TEMA 3. DESCRICION ESTATÍSTICA DE DATOS MULTIDIMENSIONAIS.

3.1. Distribucions Bidimensionais de Frecuéncias. Representacion Grafica.

No tema anterior nos ocupabamos de descrever as variables estatísticas unha por unha, cada indivíduo da mostra era descrito de acordo con unha única caractarística independentemente de outras posibles características de interese no mesmo indivíduo. Sen embargo, as vezes, necesitamos estudar nun mesmo indivíduo várias características simultáneamente, tratando de averiguar as posibles relacións entre características; ou mesmo estudar unha característica dada condicionada a certos valores posibles de outras. Por iso cuando observamos varias características simultáneamente nun mesmo indivíduo estamos a obter os valores dunha variable K-dimensional $(X_1, X_2, ...X_k)$.

Por exemplo consideremos o estudo simultáneo das cinco características en impresoras laser indicadas no exemplo abaixo. Temos entón unha variable de este tipo con cinco componentes $(X_1, X_2, ...X_5)$.

Imos comezar o estudo simultáneo de varias características polo caso en que consideremos duas. Isto é, polo que chamaremos unha variable bidimensional (X_1, X_2). Tamén se utiliza a notación (X, Y).

O primeiro que imos facer é reduzir os datos tabulando a informacion sobre os valores aparecidos e o número de vezes que aparecen no que se chama unha TABOA DE FRECUÉNCIAS.

Exemplo: Características relativas ás impresoras láser:

	X_1	X_2	X_3	X_4	X_5		X_1	X_2	X_3	X_4	X_5
1	1180	39	5	25	512	**21**	14710	42	4	50	5120
2	2550	64	10	25	716	**22**	46480	43	13	500	1024
3	2672	74	10	25	204	**23**	30780	29	32	150	1222
4	1038	6	6	15	102	**24**	10990	22	4	150	1024
5	1279	16	6	15	102	**25**	14490	27	6	150	1024
6	9878	18	6	15	512	**26**	19490	54	6	150	2048
7	2500	14	10	25	204	**27**	34990	35	8	250	2048
8	3410	22	10	25	204	**28**	23800	45	8	100	2048
9	4890	22	16	25	204	**29**	22240	29	30	250	6144
10	4700	14	3	10	256	**30**	28990	29	30	550	6144
11	1550	14	4	10	102	**31**	28000	23	10	100	2048
12	2400	14	7	25	204	**32**	14900	23	6	100	1024
13	2690	17	8	25	153	**33**	11500	23	4	100	1024
14	3150	39	8	25	204	**34**	26500	87	10	200	2048
15	5990	62	9	25	819	**35**	28500	87	16	500	4096
16	5310	62	8	25	819	**36**	10900	26	4	100	1024
17	7990	17	4	10	102	**37**	27500	45	12	250	2048
18	1122	17	4	10	153	**38**	80000	35	16	100	1024
19	2539	25	8	25	204	**39**	48145	14	18	250	2048
20	4913	17	8	40	153	**40**	21558	50	4	100	1638

Na Tab. anterior podemos ollar 5 variables medidas simultáneamente en 40 impresoras laser: X_1 = Prezo en pts., X_2 = Número de fontes de escritura, X_3 = Velocidade de Impresión, X_4 = Número de follas que pode conter o cargador de papel, X_5 = Memória en Kbytes.

Para comezar consideramos unha variable bidimensional estudando duas características nunha mesma poboación.

Sexa (x_1, y_1), ...(x_n, y_n) unha mostra de n observacions de unha variable estatística bidimensional (X,Y). Esta información pode resumir-se nunha taboa de frecuéncias onde aparecerán as modalidades respeitivas das variables (x_i, y_j) e o número de veces que aparece o par anterior será a frecuéncia absoluta n_{ij}.

Proriedades das frecuéncias:

a) $n_{ij} \geq 0$

b) $0 \leq f_{ij} \leq 1$

c) $\sum_i \sum_j n_{ij} = n$ e $\sum_i \sum_j f_{ij} = 1$

Sendo n_{ij} a frecuencia absoluta do par (x_i, y_j), e f_{ij} a frecuencia relativa do mesmo par, que poderá vir expresada en tanto por un ou en tanto por cen.

Esta taboa nomea-se Taboa de Continxéncia cando as duas variables X e Y son de tipo cualitativo, en xeral a información que aparece nela recibe o nome de ***Distribución Conxunta de X e Y***. A disposición é como segue.

