Indice

Páginas

INTRODUCCION 9

1. VARIABLES, DISTRIBUCIÓN DE FRECUENCIAS Y REPRESENTACIÓN GRÁFICA 11

   1.1. VARIABLES 13
   1.2. DISTRIBUCIONES DE FRECUENCIAS 16
   1.3. REPRESENTACIÓN GRÁFICA 19

2. MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN. ASIMETRÍA . . 31

3. PROBABILIDAD, DISTRIBUCIÓN BINOMIAL, NORMAL Y APROXIMACIÓN DE BINOMIAL A LA NORMAL 49

4. TAMAÑO MUESTRAL 93

5. ESTADÍSTICA INFERENCIAL: ESTIMACIÓN DE MEDIAS Y PROPORCIONES 109

6. PRUEBAS DE HIPÓTESIS: UNA Y DOS MUESTRAS 121

7. CONTRASTE DE HIPÓTESIS. MÁS DE DOS MUESTRAS: EL ANÁLISIS DE VARIANZA 143

8. MEDIDAS DE ASOCIACIÓN NOMINAL Y ORDINAL 163

9. REGRESIÓN Y CORRELACIÓN 201

FORMULARIO 233

ANEXO: TABLAS 245

Indice de Tablas

		Páginas
Tabla 1.	TABLA DE DISTRIBUCIÓN DE PROBABILIDAD BINOMIAL	247
Tabla 2.	ÁREAS BAJO LA CURVA NORMAL	258
Tabla 3	NÚMEROS ALEATORIOS	259
Tabla 4.	TAMAÑO MUESTRAL EN UNA POBLACIÓN INFINITA. NIVEL DE CONFIANZA DEL 95,5%	260
Tabla 5.	TAMAÑO MUESTRAL EN UNA POBLACIÓN INFINITA. NIVEL DE CONFIANZA DEL 99,7%	261
Tabla 6.	TAMAÑO MUESTRAL EN UNA POBLACIÓN FINITA. NIVEL DE CONFIANZA DEL 95,5%	262
Tabla 7.	TAMAÑO MUESTRAL EN UNA POBLACIÓN FINITA. NIVEL DE CONFIANZA DEL 99,7%	263
Tabla 8.	DISTRIBUCIÓN DE LA T DE STUDENT	264
Tabla 9.	DISTRIBUCIÓN F DE FISHER. N.s. = 0,05	265
Tabla 10.	DISTRIBUCIÓN F DE FISHER. N.s. = 0,01	266
Tabla 11.	PUNTOS DE PORCENTAJE DEL RANGO STUDENT	267
Tabla 12.	VALORES OBSERVADOS DE H EN EL ANÁLISIS DE VARIANZA DE UNA CLASIFICACIÓN POR RANGOS DE KRUSKAL-WALLIS	268
Tabla 13.	DISTRIBUCIÓN DE χ^2	271
Tabla 14.	VALORES CRÍTICOS DEL COEFICIENTE RHO DE SPEARMAN PARA LOS NIVELES DE SIGNIFICACIÓN DE 0,05 Y 0,01	273
Tabla 15.	VALORES DE Z PARA VALORES DADOS DE R	274

1

Variables, Distribución de frecuencias y Representación gráfica

INTRODUCCION

1.1. VARIABLES

Para hacer posible la medición de la realidad social que nos rodea se hace necesaria la observación de las características de personas u objetos, así como la adjudicación de números (códigos) a dichas características facilitando posteriores comparaciones.

Las características que varían de individuo a individuo, esto es, las variables, pueden clasificarse en tres grandes grupos según el nivel de medición de las mismas:

a) NOMINAL: Es el nivel más simple y permite la clasificación de los individuos u objetos en clases o categorías meramente descriptivas, que deben ser exhaustivas y mutuamente excluyentes.

b) ORDINAL: Es el nivel inmediatamente superior al anterior, pues además de clasificar, ordena, según se posea en mayor o menor grado la característica que se pretende medir.

c) INTERVALO: Este nivel no sólo clasifica y ordena, sino que además especifica las distancias existentes entre las distintas categorías.

