Pronósticos y Estadística para la Administración

Manuel Mendoza Ramírez
Departamento de Estadística y
Centro de Estadística Aplicada
ITAM, México.

Contenido

Capítulo 1

Introducción

Introducción

El término pronóstico es de uso común en el lenguaje cotidiano. En prácticamente cualquier diccionario se puede encontrar alguna de las siguientes definiciones.

pronóstico

I. Del latín *prognosticum* < gr. *prognostikon* < *progignosko* = yo conozco de antemano.

1. (sustantivo masculino). Acción y efecto de pronosticar.

2. (sustantivo masculino). Predicción de los fenómenos meteorológicos.

3. (sustantivo masculino). Predicción del médico acerca de los cambios que pueden sobrevenir durante una enfermedad, y sobre la duración y término de la misma, por los síntomas detectados.

4. (sustantivo masculino). Señal o presagio.

predicción

I. Del latín *praedictio, -onis*.

1. (sustantivo femenino). Acción y efecto de predecir o anunciar un hecho que se producirá en el futuro.

2. (sustantivo femenino). Palabras con que se predice y cosa predicha.

presagio

I. Del latín *praesagium* = conocimiento anticipado.

1. (sustantivo masculino). Señal externa o estado anímico que vaticina un suceso futuro.

2. (sustantivo masculino). Adivinación del futuro por indicios o presentimientos.

   SIN. 1. Indicio, barrunto, síntoma. 2. Pronóstico, predicción, anuncio.

pronosticar

I. De pronóstico.

1. (verbo transitivo). Prever lo futuro.
2. (verbo transitivo). Predecir, manifestar lo que va a suceder basándose en ciertos indicios.
3. (verbo transitivo). Emitir un médico su pronóstico.

predecir

I. Del lat. praedicere < *prae* = antes + *dicere* = decir.

1. (verbo transitivo). Anunciar un hecho que ocurrirá en el futuro.

FAM. Predicción, predicho, -a.
SIN. Vaticinar, pronosticar, presagiar, conjeturar.
OBS. v.irreg.; modelo decir.

Para los propósitos de este curso, un pronóstico es una afirmación sobre un evento cuya ocurrencia no es segura.

Típicamente, los pronósticos se producen sobre eventos que pueden ocurrir en el futuro.

Los fenómenos que dan origen a los eventos que son objeto de pronósticos se conocen con el nombre de fenómenos inciertos (sus resultados no se pueden anticipar con certeza).

En virtud de que los eventos para los cuales se producen pronósticos no se pueden anticipar con certeza, una característica intrínseca de todo pronóstico es que puede fallar.

La incertidumbre asociada a los fenómenos inciertos puede provenir de, al menos, dos fuentes distintas.

La falta de conocimiento.
La falta de control.

La falta de control se manifiesta a través de la variabilidad de los resultados observados.

La producción de pronósticos está inevitablemente asociada con descripción de la incertidumbre.

Producir pronósticos es relativamente fácil.

Producir pronósticos que acierten con frecuencia es difícil.

Producir pronósticos que acierten siempre es imposible.

Medir el grado de confiabilidad de un pronóstico, después de la eventual ocurrencia del evento relevante, es fácil.

Medir el grado de confiabilidad de un pronóstico, antes de la eventual ocurrencia del evento relevante, es difícil.

Existen diversos tipos de pronósticos, por ejemplo:

mágicos,
cualitativos y
cuantitativos (estadísticos)

Un pronóstico ideal debe incluir:

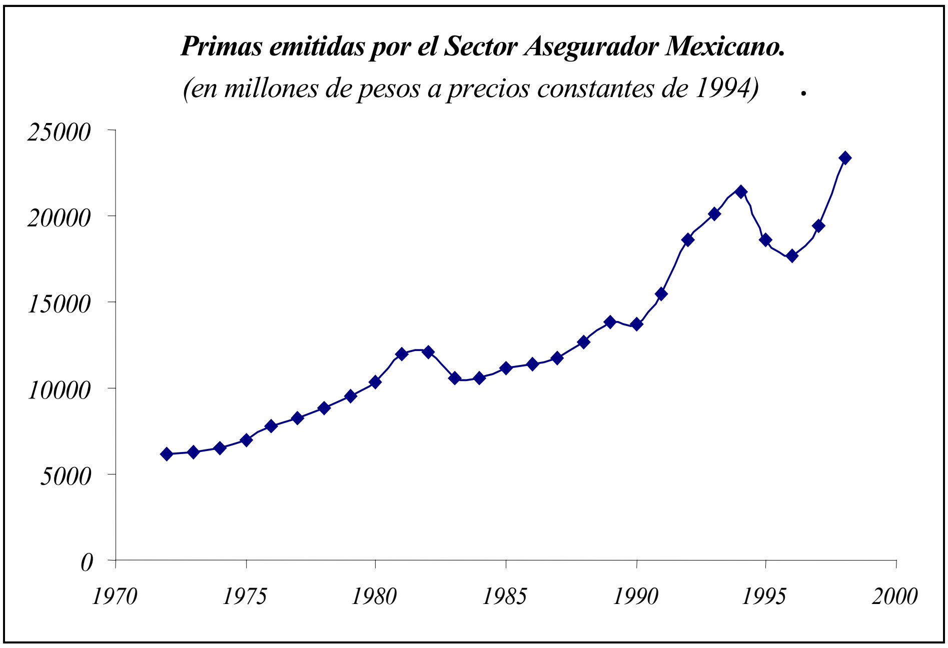
Una medida de su precisión,
Una medida de su confiabilidad (previa)

Además, debe obtenerse con un mecanismo

Reproducible

Las técnicas estadísticas de pronóstico no solamente reúnen estas tres características sino que constituyen una valiosa herramienta para la producción de pronósticos a partir del análisis de información previa.

Ejemplo 1.

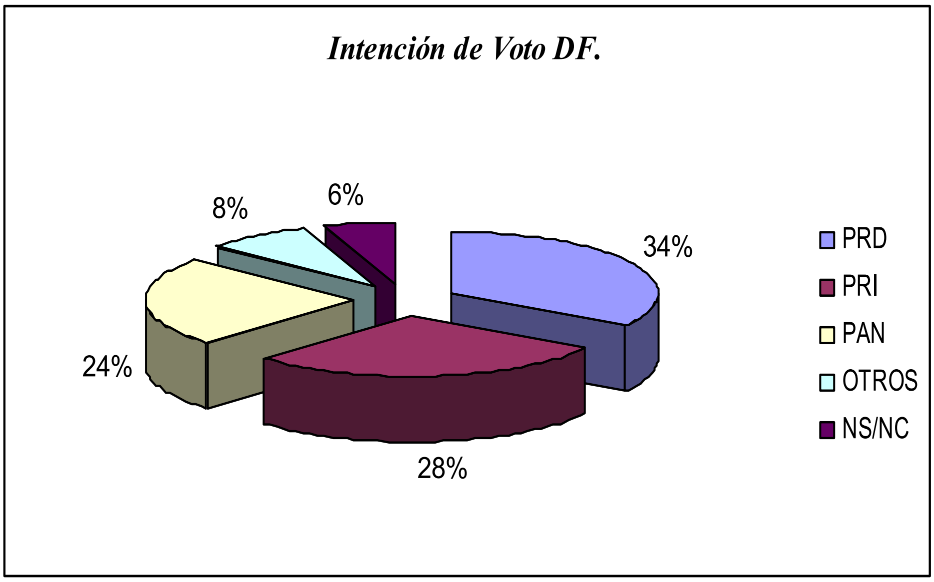


Año	Millones pesos
1990	13,686
1991	15,416
1992	18,630
1993	20,085
1994	21,370
1995	18,553
1996	17,691
1997	19,427
1998	23,379

¿Cuál sería un pronóstico razonable para 1999?

¿Con qué margen de error?
¿Con qué grado de confiabilidad?

Ejemplo 2



¿Cómo podría utilizar estos datos para producir un pronóstico?

¿Cuál sería ese pronóstico?

¿Con qué margen de error?

¿Con qué grado de confiabilidad?

Las técnicas estadísticas para la producción de pronósticos operan de acuerdo a reglas generales que, en esencia, se pueden resumir a través del siguiente algoritmo.

1. Se recolectan observaciones sobre el fenómeno.
2. Se describe el comportamiento de las observaciones.
3. Se adoptan supuestos de carácter general sobre el comportamiento de las observaciones.
4. Se establecen supuestos sobre la relación que guardan las observaciones futuras con las observaciones que se han recolectado.
5. Se describen el comportamiento futuro del fenómeno. Es decir, se producen los pronósticos cada uno de los cuales incluye una medida de su confiabilidad.