Táboa de Continxéncia que resume a Distribución Conxunta duna variable bidimensional (X, Y)

Y / X	y_1	y_2	…	y_j	…	y_m	
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1m}	$n_{1.}$
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2m}	$n_{2.}$
x_i	…	…	…	n_{ij}	…	…	$n_{i.}$
……	…	…	…	…	…	…	
x_l	n_{l1}	n_{12}	…	n_{1j}	…	n_{lm}	$n_{l.}$
	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.m}$	$n_{..}=n$

No seguinte exemplo temos unha táboa de continxéncia (variables cualitativas) onde temos as frecuéncias absolutas e relativas %.

Exemplo: Datos "Encuesta General USA 1991"

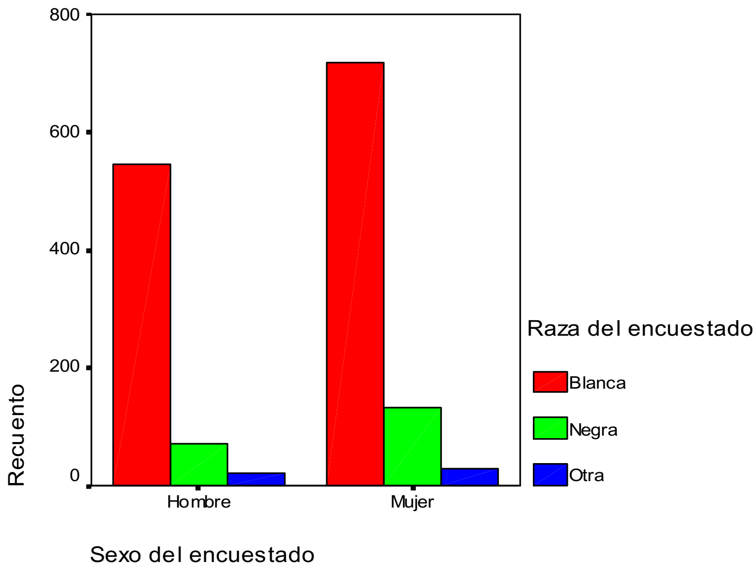
Tabla de contingencia Sexo del encuestado * Raza del encuestado

			Raza del encuestado			Total
			Blanca	Negra	Otra	
Sexo del encuestado	Hombre	Recuento	545	71	20	636
		% del total	35,9%	4,7%	1,3%	41,9%
	Mujer	Recuento	719	133	29	881
		% del total	47,4%	8,8%	1,9%	58,1%
Total		Recuento	1264	204	49	1517
		% del total	83,3%	13,4%	3,2%	100,0%

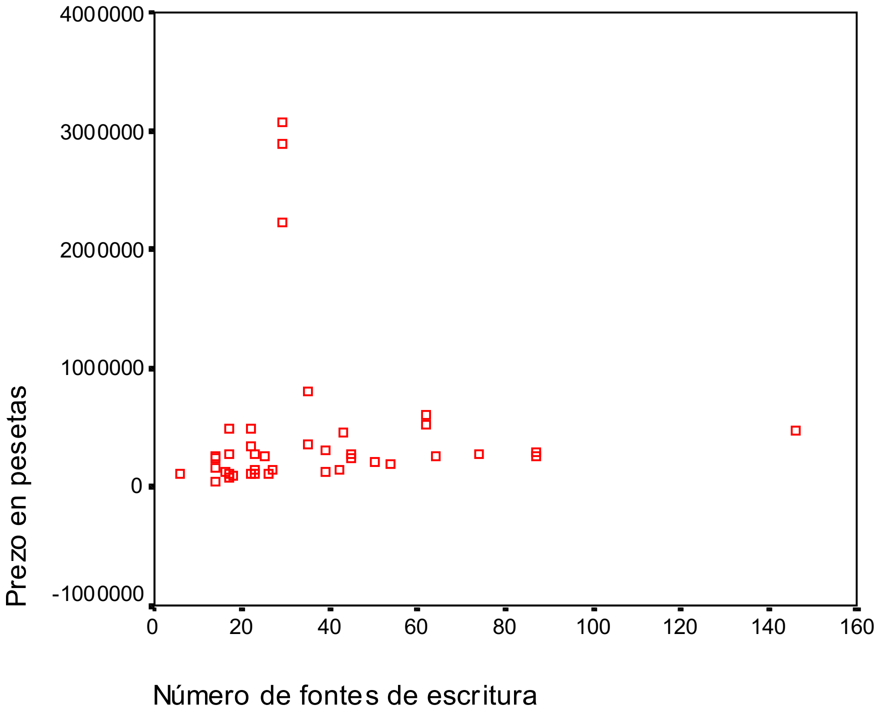
Representacions Gráficas:

A seguir podemos ollar distintas representacións gráficas de variables bidimensionais e mesmo o último gráfico que se utiliza para variables tridimensionais. Máis aló de tres variables resulta imposible representar ou interpretar os gráficos

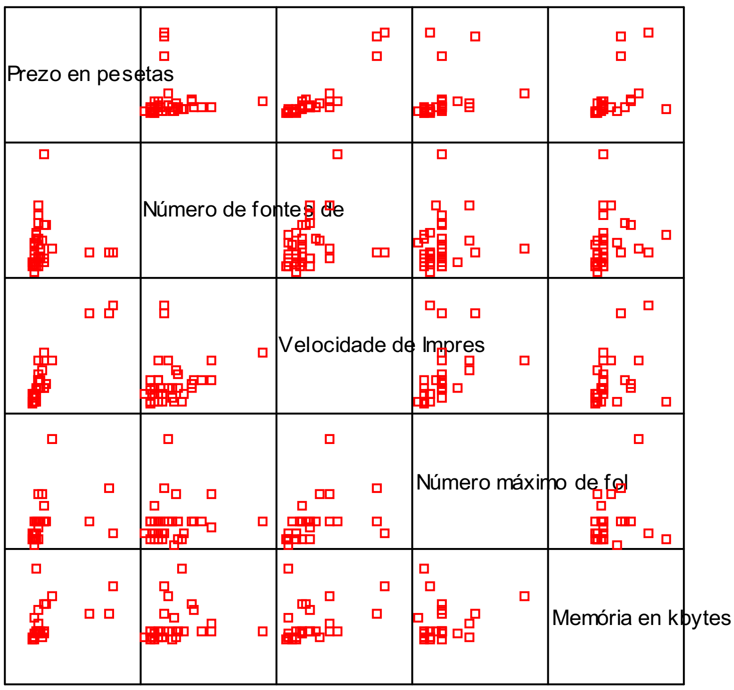
Gráfica 1. Gráfico de Barras



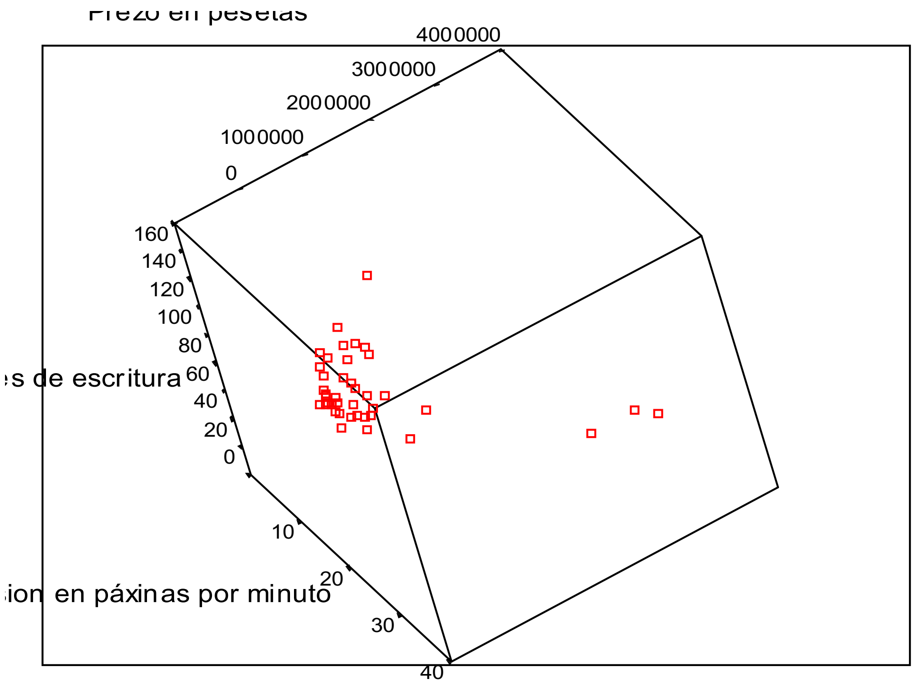
Gráfica 2. Gráfico de dispersión de dúas variables



Gráfica 3. Gráfico de dispersión simultáneo das cinco variables, tomadas de duas en duas.



Gráfica 4. Gráfico de dispersión de tres variables.



3.2. Distribucións Marxinais e Condicionadas. Independéncia.

Distribucions Marxinais.

As veces interesa estudar por separado cada variable ou mesmo estudar unha variable condicionada a certo(s) valor(es) da outra.