Existe además, el llamado nivel de razón o proporción, similar al de intervalo, pero con un «0» absoluto, que representa la ausencia total de la característica que se mide; sin embargo, este tipo de variables prácticamente no es utilizado en Ciencias Sociales, por lo que lo aglutinaremos en el nivel de medición intervalar.

Las variables también pueden ser clasificadas en cualitativas y cuantitativas. Las primeras son aquéllas que no pueden adoptar valores numéricos, lo que sí pueden hacer las segundas. Dentro de las variables cuantitativas cabe diferenciar entre discretas y continuas. Las variables DISCRETAS son aquéllas que adoptan valores que son números enteros, mientras las variables de tipo CONTINUO pueden adoptar infinitos valores fraccionados.

PROBLEMAS

PROBLEMA 1.1.

Clasifique, según el nivel de medición, cada una de las siguientes variables:

Variable

a) Estado civil (soltero, casado...)
b) Edad de los individuos
c) Las diferentes marcas de whisky de venta en España
d) Grado de satisfacción de los españoles con la Administración (muy satisfecho, bastante satisfecho...)
e) Sexo
f) Confesión religiosa
g) Estaturas de un reemplazo de soldados
h) La Renta per cápita de los países europeos
i) Partido político por el que se votó en las últimas elecciones generales
j) Nivel de estudios (no sabe leer ni escribir, lee sin estudios, primaria, bachiller...)
k) Situación laboral (ocupado, parado, jubilado...)
l) Ingresos mensuales de las familias donostiarras
m) Comunidad Autónoma de origen
n) Posición ideológica (extrema izquierda, izquierda...)

Nivel de medición

a) Nominal
b) Intervalo
c) Nominal
d) Ordinal
e) Nominal

f) Nominal
g) Intervalo
h) Intervalo
i) Nominal
j) Ordinal
k) Nominal
l) Intervalo
m) Nominal
n) Ordinal

PROBLEMA 1.2.

Determine para cada una de las siguientes variables, su carácter continuo o discreto:

Variables

a) La estatura de un grupo de alumnos
b) El número de libros prestados por día en una biblioteca
c) Edad
d) El número de discos que componen una colección de música clásica
e) El número de nacimientos por mes de una localidad española
f) El número de hijos de las familias de Aranjuez
g) Las temperaturas máximas/día registradas durante un año en Barcelona
h) El número de fallecidos en accidente automovilístico en la Comunidad Andaluza
i) El Indice General de la Bolsa de Madrid

Discreta o continua

a) Continua
b) Discreta
c) Continua
d) Discreta
e) Discreta
f) Discreta
g) Continua
h) Discreta
i) Continua

1.2. Distribuciones de frecuencias

El tratamiento estadístico de la información comienza por la clasificación y ordenación de los datos. Es decir, por la reducción de la información para que ésta sea manejable. La primera operación es el conteo. Establecemos una serie de categorías dentro de las cuales la diversidad original de los datos pertenecientes a dichas clases se pierde en aras de facilitar su interpretación, fragmentamos, por tanto, el universo en categorías homogéneas. Por ejemplo, si analizamos la edad de una población difícilmente podremos trabajar con las edades individuales (23 años y dos meses, 45 años y seis meses, 3 meses y medio...), tampoco aunque las agrupemos en edades enteras (23, 45, 0...). Lo más práctico será fragmentar la población, por ejemplo, en cuatro grupos (niños, jóvenes, adultos y ancianos), y asignar a cada uno de estos grupos un subconjunto de los valores de la variable (niños 0-15 años; jóvenes 16-30 años; adultos 31-64 años; ancianos $\geqslant$ 65 años).

Una vez contados los casos que pertenecen a cada categoría [frecuencias absolutas (f)], para facilitar la comparación entre diferentes poblaciones calculamos las frecuencias relativas, es decir, la proporción de cada subconjunto sobre el total de los datos (N):

$$fr = \frac{f}{N}$$

También podemos convertirlo en porcentajes: $fr\% = fr \cdot 100$

Otra variante de interés de las distribuciones anteriores son las frecuencias (absolutas o relativas) acumuladas, que podremos utilizar si trabajamos con variables de intervalo u ordinales. Para el cálculo de las frecuencias acumuladas sumamos los valores de la distribución anteriores a cada categoría.

