



PRÁCTICA 4. TABLAS DE FRECUENCIAS BIDIMENSIONALES. MODELOS DE REGRESIÓN.

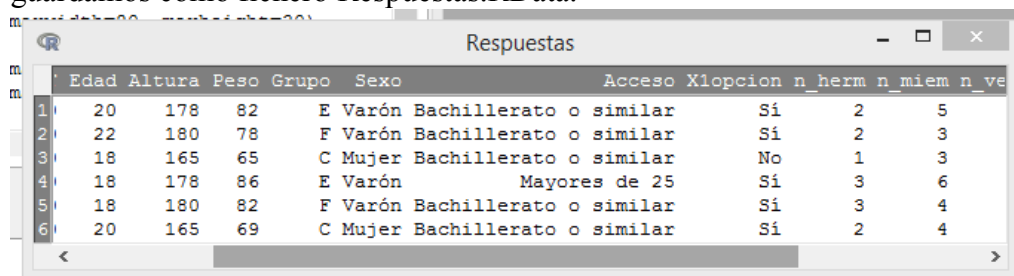
Ejecute library(Rcmdr)

Vamos a descargar los datos que se han recogido sobre los alumnos de este curso. Las variables que podemos utilizar son las siguientes:

Descripción de la variable	Nombre	Tipo
Marca temporal	T	Fecha
Edad	Edad	Discreta
Altura (en cm)	Altura	Continua
Peso (en Kg)	Peso	Continua
Grupo	Grupo	Cualitativa
Sexo	Sexo	Cualitativa
¿Cómo accediste a esta titulación?	Acceso	Cualitativa
¿El Grado en Ingeniería Informática era tu primera opción?	I_1opcion	Cualitativa
Número de hermanos (incluyéndote a ti)	n_herm	Discreta
Número de miembros de tu hogar	n_miem	Discreta
Número de vehículos en tu hogar	n_veh	Discreta
¿Has estudiado alguna vez Estadística?	Est_Est	Cualitativa
¿Qué red social prefieres para comunicarte?	Red	Cualitativa
¿Qué plataforma docente prefieres?	Plataf	Cualitativa

Ejercicio 1

- a) En primer lugar, copiamos los datos “respuestas.xls” a una carpeta en su escritorio denominada Practica4. En R-Commander, cambiamos el directorio de trabajo a dicha carpeta (Menú fichero). A continuación, importamos los datos de Excel a R-Commander mediante el Menú Datos. Los visualizamos y los guardamos como fichero Respuestas.RData.



	Edad	Altura	Peso	Grupo	Sexo	Acceso	X1opcion	n_herm	n_miem	n_veh
1	20	178	82	E	Varón	Bachillerato	o similar	Sí	2	5
2	22	180	78	F	Varón	Bachillerato	o similar	Sí	2	3
3	18	165	65	C	Mujer	Bachillerato	o similar	No	1	3
4	18	178	86	E	Varón	Mayores de 25		Sí	3	6
5	18	180	82	F	Varón	Bachillerato	o similar	Sí	3	4
6	20	165	69	C	Mujer	Bachillerato	o similar	Sí	2	4

- b) Comencemos creando una tabla de doble entrada con variables cualitativas (Sexo y Red Social Preferida). Veamos si influye el ser hombre o mujer en la preferencia. Nos vamos a Estadísticos/Tablas de contingencia/Tablas de doble entrada. Introducimos las variables Red y a continuación Sexo. Calcule



porcentajes totales, por fila y por columna, para la obtención de marginales y condicionadas de la tabla.

- c) Para estudiar el grado de relación/asociación entre las variables, se puede obtener el estadístico chi-cuadrado. Dicho estadístico está acotado entre:

$$0 \leq \chi^2 \leq N \cdot \min\{p - 1, q - 1\}$$

Siendo el 0, la mínima asociación. Estudie la tabla obtenida.

- d) Construya la tabla de doble entrada del Número de vehículos y Número de miembros de la Unidad Familiar. ¿Influye esta segunda variable en la primera?. Obtenga la distribución condicionada del número de miembros cuando las familias solamente tienen un vehículo en su hogar.

Ejercicio 2

Se quiere estudiar la posible asociación entre el nivel de estudios de un grupo de personas y el hábito de fumar. Las personas se han seleccionado de forma aleatoria y los datos se presentan en la tabla adjunta.

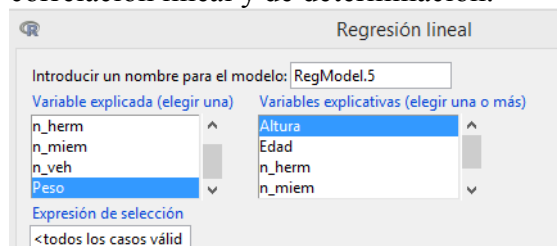
Fumar \ n_estudios	Primarios	Medios	Superiores
SI	20	10	4
NO	16	12	2

- a) Introduzca los datos (ya tabulados) en R-comander usando el Menú Estadísticos/Tablas de contingencia/Introducir una tabla de doble entrada.
b) Calcule las distribuciones marginales.
c) ¿Son dichas variables independientes?