O estudo indivudual de cada variable, independentemente da outra constitúe o que chamamos Distribución Marxinal de X ou Distribución Marxinal de Y, segundo sexa o caso. A Distribucion Marxinal da variable X está constituida polos valores que toma a variable X xunto con o número de veces que aparece cada valor. ($X=x_i$, $n_i=n_{i.}=\sum_j n_{ij}$). Estas frecuencias absolutas da variable marxinal X aparecen na última coluna da táboa de frecuencias.

Do mesmo xeito a Distribución Marxinal de Y estará constituida polos valores que toma a variable Y xunto con o número de veces que aparece cada valor. ($Y=y_j$, $n_j=n_{.j}=\sum_i n_{ij}$). Estas frecuéncias absolutas da variable marxinal Y aparecen na última fila da táboa de frecuéncias.

Podemos observar na taboa do exemplo cos datos da “Encuesta General USA 1991” toda a información das distribucións marxinais de X e de Y na última coluna e na ultima fila respeitvamente.

Distribución condicionadas.

Poderíamos estar interesados en saber como é a distribución das razas dependendo do sexo. Isto significa supeditar o estudo duna variable a certo(s) valor(es) da outra. Na Enquisa Xeral USA 1991, gráficamente aparece na Gráfica 1 unha representación simultánea das duas distribucións condicionadas: X=Raza condicionada a Y=home ou Y=muller.

Coñecer os valores que toma unha variable dada X condicionada a certo ou certos valores de outra $Y=y_j$ xunto con o número de veces que aparecen aqueles valores $n_i/Y=y_j$ chámase *distribución de X condicionada a $Y=y_j$.*

Independéncia de duas variables X e Y.

Intuitivamente a independéncia entre duas variables se dá cando o coñecemento de certos valores de unha variable non presupón a aparicion de determinados valores da outra. Isto na práctica se comproba ollando se a distribución marxinal duna variable coincide con cualquer distribución condicionada, independentemente do valor ou valores aos que condicionemos. Cuando non hai independencia resulta de interese estudar en que modo están relacionadas as variables. Agora deberemos ter en conta que non se pode medir esta relacion de igual modo para variables cualitativas ou cuantitativas. Dentro de estas últimas tamén distinguiremos entre variables de intervalo ou variables ordinais.

3.3. Momentos dunha Distribución de Frecuéncias.

Os momentos dunha Distribución de Frecuéncias son medidas que resumen a información de unha variable estatística, xa temos visto algúns exemplos de este tipo de medidas como son as medidas de centralización, dispersión ou forma. Os momentos xeneralizan estas medidas. Podemos definir momentos de distribución de frecuencias tamén para as variables unidimensionais, bidimensionais, etc., sempre que tratemos con variables numéricas ou cuantitativas.

Lembremos os momentos para unha variable unidimensional.

Momento de orde r a respeito de un ponto c, c ou r poden ser números reais:

$$\mu_{r,c} = \sum_i (x_i - c)^r f_i$$

Casos particulares:

a) se $c = \bar{x}$ o momento chama-se momento central

b) se c = 0 o momento chama-se momento a respeito da orixe.

Exemplo de momentos son a propria média, que é o momento respeito da orixe de orde r=1.

<u>*Momentos de unha varible bidimensional.*</u>

Para unha variable bidimensional (X,Y) podemos definir o momento de orde (r,s) a respeito de dous pontos <u>c</u> e <u>d</u> como: $\mu_{(r,s),(c,d)} = \sum_{i,j}(x_i - c)^r (y_j - d)^d f_{ij}$ o máis relevante é o momento de orde r=1, s=1 a respeito das médias de X e Y que se chama Covarianza $\mu_{(1,1),(\bar{x},\bar{y})} = \sum_{i,j}(x_i - \bar{x})(y_j - \bar{y}) f_{ij}$, a covarianza mide un certo tipo de relacion entre duas variables como veremos a seguir.

3.4. Covarianza e Correlación. Recta de Regresión.

Xa temos indicado que cuando non hai independencia entre variables, certo(s) valor(es) de unha das variables condicionará a frecuencia con que aparecen os valores da outra. En estos casos resulta de interese saber medir a <u>*intensidade e forma de relacion entre variables.*</u>

A forma de relacionar-se duas variables pode ser diversa, os modelos que se utilizan son modelos matemáticos que expresan unha variable en funcion da outra como poden ser: rectas, funcions exponenciais, funcions logarítmicas, etc.

O xeito mais simple de relacionarmos duas variables é unha recta: Y=a + bX, entón temos varias incógnitas por abordar como son a estimación dos parámetros a e b, saber o grao de axuste dos nosos datos a este modelo ou ecuación. Isto último nos proporcionará o grao de relacion entre as variables. En resumo deberemos averiguar a FORMA e a INTENSIDADE da relacion entre as variables implicadas.