En resumen, podemos retener los siguientes conceptos:

Frecuencia: Número de sucesos que manifiesta un fenómeno.

Distribución: Cuando un fenómeno se divide en categorías o clases, forma que adopta la repetición de sucesos.

Distribución de frecuencias: Forma que adopta un suceso que se repite y que previamente se ha organizado en categoría o clases (intervalos de frecuencia).

PROBLEMAS

PROBLEMA 1.3.

1. Planteamiento

Los miembros de una cooperativa de ganaderos tienen las siguientes edades:

56	66	21	63	60	60	60	42	38	35	45	44	57	51	57
45	38	49	50	49	47	35	30	53	32	53	42	47	41	28
47	61	35	33	48	40	38	54	63	21	20	27	53	55	41
54	28	39	45	34	39	21	42	39	27	48	48	43	35	33
27	27	54	52	38	29	30	53	53	58	23	41	63	28	43
56	59	60	40	24										

Queremos organizar la información agrupándola en diversas categorías, de manera que se facilite el análisis de la información.

Una de las posibles soluciones es la siguiente:

Edad	n
20-29	14
30-39	17
40-49	22
50-59	18
60-69	9
Total	**80**

La agrupación de los datos en categorías dependerá de los objetivos del estudio. Por ejemplo, si en el presente caso estuviésemos interesados en conocer el colectivo susceptible de acogerse a las medidas de prejubilación agraria, deberíamos separar el grupo de 55-59 años.

En la distribución anterior, calcular las frecuencias relativas, acumuladas y porcentajes.

a) Indicar el nivel de medición de la variable edad.
b) Indicar el recorrido de la variable edad.
c) Indicar el porcentaje de ganaderos menores de 40 años, el porcentaje de mayores de 50 años.
d) Indicar las marcas de clase.

2. *Cálculo*

Edad	n	fr	%	fra	% acumulado
20-29.....	14	0,175	17,5	0,175	17,5
30-39.....	17	0,212	21,2	0,387	38,7
40-49.....	22	0,275	27,5	0,662	66,2
50-59.....	18	0,225	22,5	0,887	88,7
60-69.....	9	0,113	11,3	1,000	100,0
Total.....	**80**	**1,000**	**100,0**		

a) El nivel de medición utilizado es de intervalo, ya que la variable edad es una variable continua. Puede parecer que esta medida, como sólo se indican las edades enteras, se utiliza como escala de tipo discreto. Sin embargo, hay que tener en cuenta los límites del intervalo que comprende. Por ejemplo, la categoría 30-39 años comprende las edades entre 30 y 39,99 años.

b) El recorrido de la variable edad será la diferencia entre el valor máximo y el valor mínimo que toma, en este caso 66 es la mayor edad y 20 la menor, por tanto el recorrido = 66 − 20 ≃ 46 años.

c) El porcentaje de menores de 40 años nos lo indica rápidamente la columna de los porcentajes de frecuencias acumuladas. El 38,7% de los ganaderos son menores de 40 años. El porcentaje de mayores de 50 años podríamos calcularlo obteniendo las frecuencias acumuladas en sentido descendente. De todas formas, si conocemos las frecuencias o porcentajes acumulados en un sentido podemos rápidamente convertirlos mediante fa = 1−fa. Así, si los menores de 50 años son el 66,2%, los mayores de esta edad serán 100−66,2 = 33,8%.

d) Las marcas de clase corresponden al punto medio del intervalo. Su valor es la semisuma del límite superior e inferior del intervalo.

