ACTIVITAT \_2

Nom i Cognoms …Iván Sánchez Alcázar……………………………………………….Grup …A5…..

Entreu al programa R\_Commander i carregueu la base de dades **Encuesta.Rdata**

(Seqüència: *Dades ► Carregar conjunt de dades*).

Realitzeu:

1. Breu anàlisi de totes les variables de la base de dades

 *(Estadístics ► Resums ► Conjunt de dades activa)*.

Indiqueu:

* Nombre de variables que conté la base de dades ......16..............
* Variables Numèriques ..11..................
* Dades factor . ..............5................
1. Distribució de freqüències de la variable **Lugar\_compra***. (Estadístics ► Resums ► Distribució de freqüències).*

¿Quin percentatge d'observacions correspon a “Grandes superficies”? ...52,75%.........

1. Histograma de la variable **Ingresos** *(Gràfiques ► Histograma)*.

¿Quina forma presenta aquesta distribució? Campaniforme / Asimètrica a la dreta / Asimètrica a l'esquerra ...........Asimètrica a la dreta...........................

1. Taula de freqüències de la variable **Gasto\_alim** amb les instruccions:
* *hist(Encuesta$Gasto\_alim, plot=F)*
* *Encuesta$T1=cut(Encuesta$Gasto\_alim, breaks=c(300,400,500,600,700,800,900,1000,1100,1200))*

*cbind(table(Encuesta$T1), table(Encuesta$T1)/sum(table(Encuesta$T1)))*

¿Quina freqüència absoluta i relativa presenta l'interval amb marca de classe 450? ..154(ni) ..i ...... 0.386934673 (fi)..............

1. Resum numèric d’**Edad** *(Estadístics ► Resums ► Resums numèrics)*

Indiqueu: Edat Mínim .18...... Màxim .65...... Mediana...39.... Recorregut interquartílic ..22..... i Edat Mitjana ...39.86.....

1. Stem and Leaf de **Gasto\_alim i** **Ingresos**. *(Gràfiques ► Gràfica de tija i fulles)* Indiqueu:
* Nom de la variable que presenta valors outliers ....Gasto\_alim i Ingressos........
* Despesa màxima del 40% d' enquestats que gasta menys...440.......
* Percentatge d' enquestats que gasten més de 600€ ...15,07%.......
* Ingrés mínim i màxim .....850(min) 3150,25(max)................
1. Box-plots de **Gasto\_alim** seleccionant a *Gràfica per grups* la variable **Zona***. (Gràfiques ► Diagrama de caixa)* Indiqueu:
* Quina zona presenta valors outliers? .....Barcelona.............
* Zona que té major mediana **(BCN** / Hospitalet / Sabadell)
* Zona amb menor rang interquartílic (BCN / Hospitalet / **Sabadell**)
* Zona amb major valor màxim (exclòs l’outlier) (**BCN** / **Hospitalet** / Sabadell)
* Quina zona té menor despesa mitjana? (BCN / Hospitalet / **Sabadel**l)
1. Resum numèric de **Gasto\_alim i** **Ingresos** *(Estadístics ► Resums ► Resums numèrics).* Indiqueu quina de les dues variables té una mitjana més representativa ...Gasto\_alim.......activar la opción coeficiente de variacion.... 0.2190307........
2. Estandarditzeu la variable **Edad** i comproveu amb el resum numèric que la mitjana es 0 i la variància 1. *(Dades ► Modificar variable ► Tipificar variable)*. És cert que la mitjana és 0 i la varianza 1

 mean sd IQR 0% 25% 50% 75% 100% n

 3.971649e-17 1 1.646258 -1.635782 -0.8126529 -0.06435373 0.8336052 1.881224 400

1. Obtingueu una nova variable amb nom **Tram\_Ingreso**, com a resultat de recodificar la variable **Ingresos** de la següent forma: *en estadisticos>resumenes>disdtribucion de frecuecnais*

|  |  |
| --- | --- |
| **Variable Ingresos**  | **Tram\_Ingreso** |
| 858 < **Ingresos** <= 1266.1221266.122 < **Ingresos** <= 1909.041909.04 < **Ingresos** <= 3151 | **BajoMedioAlto** |

Seqüència: *(Dades ► Modificar variables del conjunt de dades actives ►Recodificar variables)*



Indiqueu:

* Quants consumidors tenen uns **Ingressos Baixos**?.....99..................
* Quin percentatge de consumidors tenen uns **Ingressos Mitjans**?.......49.75...............
* La mitjana de la variable **Gasto\_alim** per als consumidors amb uns **Ingressos Mitjans** és igual a: ........477.0765.................
* En quin tram d’ingressos la variable **Gasto\_alim** és més homogènia?...............el tram baix...........................
1. Obtingueu la taula de doble entrada de les variables **Tram\_Ingreso** i **Lugar\_compra**

*(Estadístics ► Taules de contingencia ► Taules de doble entrada)*

* Quin percentatge d'individus tenen uns **Ingressos Alts** i compren en **Grans Superfícies**? ..s/total....11,2%..............
* Per als consumidors amb uns **Ingressos Mitjans*,*** quin percentatge compren al **Mercat Municipal**? ....s/fila.......16,8%...........
* Per als individus que compren habitualment en els **Comerços del Barri**, quin percentatge tenen uns **Ingressos Baixos**?....33%..........s/columna
1. Obtingueu el coeficient de correlació entre les variables **Ingresos** i **Gasto\_alim**. Comenteu el resultat:

*(Estadístics ► Resums ► Matriu de correlacions)*

.... El resultat és 0.8803677....i per tant la correlació entre les variables és aproximadament lineal positiva..................................................................................................................................................................................................................................................................................................................................

1. Ajusteu el model de regressió lineal de les variables **Ingresos**(Variable Independent) i **Gasto\_alim** (Variable Depenent). Respongueu a les següents qüestions:

*(Estadístics ► Ajust de Models ► Regressió lineal)*

* Quin percentatge de variació observada a **Gasto\_alim** s’explica amb la recta ajustada?........77.5%..........
* Per uns ingressos de 1500€, quina és la despesa mitjana estimada? ......464.95.......................
* Si desitgem tenir una despesa mitjana en alimentació de 550€, quin hauria de ser el nivell de ingressos? .............1935.48........................................
* Si s'incrementen en un 3% els ingressos de tots els individus, com queda afectada la recta de regressió?

..la pendent de la recta de regressió incrementa, i per tant, la recta serà més inclinada..................................................................................................................................................................

* Obtingueu el diagrama de dispersió de X=Ingresos i Y=Gasto\_alim