A Covarianza definida anteriormente pode interpretarse como unha medida da Relacion Linear entre as duas variables X e Y de tipo cuantitativo. E un valor positivo ou negativo: se a relación é directa a covarianza é un número positivo e se a relación entre as variables é inversa a covarianza é un número negativo. Por exemplo o peso e a altura de unha persoa dá unha covarianza positiva porque a relación é directa *a máis altura mais peso*. E o gasto de unha persoa en certo tipo de produtos pode ser de tipo inverso, como por exemplo o consumo de caramelos e a idade de un individuo.

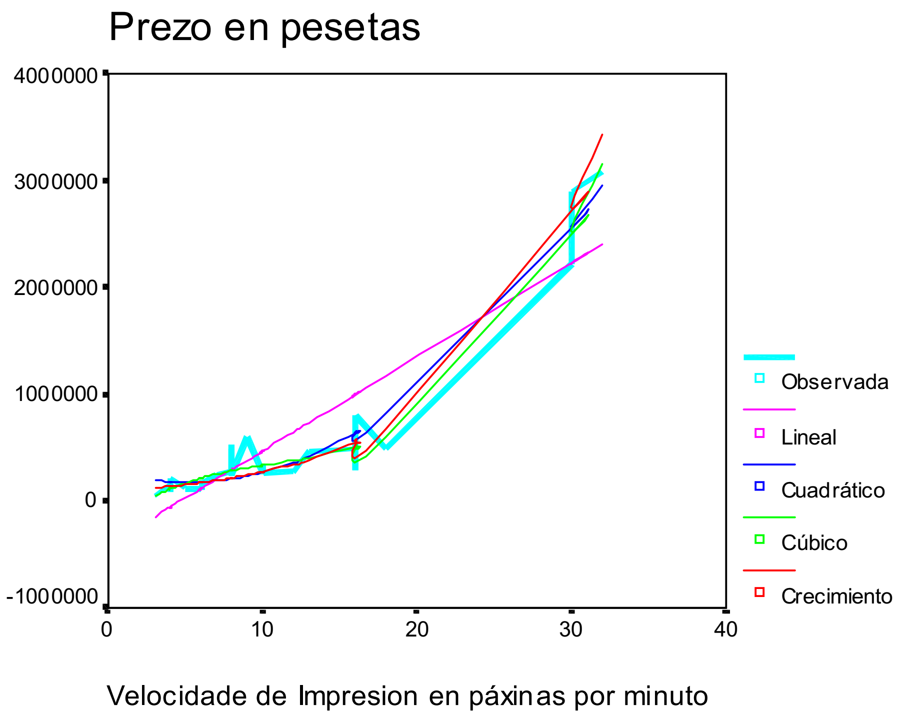
O problema da covarianza como medida de relacion é que cambiando de unidades cambia a sua magnitude e por tanto non é doada de interpretar. Unha medida alternativa sen unidades é o coeficiente de correlación linear $r=COV(X,Y)/\ \sigma_x\sigma_y$ o cuadrado do coeficiente de correlación chama-se coeficiente de determinacion r^2.

O coeficiente de correlación só mide unha relación linear entre as variables: $Y = a + bX$. O coeficiente de correlación está acoutado entre –1 e 1, cuanto máis perto de 1 ou –1 máis intensa será a relación linear entre as variables X e Y. Unha aproximación ao valor 0 do coeficiente de correlación indica unha falta de relación linear entre as variables, pero isto non presupón que non haxa outro tipo de relación como pode ser: cuadrática, cúbica, logaritmica, exponencial, etc. Para relacións non lineares debemos interpretar o coeficiente de Determinación R^2 correspondente a cada modelo. Este coeficiente de determinación pode definir-se para cualquer tipo de relacion funcional.

Abaixo temos os distintos coeficientes para diferentes modelos. No gráfico de abaixo podemos ollar diferentes modelos axustados aos datos sobre as impresoras laser. Podemos ver claramente que a Recta de Regresión en comparacion con outros modelos:

Independent: X3

Dependent	Mth	b0	b1	b2	b3
X1	LIN	-411179	88022,3		
X1	QUA	73598	-38804	3837,14	
X1	CUB	-226516	114123	-8338,1	252,363
X1	GRO	11,3559	,1154		



3.5. Táboas de Continxéncia

Cuando facemos unha análise simultánea de duas características cualitativas en estudo ou cando tratamos de avaliar conxuntamente a resposta de duas preguntas nunha enquisa, obtemos o que se chama TÁBOA DE CONTINXÉNCIA. O número de casos de cada combinación de valores das duas variables aparece nunha cela da táboa, xunto con vários percentaxes. Estas entradas nas celas proporcionan información acerca das relacións entre as variables. Tentaremos responder a preguntas como: Hai relación entre o sexo e a resposta á pregunta formulada? Hai influéncia do sexo na resposta? Son independentes?