Edad		Xc
20-29	20 + 30/2 =	25
30-39	30 + 40/2 =	35
40-49	40 + 50/2 =	45
50-59	50 + 60/2 =	55
60-69	60 + 70/2 =	65

1.3. Representación gráfica

INTRODUCCIÓN

Las técnicas de representación gráfica son otro instrumento para reducir la información original. Mediante la representación gráfica la pérdida de información es mayor, aunque presenta la ventaja de establecer diferencias entre dos distribuciones a simple vista. Los ejercicios propuestos se centran en las diferentes posibilidades que tiene el investigador para graficar la información, sin entrar en los problemas de realización de gráficos divulgativos.

Para realizar un gráfico se deberá distinguir si la variable es de tipo continuo o discreto, y si se trata de distribuciones simples o acumuladas, para seleccionar correctamente el tipo de gráfico.

La correcta realización de un gráfico necesita que se incluya en el mismo la fuente origen de los datos que permita consultas directas a la información, la escala de medida de los datos (porcentajes, miles, absolutos...), el año o período de referencia de la información y un título en el que se indique claramente la variable a que corresponde la información y el ámbito de la misma.

PROBLEMAS

PROBLEMA 1.4.

1. Planteamiento

En 1987, el calendario de la nupcialidad fue el siguiente:

	Matrimonios (miles)	Acumulados
Enero	9,9	9,9
Febrero	9,3	19,2
Marzo	13,6	32,8
Abril	14,9	47,7
Mayo	19,8	67,5
Junio	18,0	85,5
Julio	22,4	107,9

	Matrimonios (miles)	Acumulados
Agosto	27,2	135,1
Septiembre	24,7	159,8
Octubre	23,2	183,0
Noviembre	11,2	194,2
Diciembre	15,9	210,1

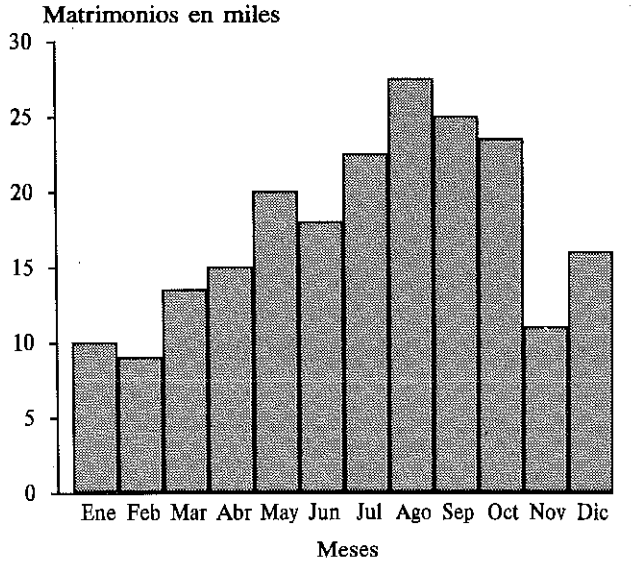
Fuente: Movimiento Natural de Población. 1987. INE.

Representar gráficamente las series de valores absolutos y acumulados.

2. *Cálculo*

Para la representación gráfica en este caso elegimos un histograma, ya que la variable está medida en una escala de intervalo (0-30 días) (30-60 días)...

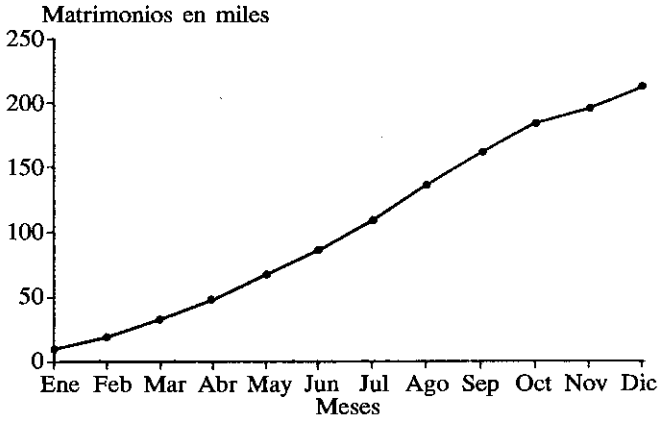
EVOLUCIÓN DE LA NUPCIALIDAD. 1987



Fuente: Movim. Natural Población. INE. 1987.