Este algoritmo se complementa con una etapa más de verificación o contraste del pronóstico. Esta etapa se lleva a cabo cuando la incertidumbre sobre la ocurrencia del evento objeto del pronóstico desaparece.

En esas condiciones, el resultado del evento se compara con el pronóstico y de esa comparación se pueden sugerir modificaciones al procedimiento de producción de los pronósticos.

Estas modificaciones, en general, afectan los supuestos que se adoptan en el proceso y tienen por objeto describir mejor tanto las observaciones disponibles como la relación de éstas con las observaciones futuras, objeto de los pronósticos.

Cuando, como resultado de la etapa de contraste, se dan modificaciones el algoritmo debe ponerse a prueba nuevamente desde el principio.

La repetición de este ciclo es la base de la mejora y adaptación continua del proceso de producción de pronósticos.

En cualquier caso, el proceso completo de producción de pronósticos estadísticos, tal como se ha indicado, se lleva a cabo a partir de la información que se recolecta sobre el fenómeno e inevitablemente requiere de supuestos.

Ahora bien, los supuestos en un proceso de análisis estadístico, como el que se emplea para producir pronósticos, se incorporan a través de lo que se conoce como modelo.

Un modelo es una descripción aproximada de la realidad.

Los modelos se utilizan con el propósito de analizar los aspectos relevantes de la realidad en condiciones simplificadas.

Así, los modelos destacan los elementos que se consideran de importancia en el estudio de un fenómeno y simultáneamente, toman en cuenta las relaciones que existen entre estos elementos a la vez que describen de forma general todos los otros aspectos del fenómeno.

Existen diferentes tipos de modelos. El Análisis Estadística hace uso preferentemente de los llamados modelos simbólicos. Esta es una clase de modelos muy utilizada en Matemáticas donde los distintos elementos del fenómeno bajo estudio se representan por medio de símbolos y sus relaciones se establecen a través de funciones.

Ya nos ocuparemos más delante de los modelos estadísticos para la producción de pronósticos pero por el momento es conveniente puntualizar que, no importa la clase a la que pertenezca, las características más importantes de todo modelo son las siguientes:

Simplicidad
Capacidad descriptiva

La simplicidad se refiere a la propiedad de que el modelo pueda ser utilizado y sus resultados puedan ser interpretados y analizados sin dificultad.

La capacidad descriptiva, por su parte, es la propiedad con que debe contar el modelo para reproducir las manifestaciones del fenómeno bajo estudio con fidelidad. En particular esta propiedad se comprueba, en muchos casos, mediante la producción de pronósticos. Un modelo con esta propiedad, produce buenos pronósticos.

Otra propiedad que complementa estas dos es la generalidad.

Un modelo cuenta con esta propiedad cuando reúne tanto la simplicidad como la capacidad descriptiva para toda una clase de fenómenos similares y no para uno solo.

Ejemplo 3

Suponga que se le pide producir un pronóstico sobre el sexo (hombre o mujer) de la última persona del grupo que ingrese a la sala la próxima sesión de este curso.

¿Cuál sería su pronóstico?

¿Qué información tomaría en cuenta?

¿Cómo la utilizaría?

¿Cuál sería el grado de confiabilidad del pronóstico?

Conjunto de Datos 1

Base de datos de 200 empleados de una compañía

Hanke, J.E. & Reitsh, A.G. (1995). Estadística para Negocios.
Madrid: Irwin. Apéndice C.

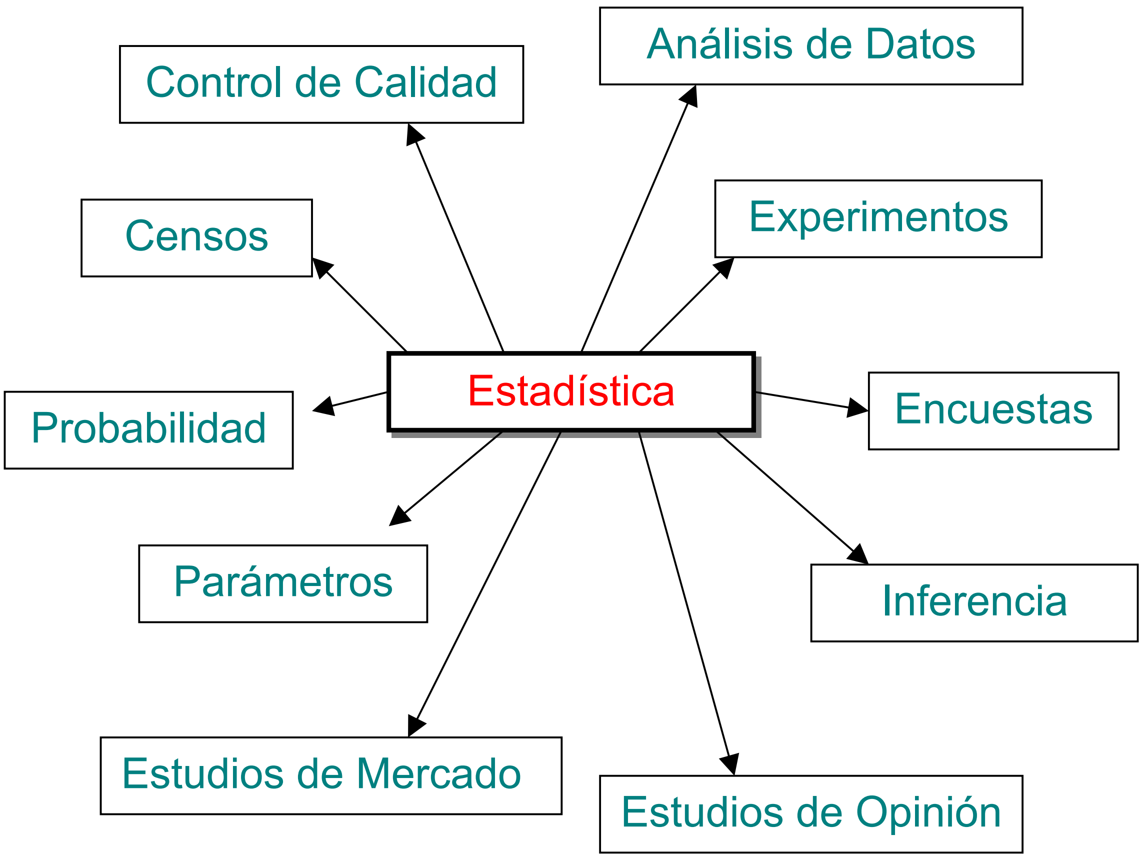
	antigüedad	horas extra	sexo	cursos	incapacidad	aptitudes	escolaridad	salario	edad
n	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	11	125	1	4	9	121.89	2	23065	44
2	24	225	2	2	2	114.20	1	27180	50
3	17	115	2	3	5	134.11	1	34875	48
4	9	117	1	1	1	113.95	1	23685	53
5	15	26	1	2	0	151.41	2	33550	62
6	6	43	1	4	3	96.65	1	22635	45
7	4	124	2	2	4	98.43	2	19575	26
8	2	71	2	1	1	110.06	1	20430	28
9	17	166	2	2	5	101.98	1	18955	33
10	17	158	1	3	2	101.01	1	25595	40
11	15	182	2	4	4	103.42	2	34975	63
12	21	81	2	3	6	106.88	2	26800	55
13	4	58	1	2	5	99.36	2	22400	50
14	12	203	1	2	3	105.66	2	31200	33
15	23	144	1	2	4	100.91	1	24750	41
16	20	179	1	3	5	73.76	2	30495	53
17	19	96	2	1	5	83.39	0	33965	58
18	12	96	2	4	7	88.41	1	30440	51

1. Describa el comportamiento de este grupo de personas en lo que se refiere a género (sexo).

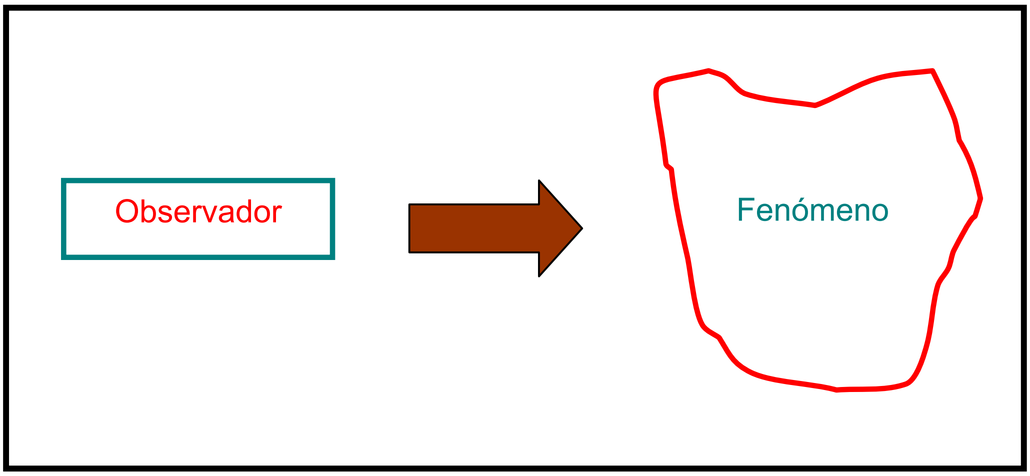
2. Describa este grupo de personas en lo que se refiere a su nivel de escolaridad.