Ejercicio 3

En este ejercicio, vamos a obtener un modelo de regresión lineal para predecir el peso en función de la altura, usando los datos “Respuestas.RData”. En este caso, graficaremos en primer lugar la nube de puntos y a continuación realizaremos el ajuste.

- a) En el menú Gráficas elija Gráfica XY e introduzca Altura, Peso. Comente dicho diagrama de dispersión.
b) En el menú Estadísticos/Ajuste de modelos/Regresión lineal, realice el ajuste del Peso en función de la Altura. Escriba el modelo e interprete el coeficiente de correlación lineal y de determinación.

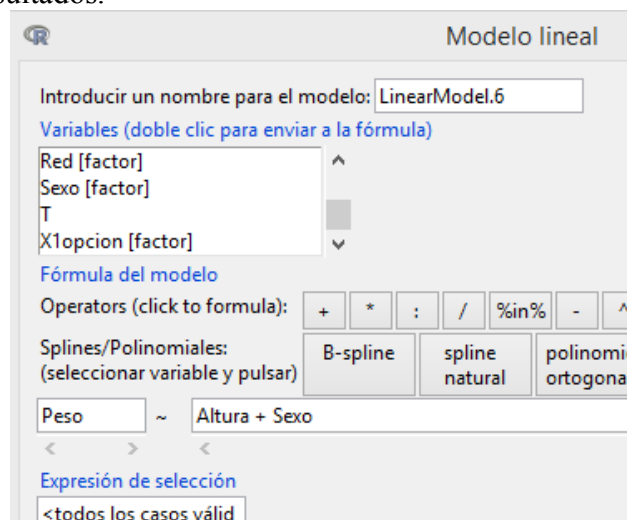


(puede cambiar la selección, escribiendo por ejemplo, `Sexo=="Varón"`, con lo que haremos el modelo solamente para los hombres).

Para graficar el modelo ajustado a la nube de puntos (e incluir una línea suavizada, lowess line), escribimos en el Script:

```
plot(Respuestas$Altura, Respuestas$Peso, main="Nube de puntos y ajuste",  
xlab="Altura", ylab="Peso")  
abline(lm(Respuestas$Peso ~ Respuestas$Altura), col="red")  
lines(lowess(Respuestas$Peso, Respuestas$Altura), col="blue")
```

- c) Puede ver el ajuste por sexos introduciendo la orden en el Script:
`scatterplot(Peso ~ Altura | Sexo, xlab="Altura", ylab="Peso", data=Respuestas)`
- d) También se puede acceder en R-commander a otros modelos (el que hemos visto en el apartado anterior es un modelo de regresión lineal simple). En el menú Estadísticos/Ajuste de modelos/Modelo lineal se pueden ajustar modelos más complejos, introduciendo mayor número de variables explicativas (regresión múltiple), que incluso pueden ser cualitativas (factores) que serán convertidos a variables dicotómicas. El sexo de un individuo seguramente influya en su Peso. Intente ajustar el Peso en función de la Altura y el sexo y comente los resultados.



- e) Intente ahora incluir también el número de hermanos(`n_herm`) y comente los resultados obtenidos. ¿Cree que aporta algo dicha variable al modelo construido?
- f) También se pueden realizar modelos “linealizables”, como por ejemplo, el modelo exponencial $Y = \exp(a + b \cdot X)$, el modelo logarítmico $Y = a + b \cdot \ln(X)$ o modelo multiplicativo $Y = a \cdot X^b$, entre otros. Vamos a ajustarlos al Peso en función de la Altura y elegiremos el que mejor ajuste nos ofrezca.
 - i. El modelo exponencial se puede escribir como $\ln(Y) = a + b \cdot X$. Por tanto, en el script de R-comander pondremos:
`Exponencial <- lm(log(Peso) ~ Altura, data=Respuestas)`
`summary(Exponencial)`
Para graficarlo:



```
scatterplot(log(Peso)~Altura,reg.line=lm,  
xlab="Altura", ylab="Peso", data=Respuestas)
```

- ii. El modelo logarítmico se puede obtener de la siguiente forma:

```
Logaritmico <- lm(Peso ~ log(Altura), data=Respuestas)  
summary(Logaritmico)
```

Para graficarlo:

```
scatterplot(Peso~log(Altura),reg.line=lm,  
xlab="Altura", ylab="Peso", data=Respuestas)
```

- iii. El modelo multiplicativo se puede escribir $\ln(Y) = \ln(a \cdot X^b) = \ln(a) + b \cdot \ln(X)$

Por tanto, en el script de R-comander pondremos:

```
Multiplicativo <- lm(log(Peso) ~ log(Altura), data=Respuestas)  
summary(Multiplicativo)
```

Tenga en cuenta que el valor del “intercept” será $\ln(a)$ y por tanto, para hallar “a”, tendremos que exponenciar su valor.

Para graficarlo:

```
scatterplot(log(Peso)~log(Altura),reg.line=lm,  
xlab="Altura", ylab="Peso", data=Respuestas)
```