Primeiro imos botar unha ollada á táboa de continxéncia, ás suas percentaxes e o que nos queren dicer:

Tabla de contingencia ¿POSEE ALGUNA INFORMACIÓN ? * SEXO

			SEXO		Total
			HOMBRE	MUJER	
¿POSEE ALGUNA INFORMACIÓN ?	SI	Recuento	17	13	30
		% de ¿POSEE ALGUNA INFORMACIÓN ?	56.7%	43.3%	100.0%
		% de SEXO	85.0%	65.0%	75.0%
		% del total	42.5%	32.5%	75.0%
	NON	Recuento	3	7	10
		% de ¿POSEE ALGUNA INFORMACIÓN ?	30.0%	70.0%	100.0%
		% de SEXO	15.0%	35.0%	25.0%
		% del total	7.5%	17.5%	25.0%
Total		Recuento	20	20	40
		% de ¿POSEE ALGUNA INFORMACIÓN ?	50.0%	50.0%	100.0%
		% de SEXO	100.0%	100.0%	100.0%
		% del total	50.0%	50.0%	100.0%

Contido das celas e marxinais

Os números situados á direita e debaixo da táboa coñecen-se como ***marxinais*** e manifestan as variables por separado, independentemente unha da outra.

Dentro das celas podemos considerar, por orde de aparición cara abaixo:

i) o número de casos ou frecuencia absoluta

ii) a percentaxe de fila, percenctaxe calculado sobre o total de esa fila.

iii) a percentaxe de coluna, percentaxe calculado sobre o total de esa coluna.

iv) a percentaxe sobre o total da táboa

Cál das percentaxes nos interesa?

Eleición das Percentaxes

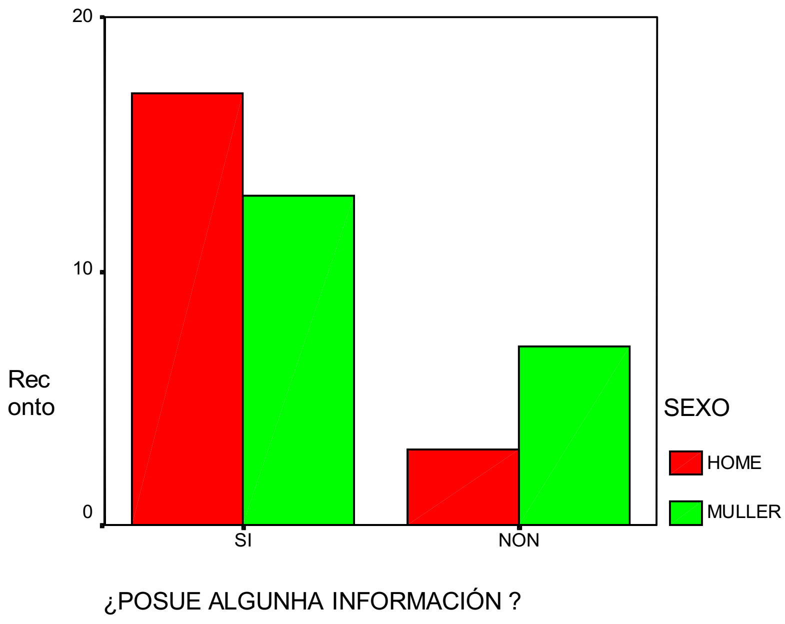
Se unha das variables está baixo control experimental se chamará variable independente. Supon-se que a variable independente afectará á resposta ou variable dependente. No noso caso a Sexo é a variable independente e a resposta á pregunta a variable dependente. Como REGRA XERAL seleccionamos a percentaxe da fila se a variable independente é a variable fila, e a percentaxe coluna se a variable independente é a variable coluna. No noso caso a información relevante está a percentaxe coluna posto que ao facermos a táboa temos situado a variable independente (sexo) nas colunas.

Adición dunha variable de control

As veces necesitamos ver se grupos concretos de homes e mulleres difiren, en algunha resposta a outra(s) preguntas(s), teremos entón, que incluir variables adicionais na táboa de continxéncia. Por exemplo se desexamos ver se o estado civil afecta ás respostas de homes e mulleres, poderíamos facer unha táboa de continxéncia de sexo e a resposta para cada unha das categorías de interese do estado civil.

Representación gráfica das táboas de continxéncia.