3. Describa este grupo de personas en lo que se refiere a su edad.

La Naturaleza del Análisis Estadístico



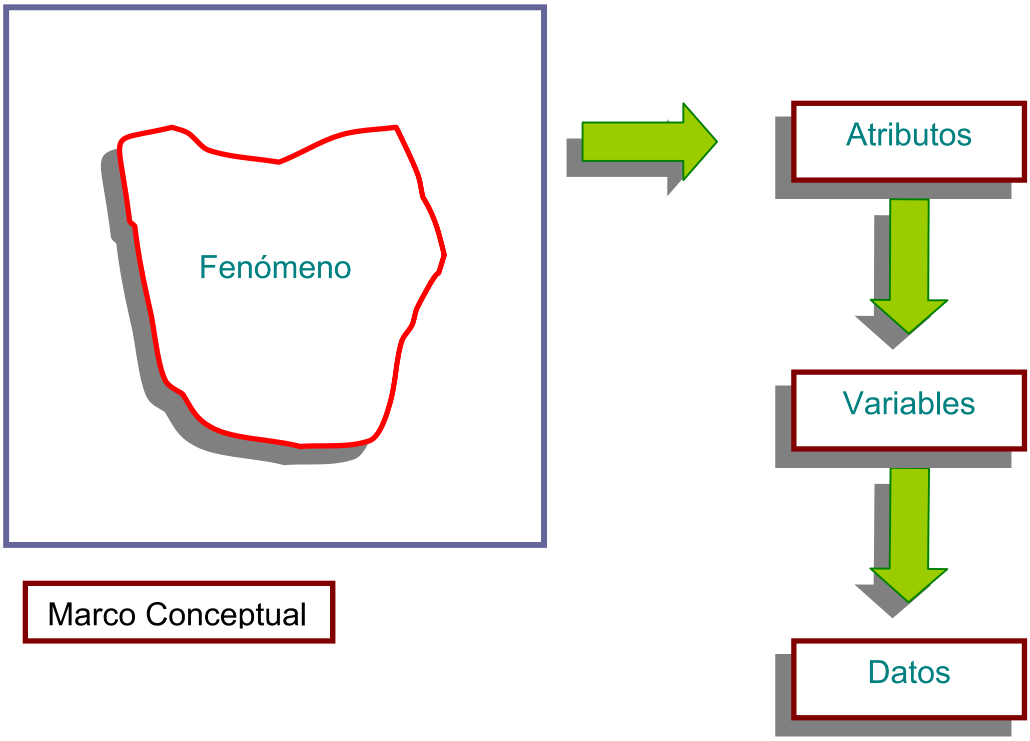
Objeto: Estudio de Fenómenos Inciertos (Aleatorios).



Fenómeno Incierto: Aquél cuyos resultados no se pueden predecir con certeza.

El estudio se realiza a partir del conocimiento previo y de observaciones que se realizan sobre el fenómeno.

El Proceso de Observación.



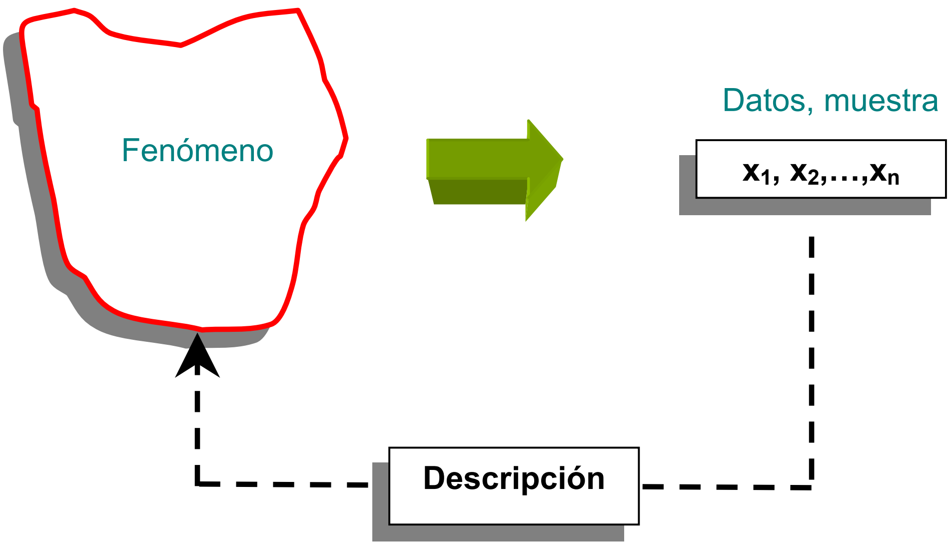
Atributo: Manifestación de interés.

Variable: Codificación numérica de un atributo.

Dato: Registro (numérico) de la observación de un atributo a través de una variable (valor concreto).

Variables: X, Y, Z, W,…

Datos: x, y, z, w,…



Estadística: Familia de técnicas para describir un fenómeno, a partir de un conjunto de datos que presenta variabilidad.

Conclusión: Toda la Estadística es descriptiva!

Casos:

I. Se cuenta con todos los datos posibles del fenómeno.

Descripción Exacta

Análisis Exploratorio de Datos

II. Se cuenta con una parte sólo de todos los datos posibles.

Descripción Aproximada

Inferencia Estadística

Capítulo 2

Análisis Exploratorio de Datos

Análisis Exploratorio de Datos

- ✓ Todos los datos, objeto de un análisis estadístico, son numéricos.
- ✓ Esto ocurre así debido a que todas las variables en un estudio estadístico son codificaciones numéricas.
- ✓ Es necesario, sin embargo, reconocer que no todos los datos son intrínsecamente numéricos.
- ✓ Existen diversas clases de variables que dan lugar a los correspondientes tipos de datos.

Datos (Variables)

- Cualitativos
  - Nominales
  - Ordinales
- Cuantitativos
  - Discretos
  - Continuos

Datos Cualitativos. Son aquellos que no son intrínsecamente numéricos.

- De Tipo Nominal. Los números en la codificación funcionan únicamente como etiquetas (nombres).

  Ejemplo: Sexo en el conjunto de datos 1 (1=mujer, 2=hombre).

- De Tipo Ordinal. Los números en la codificación son etiquetas pero los valores se asignan de acuerdo a un orden que contiene información sobre la intensidad del atributo.

  Ejemplo: Escolaridad en el conjunto de datos 1 (0=bachillerato, 1=lic. s/tit., 2=lic. c/tit.)

Datos Cuantitativos. Son aquellos que sí son intrínsecamente numéricos.

- De Tipo Discreto. Solamente pueden producirse valores **aislados** y, con frecuencia, describen conteos.

  Ejemplo: Incapacidad en el cd1: (días de incap. en los últimos 6 meses)

- De Tipo Continuo. Se puede producir cualquier valor en un **intervalo**.

  Ejemplo: ¿Aptitudes?, ¿Salario?

El Insumo del Análisis Exploratorio de Datos

La materia prima del Análisis Exploratorio de Datos (AE) son los propios datos y estos se organizan en Bancos de Datos.

Columnas = Variables

Hanke, J.E. & Reitsh, A.G. (1995). Estadística para Negocios.
Madrid: Irwin. Apéndice C.

