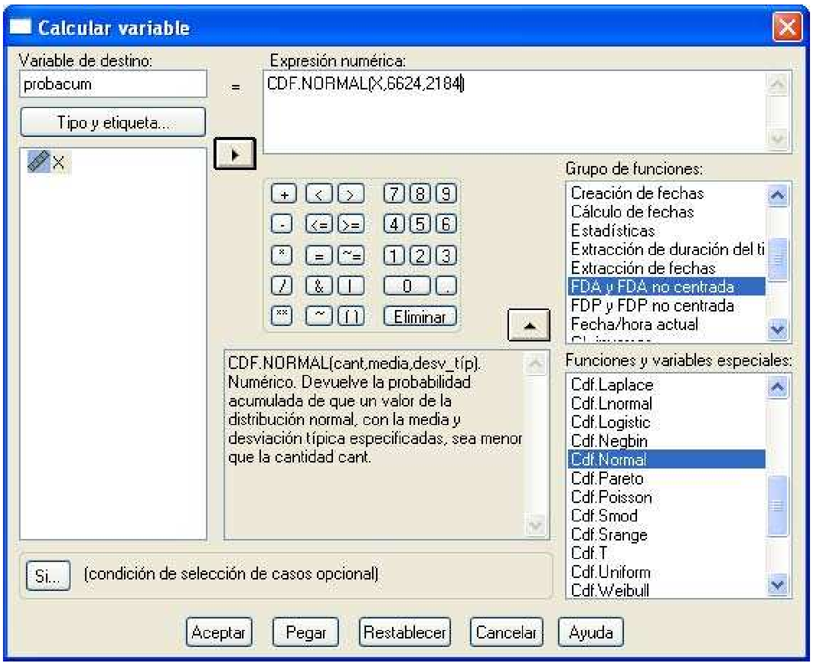
- Pas 1 En un arxiu nou de dades, introduïrem en una nova variable, que podem anomenar x, els valors dels quals volem calcular les probabilitats (els %). En el nostre cas introduïrem els valors numèrics que apareixen a les qüestions que ens han plantejat.



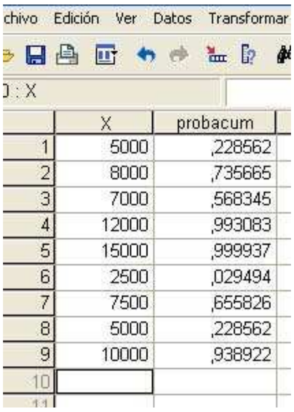
- Pas 2 Seguidament anem al menú

Transformar $\rightarrow$ **Calcular**,

on com a **Variable de destino** introduim **probacum** i en el menú de **Grupo de funciones** seleccionem **FDA y FDA no centrada**. Després en el menú anomenat **Funciones y variables especiales** busquem la funció **Cdf.Normal** . Mitjançant la fletxa, la desplacem a la finestra d'**Expresión numèrica** . Només ens falta omplir els interrogants que apareixen amb les dades: els dos últims corresponen a la mitjana (6624) i a la desviació típica (2184), mentre que el primer cal omplir-lo amb els valors dels quals volem calcular la probabilitat acumulada, és a dir, en aquest cas amb la variable X que hem creat.



De manera que es crea la variable **probacum** a la base de dades. Sel.leccionem 4 o més decimals per aquesta variable (per exemple, 6 decimals) i temim:



	X	probacum
1	5000	,228562
2	8000	,735665
3	7000	,568345
4	12000	,993083
5	15000	,999937
6	2500	,029494
7	7500	,655826
8	5000	,228562
9	10000	,938922
10		

Hem de recordar que la funció **CDF** ens dóna la probabilitat acumulada fins al valor X; és a dir, l'àrea a l'esquerra del valor.

▶ QÜESTIONS: Ara ja podem respondre a les qüestions anteriors,

- Quin % de llars tenen una despesa inferior a 5000 euros? (equivalentment, expressat proporció: quina és la probabilitat que una llar escollida a l'atzar tingui una despesa inferior a 5000 euros)

  La probabilitat d'inferior a 5000 (o percentatge acumulat fins al valor 5000) és el valor de **probacum** calculada al costat del valor 5000: 0.228562.

  Per tant, el % de llars amb despesa inferior a 5000 és aproximadament del 22.9%.
- Quin % de llars tenen una despesa inferior a 8000 euros?
- Quin % de llars tenen una despesa superior a 7000 euros?

  Si el % inferior a 7000 és ...., el percentatge superior a 7000 és ....

- Quin % de llars tenen una despesa superior a 12000 euros?
  Si el % inferior a 12000 és ...., el percentatge superior a 12000 és ....
- Quin % de llars tenen una despesa superior a 15000 euros?
  Si el % inferior a 15000 és ...., el percentatge superior a 15000 és ....
- Quin % de llars tenen una despesa entre 2500 i 7500 euros?
  Si el % inferior a 7500 és .... i l'inferior a 2500 és ...., aleshores el % entre 2500 i 7500 és ....
- Quin % de llars tenen una despesa entre 5000 i 10000 euros?
  Si el % inferior a 10000 és .... i l'inferior a 5000 és ...., aleshores el % entre 5000 i 10000 és ....

▶ EXERCICI: Considereu que la variable X corresponent al salari mensual dels professors de secundària a Espanya segueix una distribució normal de mitjana 1776 euros i desviació típica 142 euros. Calculeu,

- Quin percentatge aproximat de professors té un salari mensual de menys de 1500 euros?
- Quin percentatge aproximat de professors té un salari mensual de més de 2000 euros?
- Quin percentatge aproximat de professors té un salari mensual entre 1600 i 1800 euros?

La llei de distribució normal de mitjana μ i desviació típica σ compleix que un aproximadament un 95% dels valors es troben entre $\mu - 1.96\sigma$ i $\mu + 1.96\sigma$, aproximadament un 95.5% entre $\mu - 2\sigma$ i $\mu + 2\sigma$ i aproximadament un 99.7% entre $\mu - 3\sigma$ i $\mu + 3\sigma$.

▶ EXERCICI: Considereu que la variable X de l'exercici anterior (salari mensual dels professors de secundària a Espanya) que segueix una distribució normal de mitjana 1776 euros i desviació típica 142 euros. Calculeu,

- Quin percentatge aproximat de professors té un salari mensual entre $1776 - 142$ i $1776 + 142$ euros?
- Quin percentatge aproximat de professors té un salari mensual entre $1776 - 1.96 \cdot 142$ i $1776 + 1.96 \cdot 142$ euros?
- Quin percentatge aproximat de professors té un salari mensual entre $1776 - 2 \cdot 142$ i $1776 + 2 \cdot 142$ euros?
- Quin percentatge aproximat de professors té un salari mensual entre $1776 - 2.58 \cdot 142$ i $1776 + 2.58 \cdot 142$ euros?
- Quin percentatge aproximat de professors té un salari mensual entre $1776 - 3 \cdot 142$ i $1776 + 3 \cdot 142$ euros?