Igual que as táboas de frecuéncias, a representación visual de unha táboa de continxéncia simplifica, a miúdo, a busca de asociacións. Na gráfica que segue podemos ollar claramente a diferencia de opinión dos homes e das mulleres e a sua relación segundo manifestan que "posúen algunha información"



Utilización da táboa de continxéncia para o filtrado de datos

Os erros e valores inusuais da entrada de datos que non poden representar-se con táboas de frecuéncias as veces poden identificar-se utilizándo unha táboa de continxéncia. Por exemplo, un caso identificado como home con un historial de tres embarazos non se identificaría como suspeitoso nas táboas de frecuéncias de sexo e número de embarazos. Cuando se consideran por separado, o código home é aceptable para sexo e o valor 3 é aceptable para o número de embarazos. Sen embargo, a combinación é inesperada.

Estatísticos das táboas de continxéncia.

Ainda que un exame das distintas percentaxes de fila e coluna nunha táboa de continxéncia é un primeiro paso útil no estudo das relacións entre duas variables, as percentaxes de fila e de coluna non permiten a cuantificación ou comprobación de dita relación. Para esta finalidade, é útil considerar distintos índices que miden o grao de asociación, así como as probas estatísticas de hipóteses de que non existe asociación. Xa indicamos que cando as variables son de tipo cuantitativo existe unha medida do grao de relación entre as mesmas: o coeficiente de determinación R^2. Análogamente podemos medir o grao de Asociacion entre duas variables cualitativas polos correspondentes coeficientes. Pero distinguiremos entre datos de tipo ordinal ou datos de tipo nominal.

Cando os datos indican un orde o coeficiente que mide o grao de asociación deriva do coeficiente de correlacion linear e chama-se coeficiente de correlacion por rangos de SPEARMAN ou coeficiente de correlación ordinal:

$$\rho = 1 - \frac{6\sum_i d_i^2}{n^3 - n}$$

sendo n o número total de datos, $d_i = x_i - y_i$, con x_i e y_i os rangos correspondentes a dous critérios de ordenación A e B para cada valor observado:

Exemplo: Rangos de cinco estudantes segundo as Cualificacións en Direito Administrativo e Direito Constitucional

Alunos	D. Administrativo (x)	D. Constitucional (y)	$d_i = x_i - y_i$	d_i^2
A	1	3	-2	4
B	2	2	0	0
C	3	1	2	4
D	4	5	-1	1
E	5	4	1	1

O coeficiente de correlación ordinal é $\rho = 0.5$. A sua interpretación é igual que o coeficiente de correlación linear: -1 significa unha discordáncia perfecta, 1 significa unha concordáncia perfecta.

Exercício: calcula o coeficiente de correlación por rangos de Spearman.

Alunos	D. Administrativo (x)	D. Constitucional (y)	$d_i = x_i - y_i$	d_i^2
A	1	5		
B	2	4		
C	3	3		
D	4	1		
E	5	2		

Coeficientes de continxencia.

Agora imos estudar as medidas de asociación para duas características de tipo cualitativo. O grao de asociación para taboas de continxéncia ven medido polo coeficiente de continxéncia Xi-cuadrado χ^2 ou o cuadradado médio da continxéncia φ^2. Ambos coeficientes son positivos e o seu valor indica o grao de asociación. Canto máis perto de cero indicarán unha independéncia entre os dous caracteres. O problema de estes coeficientes é que non están acoutados e por tanto non teñen doada interpretación. Outros coeficientes de continxéncia evitan este problema.

$$\chi^2 = \sum_i \sum_j \frac{(n'_{ij} - n_{ij})^2}{n'_{ij}} = \sum_i \sum_j \frac{(Esperadas - Observadas)^2}{Esperadas}$$

$$n'_{ij} = frecuencia \;\; esperadas = n_i . n_j / n$$

$$\varphi^2 = \frac{\chi^2}{n}$$

Exemplo: De unha poboación de 100 persoas, temos observado a seguinte tab.:

	Parados	Ocupados	Total
Pais parados	11	29	40
Pais ocupados	19	41	60
Total	30	70	100

Calcular o coeficiente Xi-cuadrado

De outra poboación distinta de 150 persoas, temos observado a seguinte tab.:

	Parados	Ocupados	Total
Pais parados	31	29	60
Pais ocupados	19	71	90
Total	50	100	150

Calcular o coeficiente Xi-cuadrado e comparar coa primeira das poboacions indicando a qué conclusión se chega.