Renglones = Casos

	antigüedad	horas extra	sexo	cursos	incapacidad	aptitudes	escolaridad
n	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	11	125	1	4	9	121.89	2
2	24	225	2	2	2	114.20	1
3	17	115	2	3	5	134.11	1
4	9	117	1	1	1	113.95	1
5	15	26	1	2	0	151.41	2
6	6	43	1	4	3	96.65	1
7	4	124	2	2	4	98.43	2
8	2	71	2	1	1	110.06	1
9	17	166	2	2	5	101.98	1
10	17	158	1	3	2	101.01	1
11	15	182	2	4	4	103.42	2
12	21	81	2	3	6	106.88	2
13	4	58	1	2	5	99.36	2
14	12	203	1	2	3	105.66	2
15	23	144	1	2	4	100.91	1
16	20	179	1	3	5	73.76	2
17	19	96	2	1	5	83.39	0
18	12	96	2	4	7	88.41	1
19	5	157	2	4	8	98.19	1
20	11	27	2	2	4	101.72	1

El Propósito del Análisis Exploratorio de Datos

El objetivo general que persigue el AE, cuando se enfrenta a un banco de datos, es describir la información contenida en el banco.

Estas dos columnas provienen del conjunto de datos 1 y constituyen solamente una pequeña fracción del banco (1 de 9 variables y 68 de 200 casos).

	sexo
n	X_3
1	1
2	2
3	2
4	1
5	1
6	1
7	2
8	2
9	2
10	1
11	2
12	2
13	1
14	1
15	1
16	1
17	2
18	2
19	2
20	2
21	1
22	2
23	1
24	1
25	2
26	1
27	1
28	1
29	2
30	2
31	2
32	1
33	1
34	1
35	1
36	2
37	1
38	2
39	2
40	2
41	2
42	1
43	1
44	2
45	2
46	2
47	1
48	1
49	2
50	2
51	1
52	2
53	1
54	1
55	1
56	2
57	2
58	2
59	1
60	2
61	2
62	1
63	1
64	1
65	1
66	2
67	2
68	1

No existe una descripción más completa del banco que el banco mismo.

Esta descripción, sin embargo, no es eficiente.

La solución es producir resúmenes (resumir la información contenida en el banco)

Un resumen eficiente debe satisfacer dos propiedades:

- Preservar la información relevante (Suficiente)
- Eliminar toda la información irrelevante (Minimal)

La clasificación de la información en relevante o irrelevante depende de los objetivos particulares de cada estudio.

Un resumen puede ser eficiente en un estudio y completamente inapropiado en otro.

En cualquier caso, el objetivo general del AE es la producción de resúmenes eficientes de la información contenida en un banco de datos.

Algunos propósitos más específicos del AE:

- Identificar grupos
- Establecer tendencias
- Revelar asociaciones
- Aislar observaciones atípicas

Para alcanzar sus propósitos, el AE produce básicamente resúmenes de dos tipos:

- Numéricos.
- Gráficos.

Las técnicas del AE se pueden clasificar de acuerdo con distintos criterios. En particular, con respecto al número de variables que consideran simultáneamente.

Análisis Exploratorio Multivariado.

Cuando las técnicas del AE se aplican a dos o más variables de un mismo banco de datos simultáneamente.

Análisis Exploratorio Univariado.

Cuando se aplican las técnicas del AE a una sola variable de un banco de datos.

Las técnicas estadísticas también se pueden clasificar si se toma en cuenta la naturaleza del análisis que se realiza.

Análisis Exploratorio Comparativo.

Cuando las técnicas de AE se utilizan para contrastar el comportamiento de un mismo grupo de variables en dos o más bancos de datos.

Análisis Exploratorio de Asociación.

Cuando el AE se emplea para describir la relación que guardan entre sí dos o más variables del mismo banco de datos.

En cualquier caso, la elección de un resumen apropiado (numérico o gráfico) depende de los objetivos generales y particulares de cada estudio.

El otro aspecto fundamental es que la elección de un resumen apropiado depende, también, del tipo de datos (variable) de que se trate.

Los estudios más complicados son aquellos en los que se consideran diferentes objetivos, con distintos bancos de datos y a través de los cuales se pretende describir simultáneamente el comportamiento de dos o más variables.

Por facilidad, en este curso, iniciaremos con las técnicas del Análisis Exploratorio Univariado.

El propósito es, entonces, describir el comportamiento de una variable a partir de la información correspondiente, contenida en un banco de datos.

Así, la información está formada por los datos que se encuentran en una columna del banco y la principal característica de esos datos es su variabilidad.

Datos Cualitativos Nominales

Estos datos no son intrínsecamente numéricos y los valores con que se codifican desempeñan el papel de etiquetas.

Sugiera y comente algunos ejemplos de datos (variables) de este tipo.

Los datos de este tipo, en general, toman sus distintos valores de una colección relativamente reducida de posibilidades.

Considere el ejemplo de la variable Sexo (X_3) del banco de datos 1. Ahí, se tienen 200 datos con valores 1 ó 2. (1 = mujer, 2 = hombre).

Estos datos presentan variabilidad (no todos son iguales) y su descripción completa requiere 200 números.

En una gran cantidad de aplicaciones, no interesa conocer cual individuo (caso) concreto dio origen a cada uno de los datos en el banco. Entonces, se dice que las observaciones son intercambiables.

Cuando se tienen observaciones intercambiables, la información que típicamente es relevante es la que se refiere a:

1. Los valores distintos que se presentan en el banco de datos.

2. La frecuencia con que cada uno de esos valores se presenta en el banco de datos.

Extendiendo esta idea, se puede construir un resumen con la información relativa a:

1. Los valores distintos que se podrían presentar en el banco de datos.

2. La frecuencia con que cada uno de esos valores se presenta en el banco de datos.

De esta manera, para una variable nominal, que puede producir k diferentes valores, y un banco de datos con n casos el resumen apropiado consta de k números:

$$f_1, f_2, \ldots f_k$$

donde f_i es la frecuencia (número de veces) con que se presenta el i-ésimo valor (x_i) de la variable X en el banco de datos.

Las frecuencias observadas f_1, $f_2, \ldots$ f_k satisfacen tres propiedades:

1. f_i es un número entero, para i = 1,..., k.

2. $f_i \geq 0$, para i = 1,..., k.

3. $\sum_{i=1}^{k} f_i = n$.

De esta forma, una columna con n datos se puede reemplazar por un resumen de k números. Entre más grande sea la diferencia entre n y k, más dramático será el efecto del resumen.

Por otra parte, se puede observar que el resumen

$$f_1, f_2,... f_k$$

es equivalente al que se obtendría sustituyendo una cualquiera de las frecuencias observadas por el tamaño del banco n. Por ejemplo, al resumen

$$n, f_2,... f_k.$$

Con el fin de concentrar la información obtenida del banco con la que proviene de la definición de la variable en estudio, las frecuencias observadas se suelen presentar en compañía de los valores posibles correspondientes x_1, x_2,... x_k.

El resultado se conoce con el nombre de Tabla de Frecuencias. Se trata de un arreglo en donde la información, tanto de los valores posibles como de las frecuencias observadas, se despliega por columnas.

X	F
x_1	f_1
x_2	f_2
.	.
.	.
.	.
x_k	f_k

Esta Tabla es un resumen numérico que contiene toda la información relevante y, en las condiciones que se comentan, es minimal (salvo por la información de los valores de la variable X en la primera columna).

En ocasiones, la información de este resumen se complementa con otras columnas que se pueden obtener a partir de la primera. Por ejemplo, es común que se incluya otra columna con las frecuencias expresadas en términos relativos.

X	f	fr
x_1	f_1	fr_1
x_2	f_2	fr_2
.	.	.
.	.	.
.	.	.
x_k	f_k	fr_k

Recordando que

$$\sum_{i=1}^{k} f_i = n,$$

las frecuencias relativas se definen como fr = f / n.

Las frecuencias representan la proporción de casos que aparecen en el banco de datos con el valor correspondiente de la variable bajo estudio y satisfacen la relación

$$\sum_{i=1}^{k} fr_i = 1.$$

Alternativamente, las frecuencias relativas se reportan como porcentajes [%fr = (100) $\times$ fr] y en ese caso se tiene

$$\sum_{i=1}^{k} \%fr_i = 100.$$

Tabla con frecuencias absolutas, relativas y porcentuales

X	f	fr	%fr
x_1	f_1	fr_1	$\%fr_1$
x_2	f_2	fr_2	$\%fr_2$
.	.	.	.
.	.	.	.
.	.	.	.
x_k	f_k	fr_k	$\%fr_k$
Suma	**n**	**1**	**100**

Para el caso de la variable Sexo (X_3) del conjunto de datos 1, se tiene la siguiente tabla.