La serie acumulada la representamos en un polígono de frecuencias.

EVOLUCIÓN DE LA NUPCIALIDAD. 1987 (VALORES ACUMULADOS)



Fuente: Movim. Natural Población. INE. 1987.

PROBLEMA 1.5.

1. Planteamiento

Se quiere representar gráficamente la estructura de las viviendas familiares según su número de habitaciones.

VIVIENDAS FAMILIARES SEGÚN NÚMERO DE HABITACIONES

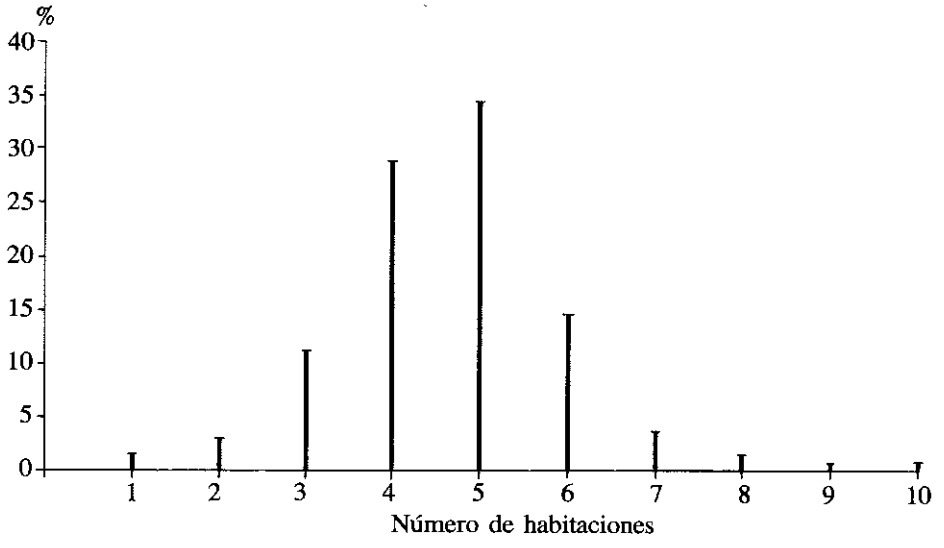
Número de habitaciones	Viviendas (miles)	%	% acumulados
1	42	0,4	0,4
2	295	2,9	3,3
3	1.164	11,3	14,6
4	2.978	29,0	43,6
5	3.537	34,4	78,0
6	1.520	14,8	92,8
7	402	3,9	96,7
8	187	1,8	98,5
9	71	0,7	99,2
10	82	0,8	100,0
Total	**10.278**	**100,0**	

Fuente: Censo de Viviendas. 1981. INE.