3.6. Variables Multidimensionais.

O estudo de mais de duas variables implica máis “complicación” en todos os aspectos. Para tres variables podemos obter gráficos en tres dimensións: gráficos de dispersión. Para máis variables isto resulta imposible, unicamente poderemos obter visions parciais de todo o conxunto utilizando proxeccions tridimensionais. En cuanto a medidas que resumen a información estatística contamos con: Vector de Médias, Matriz de Varianzas e Covarianzas, e Matriz de Correlacións. Por razóns obvias é de utilidade contar con un programa estatístico que nos axude a facer este tipo de cálculos. Estudaremos este tipo de variables utilizando un paquete de software estatístico que nos axude a tratar este tipo de información.

Exercícios.

Exercicio 1. Os seguintes datos son os prezos de venda en miles de pesetas (Y) de un modelo de automóvil usado durante X anos:

X	1	2	2	3	5	5
Y	635	570	575	540	499	490

a) Estimar varios modelos que se axusten aos datos anteriores.
b) Estimar os parámetros a e b da curva $y=ab^x$ utilizando a información da mostra anterior. Comentar o coeficiente de determinacion.
c) Estímese o prezo de venda de un veículo que ten catro anos de uso.

Exercicio 2. Os seguintes proporcionan información sobre a renda mensual familiar (X) e os importes mensuais dos seus recibos de teléfono, en miles de pesetas, (Y) nunha mostra de 10 fogares:

X	160	450	360	320	300	130	410	150	360	400
Y	3.5	7.8	10.2	5.6	7.5	2.6	13.0	4.2	5.9	8.5

a) Representar gráficamente os datos nun diagrama de dispersión
b) Determinar a ecuación da recta de regresion entre as variables X e Y

Exercício 3. A unha mostra de 200 persoas de ambos sexos deu-se-lles a probar margarina en manteiga e pediu-se-lles que indicasen a sua preferéncia, cos resultados da táboa. Calcula o valor do Xi-cuadrado. (Grao de asociación).

	Margarina	Manteiga
Homes	42	58
Mulleres	65	35

Exercício 4. Nunha enquisa entre estudantes sobre a sua créncia na percepcion ultrasensorial encontraron-se os datos da táboa. Indicar o grao de asociacion entre as duas características Xi-cuadrado.

	Cren totalmente	A medias	En absoluto
Enseñaría	30	50	20
Económicas	50	109	11
Humanidades	48	93	9

Exercício 5. Na táboa seguinte presentan-se as cualificacións médias de un grupo de estudantes en duas matérias. Calcula o grao de asociación.

	Cualificación Baixa	Cualificación Média	Cualificación Alta
Matéria A	15	18	7
Matéria B	2	22	23

Exercício 6. Nunha enquisa de presupostos familiares feita a200 famílias de unha cidade ten-se obtido a seguinte taboa de correlacion sobre os seus ingresos e gastos de consumo referidos ao derradeiro mes (os intervalos están expresados en miles de pesetas):

G = Gastos de consumo Ingresos = I	60-80	80-100	100-120	120-140	140-160
60-80	30	2			
80-100	25	10	3		
100-120	15	25	10		
120-140	5	18	30	3	
140-160			8	15	1

a) Estudar por separado as variables G e I calculando a distribución de frecuéncias média, desviación típica.
b) Estudar a distribucion de Gastos condicionada aos distintos rangos de valores dos ingresos. Calculando para cada condicionada a distribución de gastos, a média condicionada e a varianza condicionada.
c) Calcular a ecuación linear que liga estas duas variables. (utilizar unha calculadora)
d) Cal é o grao de axuste dos pontos á esta recta?

Exercício 7. A partir dos seguintes datos asustar duas rectas para expresar a variable Y en funcion das variables X_1 e X_2 respeitivamente.

Y_i	3	4	8	5	5
X_{1i}	1	2	3	4	5
X_{2i}	8	9	16	9	8

Onde é maior o grao de axuste ao modelo linear?

Exercício 8. O consumo de enerxia eléctrica en kwh per capita en 1985 e a renda per capita, expresada en dólares, foron os seguintes para os países da enton Comunidade Económica Europea:

País	Alem.	Bel	Din	Esp	Fr	Gr.	It.	Ir.	Lx.	Hol.	Por.	GB.
Consumo de Enerxia electrica	5730	4920	4960	2680	4510	2400	3050	2760	10450	4250	1720	4280
Renda	11090	8430	11290	4470	9860	3740	6440	4950	13650	9430	1970	8530

a) Estimar os parámetros dun modelo linear que exprese a demanda en funcion da renda do país, calculando o coeficiente de correlacion.
b) Calcular a demanda de enerxía electrica para un suposto país con renda de 6000 $.
c) Calcular os valores estimados para cada un dos paises xunto cos residuos. Interpretar os resultados.