X	f	fr	%fr
1 (mujer)	81	0.405	40.5
2 (hombre)	119	0.595	59.5
Suma	**200**	**1**	**100 %**

La información verdaderamente relevante en esta tabla se concentra en dos números: 81 (mujeres) y 119 (hombres).

X	f
1 (mujer)	81
2 (hombre)	119
n	**200**

La información verdaderamente relevante en esta tabla se concentra en dos números: 81 (mujeres) y 119 (hombres). De esta manera, un resumen numérico de dimensión 2 reemplaza y describe un banco de datos con 200 observaciones.

En este caso es interesante notar que un resumen numérico aún más reducido puede ser de interés aunque ya no sea suficiente. Por ejemplo, se puede resumir la información en el banco indicando que " el sexo más abundante es **hombre** con un 59.5% de los casos".

Esta forma de resumir la información en un banco de datos es bien conocida y recibe el nombre de **Moda.** En general, la (s) moda (s) de un conjunto de datos es (son) el (los) valor (es) que aparece (n) más frecuentemente en el banco.

En el ejemplo, se tiene que la moda es X = 2 (hombre) con una frecuencia relativa de 0.595.

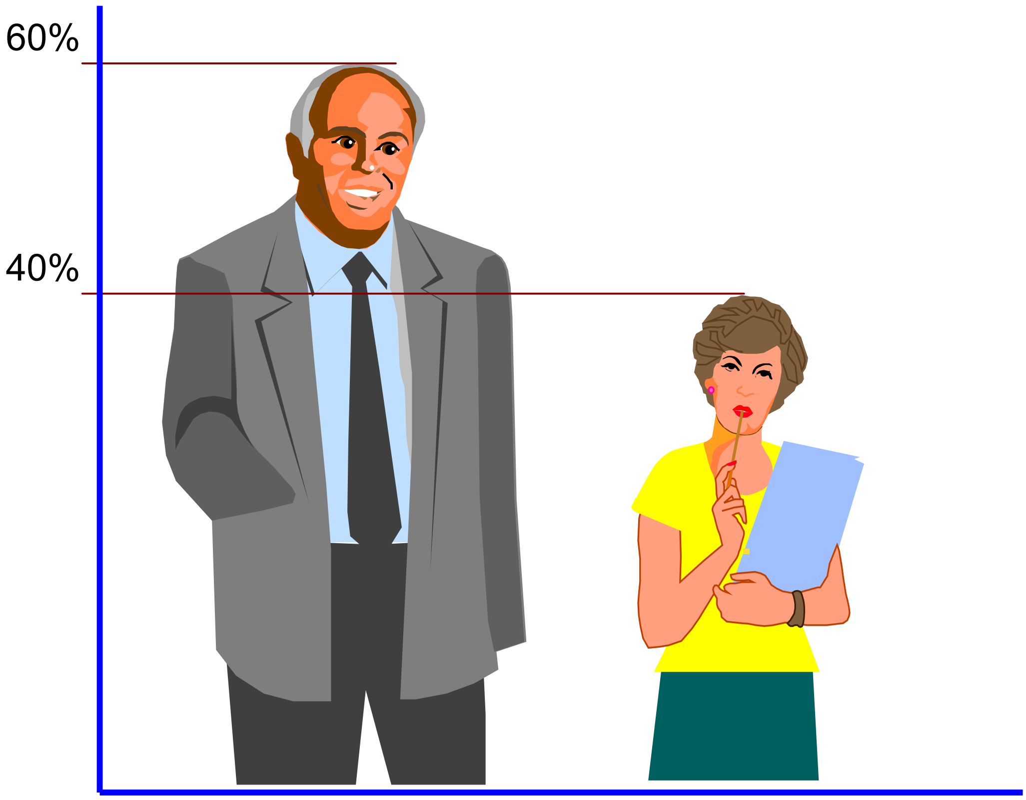
Siempre es conveniente acompañar el valor de la moda con su frecuencia relativa o porcentual debido a que, salvo en el caso de dos categorías, no es posible asegurar que la mayoría de los casos presenten el valor de la moda. Observe la siguiente tabla.

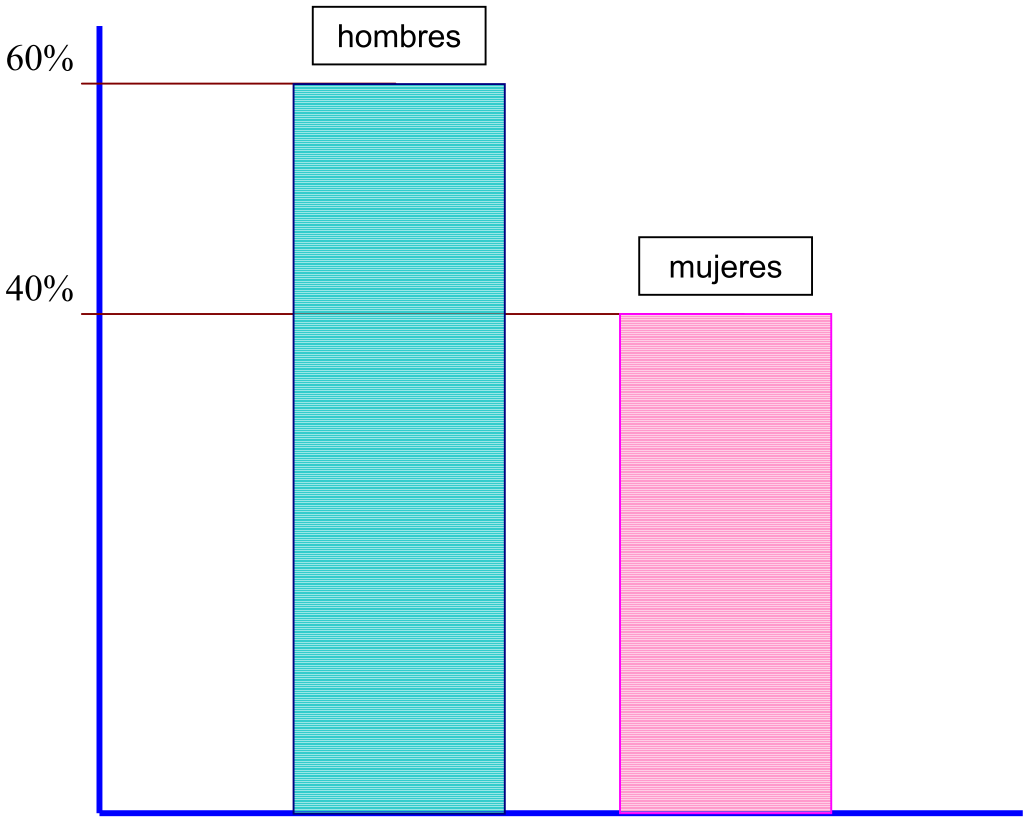
X	fr	%fr
1	20	32.8
2	20	32.8
3	21	34.4
Suma	**61**	**100.0**

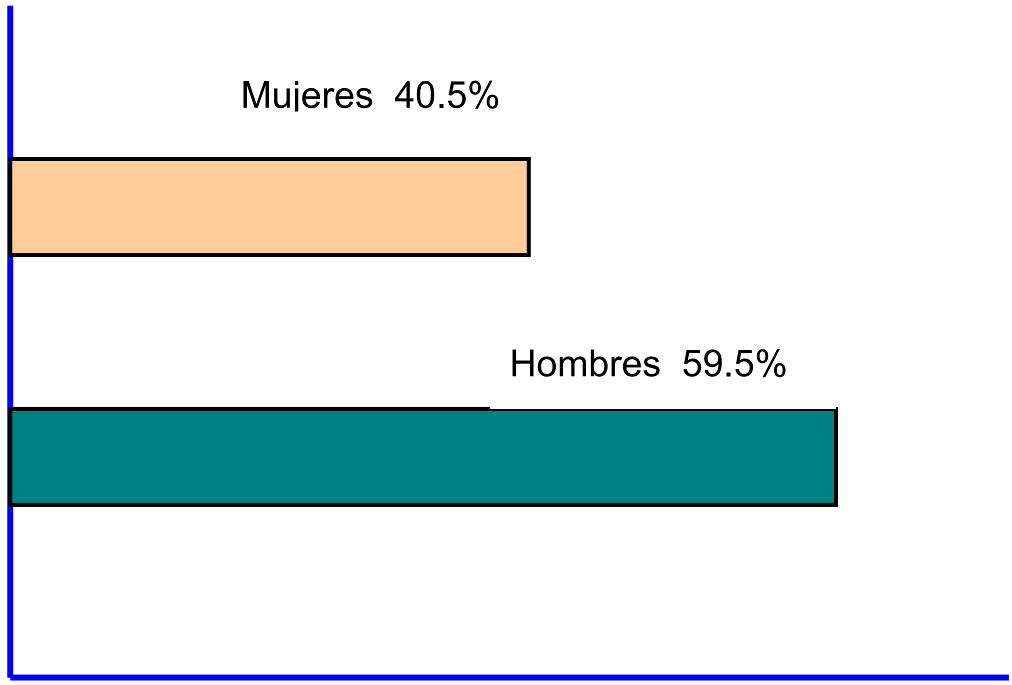
En este caso, la información relevante se refiere a como están distribuidos los casos entre los diferentes valores de X y cuantos casos hay en total en el banco. Así, los k números del resumen podrían ser n y cualesquiera k-1 de las frecuencias porcentuales (o relativas).