2. *Cálculo*

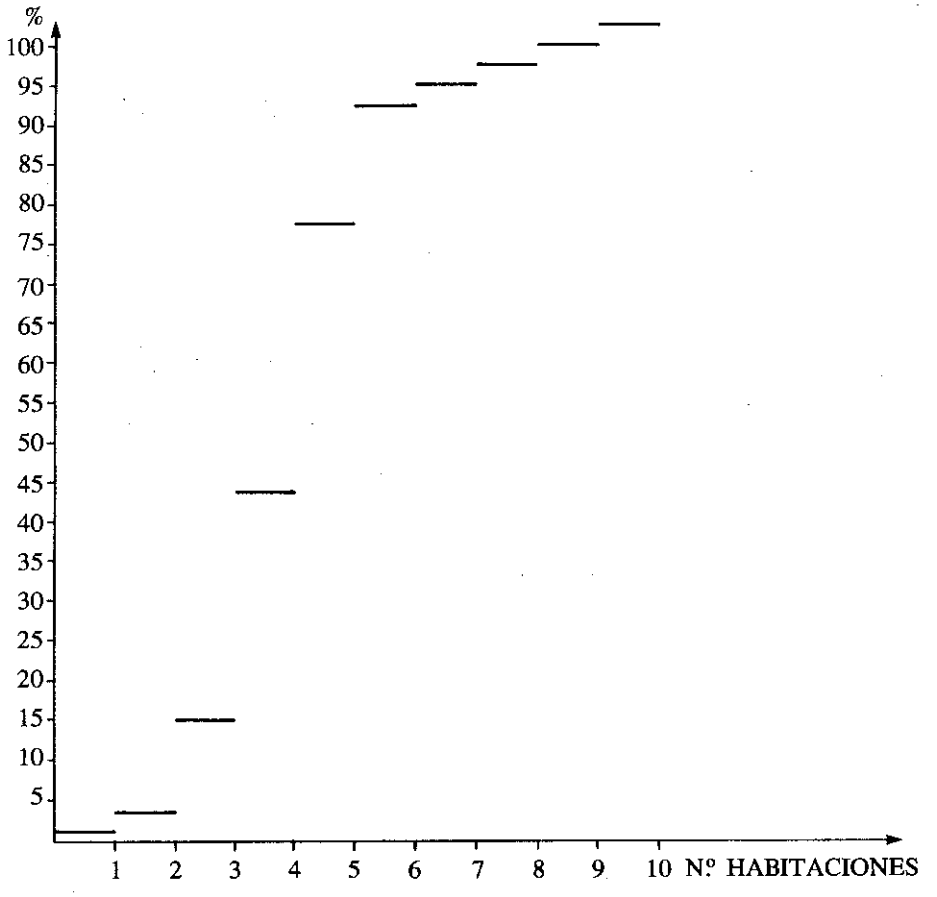
Como se trata de una variable discreta, sólo toma valores enteros y la representación gráfica será mediante un diagrama de barras. La representación de los porcentajes acumulados se realizará mediante un diagrama de escalera.

ESTRUCTURA DE LAS VIVIENDAS FAMILIARES
SEGÚN NÚMERO DE HABITACIONES



Fuente: Censo de viviendas.

ESTRUCTURA DE LAS VIVIENDAS FAMILIARES
SEGÚN NÚMERO DE HABITACIONES
(PORCENTAJES ACUMULADOS)



Fuente: Censo de viviendas. 1981. INE.

PROBLEMA 1.6.

1. Planteamiento

La distribución según la edad de la población española según los resultados de la rectificación padronal de 1987 es la siguiente:

Edad	Población
< 16	9.294.270
16-24	5.834.682
25-44	10.152.992
45-64	8.501.977
⩾ 65	4.689.410
Total	**38.473.331**

Representar dicha distribución gráficamente.

2. *Cálculo*

La representación gráfica más adecuada es un histograma. En primer lugar, calculamos los porcentajes de cada cohorte o grupo de edad, que constituyen la proporción en superficie que ocupará cada uno de ellos dentro del gráfico.

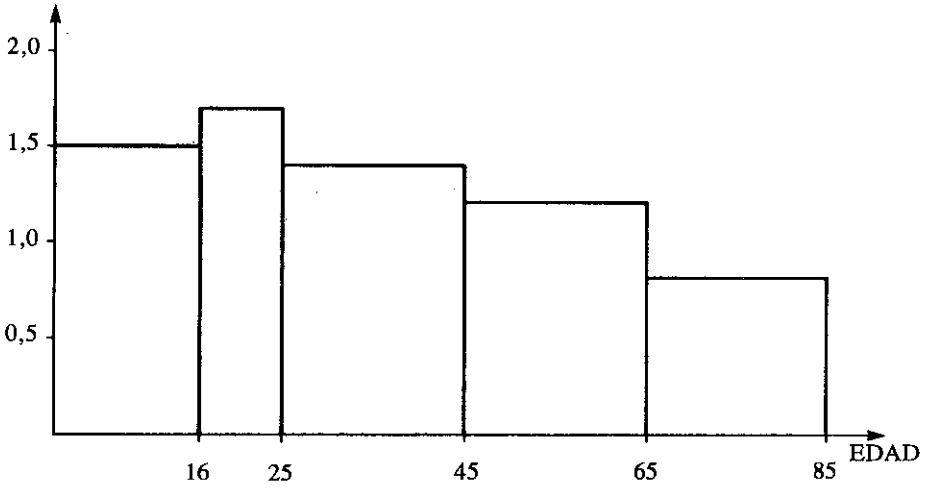
Como los intervalos no son iguales, no podremos utilizar este porcentaje como altura. En el histograma la representación de los datos es proporcional a la superficie. Por lo tanto, como las bases de los polígonos son diferentes, deberemos calcular las alturas de los mismos de tal manera que la superficie sea proporcional al porcentaje.