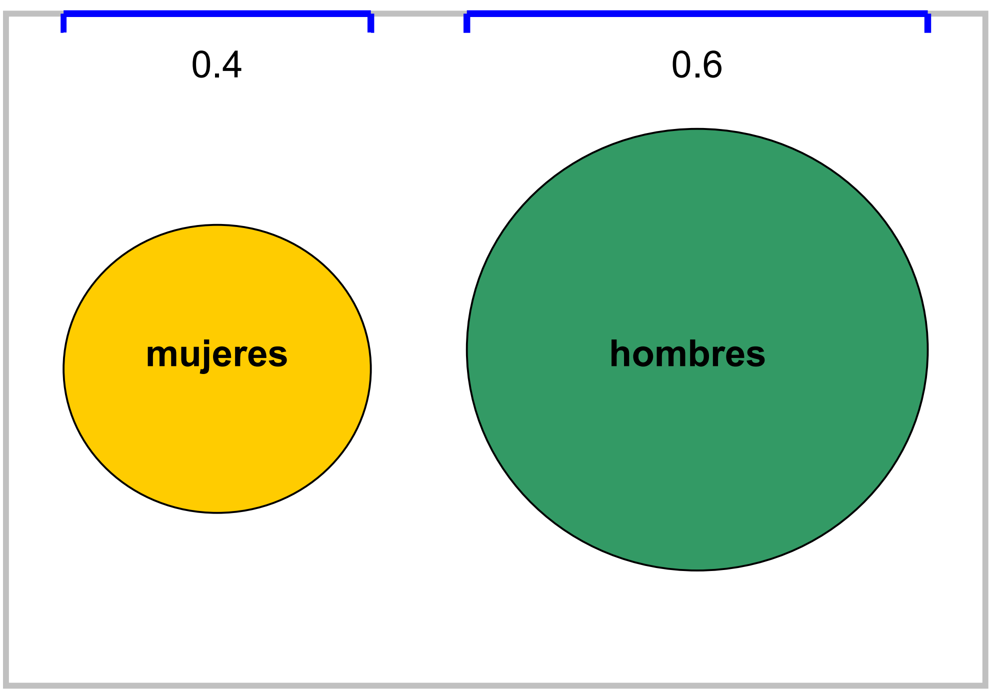
La evidencia que resalta en el resumen es la que se refiere a los valores más abundantes y, en general, la atención se concentra en la abundancia relativa de los diferentes valores.

Estas consideraciones sugieren un resumen gráfico que represente cada valor de X con una imagen cuyo tamaño sea proporcional a la frecuencia relativa correspondiente.


60%
40%


hombres
mujeres
60%
40%


Mujeres 40.5%
Hombres 59.5%


0.4
0.6
mujeres
hombres

(Areas)

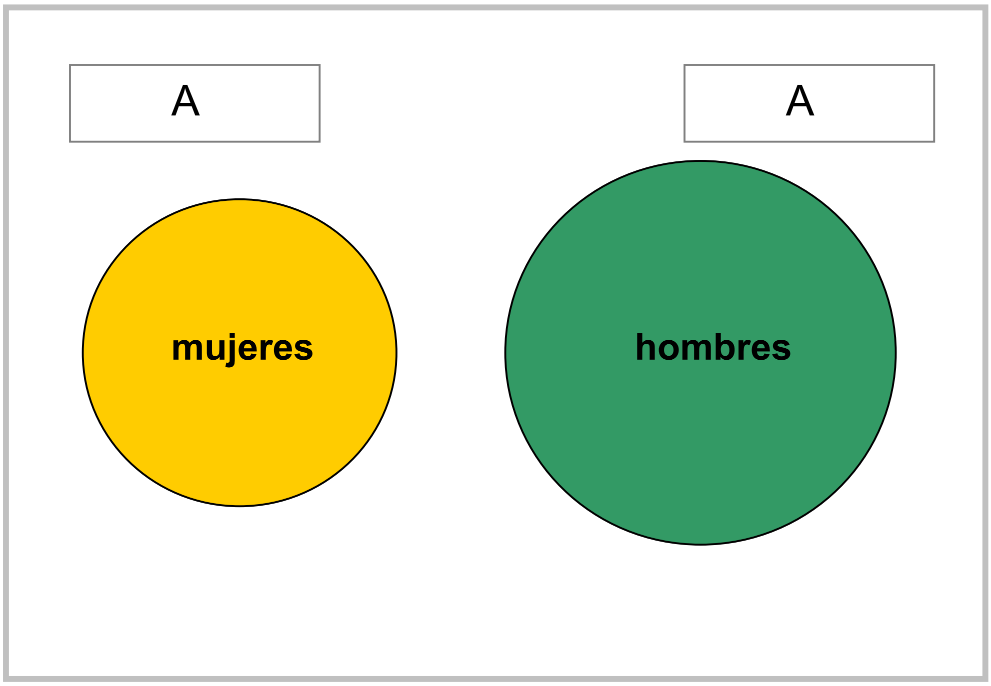
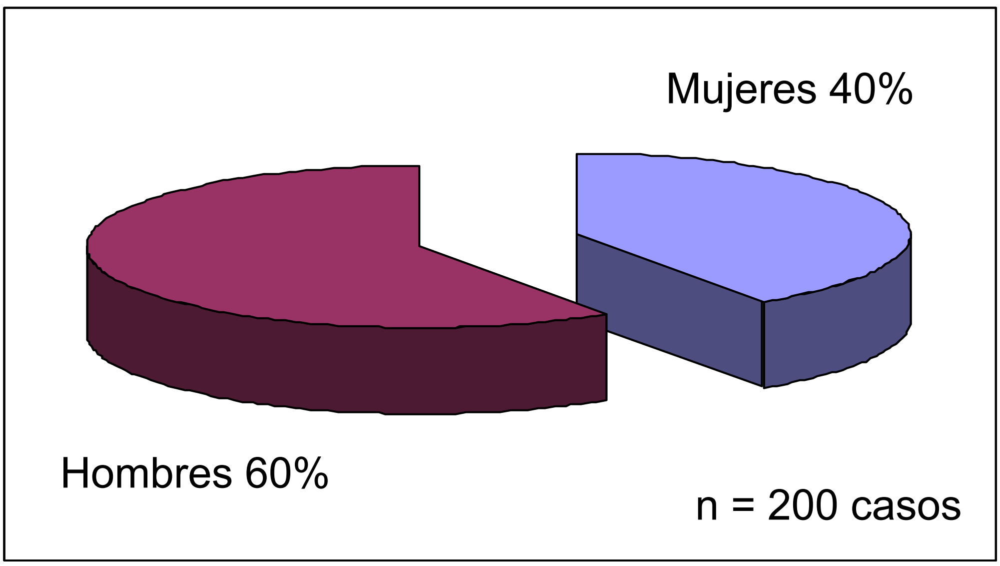


Diagrama Circular



El Diagrama Circular es el resumen gráfico más común en el caso de variables nominales. Existen distintas variantes de esta representación gráfica pero todas tienen las siguientes características:

1. Las frecuencias se representan a través de áreas.

2. Las figuras que corresponden a cada categoría se disponen de manera que se eviten interpretaciones equívocas respecto a un posible orden de las categorías.

Datos Cualitativos Ordinales

Tampoco son intrínsecamente numéricos; sin embargo, los valores con que se codifican tienen, además del papel de etiquetas, una asignación que se corresponde con el grado de intensidad con que presentan el atributo.

Al igual que los ordinales, los datos de este tipo típicamente toman sus distintos valores de una colección reducida de posibilidades.

Considere el ejemplo de la variable Escolaridad (X_7) del banco de datos 1. Ahí, se tienen 200 datos con valores 0, 1, 2 ó 3.

0 = bachillerato,
1 = licenciatura sin título;
2 = licenciatura con título y
3 = postgrado.

Si se cuenta con observaciones intercambiables, la información que, de nuevo, es relevante es la que se refiere a:

1. Los valores distintos que se presentan en el banco de datos.

2. La frecuencia con que cada uno de esos valores se presenta en el banco de datos.

Así, el resumen eficiente para este tipo de variables es, de nuevo, la Tabla de Frecuencias.

X	F	fr	%fr
0	34	0.170	17.0
1	117	0.585	58.5
2	47	0.235	23.5
3	2	0.010	01.0
Suma	**200**	**1**	**100 %**

A diferencia de lo que ocurre con los datos nominales, en este caso el orden en que se presentan los distintos valores de X en la tabla sí es informativo y no arbitrario.

Como consecuencia, las frecuencias se pueden acumular en un orden que tiene una interpretación útil en el contexto del estudio.

X	f	fr	%fr	F	Fr	%Fr
0	34	0.170	17.0	34	0.170	17.0
1	117	0.585	58.5	151	0.755	75.5
2	47	0.235	23.5	198	0.990	99.0
3	2	0.010	01.0	200	1.000	100
Suma	**200**	**1**	**100 %**			

La frecuencia acumulada F se calcula, para el valor X_i, como

$$F_i = \sum_{j=1}^{i} f_j$$

y se interpreta como el número de casos con un valor menor o igual que X_i .

Las frecuencias acumuladas relativas se obtienen de las acumuladas simples dividiendo por el número total de casos (n) y se interpretan como la proporción de casos menores o iguales que el correspondiente valor.

Por su parte, las frecuencias acumuladas porcentuales resultan de la expresión de las relativas en escala porcentual.

Por lo que se refiere a otros resúmenes numéricos más breves, que no tienen la propiedad de suficiencia, se puede reportar la **Moda** que en este caso es única y resulta el valor 1 con un 58.5% de los casos. Aquí, como coincidencia, sí ocurre que la mayoría de los casos corresponden con personas que tienen licenciatura sin título, pero no es esta una situación general en todo banco de datos.

Cuando se tienen datos ordinales, y precisamente debido al orden implícito en los valores, existe otro resumen que se puede utilizar. Se trata de los **Cuantiles**. Se dice que el valor $X_{(q)}$ es el cuantil de orden q de un conjunto de datos si satisface simultáneamente las siguientes dos propiedades:

1. Los datos, en un porcentaje de al menos $100 \times q$ %, son menores o iguales que $X_{(q)}$.

2. Los datos, en un porcentaje de al menos $100 \times (1-q)$ %, son mayores o iguales $X_{(q)}$.

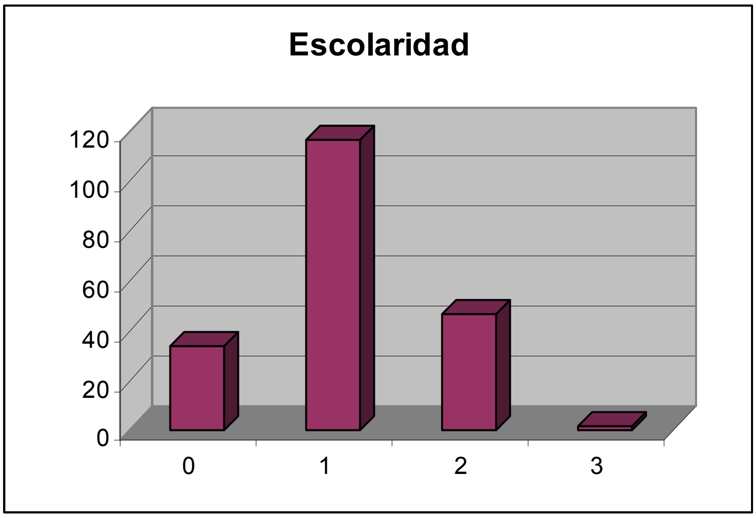
Un caso particularmente popular es el cuantil de orden 0.5 que también se conoce como la **Mediana**. Para los datos del ejemplo se tiene que la mediana coincide con el valor X = 1. Así, se puede decir que al menos el 50% de los individuos tienen estudios de cuando más licenciatura sin título y, simultáneamente, el 50% de los individuos tienen estudios de al menos licenciatura sin título.

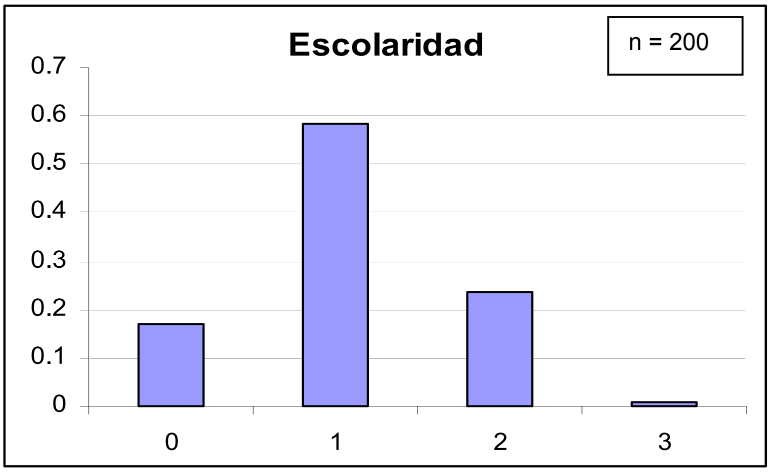
Otros cuantiles que de uso común son los cuartiles: $X_{(0.25)}$, $X_{(0.5)}$ y $X_{(0.75)}$ que, en términos generales, dividen el banco de datos en cuartas partes.

En el ejemplo tanto la mediana como el primer y el tercer cuartiles coinciden con el valor X = 1 como un reflejo de la concentración de casos en ese valor de X.

Finalmente, para datos ordinales suele reportarse, en ocasiones el valor de dos resúmenes más: el **Mínimo** $X_{[1]}$ y el **Máximo** $X_{[n]}$. Para los datos del ejemplo, $X_{[1]} = 0$ y $X_{[n]} = 3$.

Por lo que se refiere a los resúmenes gráficos, el orden en los valores sugiere la adopción de una representación que incorpore esta información en el desplegado. El gráfico más común en estos casos es el diagrama de barras.





Datos Cuantitativos Discretos

Son intrínsecamente numéricos; toman valores aislados y habitualmente se registran como resultado de un conteo. En virtud de su naturaleza numérica, es posible operar con ellos utilizando las herramientas aritméticas.

Considere el ejemplo de la variable Cursos (X_4) del banco de datos 1. Ahí, se tienen 200 datos con valores que van de 0 a 9 cursos de educación continua terminados.

Este tipo de variable, por su naturaleza numérica, tiene en particular todas las propiedades de los datos ordinales. Así, el resumen suficiente nuevamente es la tabla de frecuencias.

Tabla de Frecuencias de la variable Cursos

X	f	fr	%fr	F	Fr	%Fr
0	27	0.135	13.5	27	0.135	13.5
1	41	0.205	20.5	68	0.340	34.0
2	43	0.215	21.5	111	0.555	55.5
3	41	0.205	20.5	152	0.760	76.0
4	27	0.135	13.5	179	0.895	89.5
5	19	0.095	9.5	198	0.990	99.0
6	1	0.005	0.5	199	0.995	99.5
7	0	0.000	0.0	199	0.995	99.5
8	0	0.000	0.0	199	0.995	99.5
9	1	0.005	0.5	200	1.000	100
Suma	**200**	**1**	**100 %**			

En el caso de las variables cuantitativas, suele ocurrir que la colección de valores es más grande si se compara con lo que ocurre con las variables cualitativas. Este hecho puede restar efectividad a la tabla como resumen. En esas circunstancias es conveniente revisar los resúmenes numéricos complementarios.

Específicamente es conveniente considerar dos clases de resúmenes: las medidas de **localización** y las medidas de **variabilidad** o dispersión.

Las medidas de **localización** son resúmenes numéricos que tienen como propósito auxiliar en la respuesta la pregunta:

¿Dónde están los datos?

Las medidas de **dispersión** son resúmenes numéricos que tienen como propósito auxiliar en la respuesta la pregunta:

¿Qué tan homogéneos son los datos?