La altura de un polígono es igual a $h = S/b$, siendo h la altura, b la base y S la superficie.

Edad	Población	% (S)	b	h = S/b
< 16	9.294.270	24,2	16	1,5
16-24	5.834.682	15,2	9	1,7
25-44	10.152.992	26,4	20	1,3
45-64	8.501.977	22,1	20	1,1
⩾ 65	4.689.410	12,2	20	0,6
Total	**38.473.331**	**100,0**		

El intervalo abierto > 65 años podemos cerrarlo en 84 años (límite superior 85). La esperanza media de vida a los 65 años es de 16,52 años.

POBLACIÓN ESPAÑOLA SEGÚN EDAD. 1987



Fuente: Rectificación Padrón Municipal de Habitantes. 1987. INE.

PROBLEMA 1.7.

1. Planteamiento

Según el Censo de Población de 1981, la población activa según los diferentes sectores de actividad era:

POBLACIÓN ACTIVA
SEGÚN SECTORES ECONÓMICOS. 1981
(EN MILES)

Sectores Económicos	Mujeres	Varones
Agricultura	247,2	1.607,2
Industria	667,7	2.666,8
Construcción	26,7	1.232,8
Servicios	1.778,9	3.545,1
Total	**2.720,5**	**9.051,9**

Fuente: Censo de Población 1981. INE.

Representar gráficamente estos resultados de manera que puedan apreciarse las diferentes estructuras ocupacionales de mujeres y varones.

2. Cálculo

Un tipo de representación adecuado puede ser el círculo de sectores o ciclograma, ya que las categorías son pocas.

Primeramente calculamos los porcentajes.

POBLACIÓN ACTIVA SEGÚN SECTORES ECONÓMICOS. 1981 (EN PORCENTAJES)

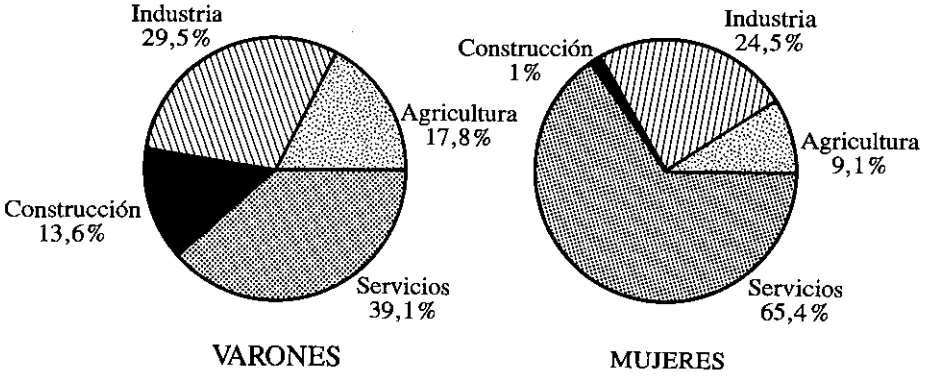
Sectores Económicos	Mujeres	Varones
Agricultura	9,1	17,8
Industria	24,5	29,5
Construcción	1,0	13,6
Servicios	65,4	39,1
Total	**100,0**	**100,0**

Como la circunferencia tiene 360° a cada 1% le corresponderán 3,6°. Por lo tanto, multiplicando los porcentajes obtenidos por 3,6°, obtendremos el ángulo que determinará el sector de cada categoría.

POBLACIÓN ACTIVA SEGÚN SECTORES ECONÓMICOS. 1981 (EN GRADOS)

Sectores Económicos	Mujeres	Varones
Agricultura	32,8	63,7
Industria	88,2	106,2
Construcción	3,6	49,0
Servicios	235,4	141,1
Total	**360,0°**	**360,0°**

POBLACIÓN ACTIVA SEGÚN SECTOR



Fuente: Censo de población. 1981. INE.

PROBLEMA 1.8.

1. Planteamiento

A partir de los resultados del Censo Agrario de 1982 hemos obtenido para las diferentes Comunidades Autónomas los siguientes resultados referentes al estatus socioprofesional de la mujer en la actividad agraria. Deseamos hacer una representación gráfica que ponga en evidencia las similitudes y disparidades entre las diferentes CCAA.

ESTATUS SOCIOPROFESIONAL DE LA MUJER EN LA AGRICULTURA (EN PORCENTAJES)

CCAA	Empresarias	Ayudas familiares	Trabajadoras asalariadas
Andalucía	13	29	58
Aragón	48	36	16
Asturias	41	56	3
Baleares	47	20	33
Canarias	12	11	77
Cantabria	40	59	1
Castilla-León	32	64	4
Castilla-La Mancha	26	30	44
Cataluña	35	39	26
Extremadura	17	48	35
Galicia	36	60	4

ESTATUS SOCIOPROFESIONAL DE LA MUJER EN LA AGRICULTURA (EN PORCENTAJES) (CONT.)

CCAA	Empresarias	Ayudas familiares	Trabajadoras asalariadas
Madrid	30	9	61
Murcia	8	14	78
Navarra	27	62	11
País Vasco	56	36	8
Rioja	25	64	11
Valencia	14	17	69

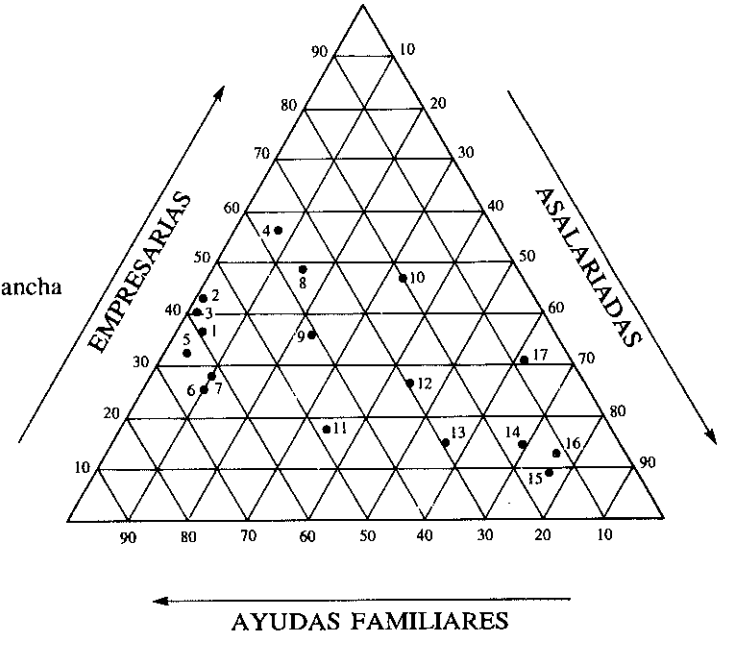
Fuente: Censo Agrario 1982. INE.

2. *Cálculo*

Como se trata de una clasificación en tres categorías, podemos representar dicha distribución en un gráfico triangular.

ESTATUS SOCIOPROFESIONAL DE LAS AGRICULTORAS. 1982

1. Galicia
2. Asturias
3. Cantabria
4. País Vasco
5. Castilla-León
6. Rioja
7. Navarra
8. Aragón
9. Cataluña
10. Baleares
11. Extremadura
12. Castilla-La Mancha
13. Andalucía
14. Valencia
15. Murcia
16. Canarias
17. Madrid



Fuente: Censo Agrario. 1982. INE.