Una lista muy común, aún cuando no exhaustiva, de medidas de **localización** incluye los resúmenes que ya se han presentado:

1. La Moda.
2. La Mediana.
3. Los Cuantiles.
4. El Máximo.
5. El Mínimo.

La adición más importante a esta lista es la **Media**. Este resumen se define y calcula como el promedio aritmético de los datos en el banco:

$$\overline{X} = \tfrac{1}{n}\sum_{i=1}^{n} x_i$$

La **Media**. Indica un valor en torno al cual se encuentran las observaciones.

Para la variable cursos, los valores de estos resúmenes son los siguientes:

Moda = 2 (21.5% de casos)

Mediana = 2

Primer Cuartil = 1

Tercer Cuartil = 3

Mínimo = 0

Máximo = 9

Media = 2.34

Por su parte, las medidas de dispersión más simples son:

1. El Rango (**R**).
2. El Rango entre cuartiles (**REC**).

El rango se define como la diferencia entre el máximo y el mínimo.

$$\mathbf{R} = X_{[n]} - X_{[1]}$$

En el caso del ejemplo, **R** = 9.

A su vez, el rango entre cuartiles se define como la diferencia entre el tercer y el primer cuartiles.

$$\mathbf{REC} = X_{(0.75)} - X_{(0.25)}$$

En el caso del ejemplo, **REC** = 2.

Otras medidas de dispersión se calculan a partir de las desviaciones que se presentan en los datos, con respecto a un valor de referencia, usualmente una medida de localización. Tres ejemplos de medidas de este tipo son:

1. El Error Medio (**EM**).
2. La Varianza ($\mathbf{S}^2$).
3. La Desviación Estándar (**S**).

Si se utiliza como valor de referencia la media, el Error Medio se calcula como:

$$EM = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{X}\right|$$

En el caso del ejemplo, **EM** = 1.327

Si en cambio se utiliza la mediana o la moda (que coinciden) se tiene **EM** = 1.29.

En el caso de la Varianza (y de la Desviación Estándar) esta medida siempre se calcula respecto al Media. La Varianza se define como sigue:

$$\mathbf{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

Para el ejemplo se tiene que $\mathbf{S}^2 = 2.554$

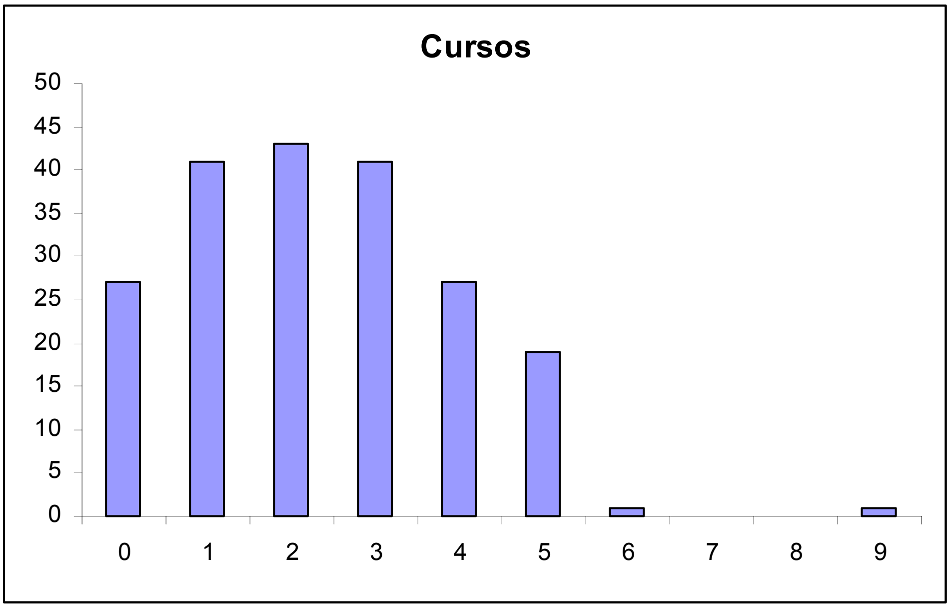
La Desviación Estándar se define como la raíz cuadrada de la varianza:

$$\mathbf{S} = \{ \mathbf{S}^2 \}^{1/2}$$

$$\mathbf{S} = \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^2 \right]^{1/2}$$

Para el ejemplo, se tiene que **S** = 1.598

Finalmente, en referente a los resúmenes gráficos, el gráfico más común en estos casos es, también, el diagrama de barras.



Datos Cuantitativos Continuos

Son intrínsecamente numéricos; potencialmente, al menos, podrían tomar cualquier valor en un intervalo predeterminado y habitualmente se registran como resultado de una medición.

Considere el ejemplo de la variable Edad (X_9) del banco de datos 1. Ahí, se tienen 200 datos con valores que van de 20 a 70 años cumplidos.

Como en los casos anteriores, el resumen eficiente es la tabla de frecuencias. Sin embargo, el volumen de posibles valores distintos constituye un problema en términos de efectividad del resumen.

Tabla de Frecuencias de la variable Edad

X	f	fr	%fr	F	Fr	%Fr
20	2	0.010	1.0	2	0.010	1.0
21	1	0.005	0.5	3	0.015	1.5
22	1	0.005	0.5	4	0.020	2.0
23	1	0.005	0.5	5	0.025	2.5
24	0	0.000	0.0	5	0.025	2.5
25	6	0.030	3.0	11	0.055	5.5
26	4	0.020	2.0	15	0.075	7.5
27	3	0.015	1.5	18	0.090	9.0
28	8	0.040	4.0	26	0.130	13.0
29	7	0.035	3.5	33	0.165	16.5
30	4	0.020	2.0	37	0.185	18.5
31	8	0.040	4.0	45	0.225	22.5
32	3	0.015	1.5	48	0.240	24.0
33	4	0.020	2.0	52	0.260	26.0
34	6	0.030	3.0	58	0.290	29.0
35	9	0.045	4.5	67	0.335	33.5
36	1	0.005	0.5	68	0.340	34.0
37	4	0.020	2.0	72	0.360	36.0
38	6	0.030	3.0	78	0.390	39.0
39	6	0.030	3.0	84	0.420	42.0
40	7	0.035	3.5	91	0.455	45.5
41	11	0.055	5.5	102	0.510	51.0
42	5	0.025	2.5	107	0.535	53.5
43	0	0.000	0.0	107	0.535	53.5
44	9	0.045	4.5	116	0.580	58.0
45	5	0.025	2.5	121	0.605	60.5

Tabla de Frecuencias de la variable Edad
(continuación)

X	f	fr	%fr	F	Fr	%Fr
46	6	0.030	3.0	127	0.635	63.5
47	5	0.025	2.5	132	0.660	66.0
48	2	0.010	1.0	134	0.670	67.0
49	5	0.025	2.5	139	0.695	69.5
50	3	0.015	1.5	142	0.710	71.0
51	3	0.015	1.5	145	0.725	72.5
52	4	0.020	2.0	149	0.745	74.5
53	3	0.015	1.5	152	0.760	76.0
54	3	0.015	1.5	155	0.775	77.5
55	3	0.015	1.5	158	0.790	79.0
56	2	0.010	1.0	160	0.800	80.0
57	5	0.025	2.5	165	0.825	82.5
58	4	0.020	2.0	169	0.845	84.5
59	0	0.000	0.0	169	0.845	84.5
60	1	0.005	0.5	170	0.850	85.0
61	3	0.015	1.5	173	0.865	86.5
62	4	0.020	2.0	177	0.885	88.5
63	7	0.035	3.5	184	0.920	92.0
64	1	0.005	0.5	185	0.925	92.5
65	4	0.020	2.0	189	0.945	94.5
66	3	0.015	1.5	192	0.960	96.0
67	2	0.010	1.0	194	0.970	97.0
68	3	0.015	1.5	197	0.985	98.5
69	2	0.010	1.0	199	0.995	99.5
70	1	0.005	0.5	200	1.000	100.0
Suma	**200**	**1**	**100 %**			

Aun si se suprimiesen las columnas que no constituyen el resumen básico, la tabla es lo suficientemente grande como para transmitir con efectividad las características de los datos en el banco.

Una solución, a costa de la suficiencia, consiste en reportar algunos de los resúmenes numéricos parciales (de localización y dispersión).

Para la variable Edad, los valores de las medidas de localización son los siguientes:

Moda = 41 años (5.5% de casos)

Mediana = 41 años

Primer Cuartil = 33 años

Tercer Cuartil = 53 años

Mínimo = 20 años

Máximo = 70 años

Media = 43.34

En lo que toca a las medidas de dispersión, para la variable Edad, los valores son los siguientes:

Rango = 50 años

Rango entre cuartiles = 20 años

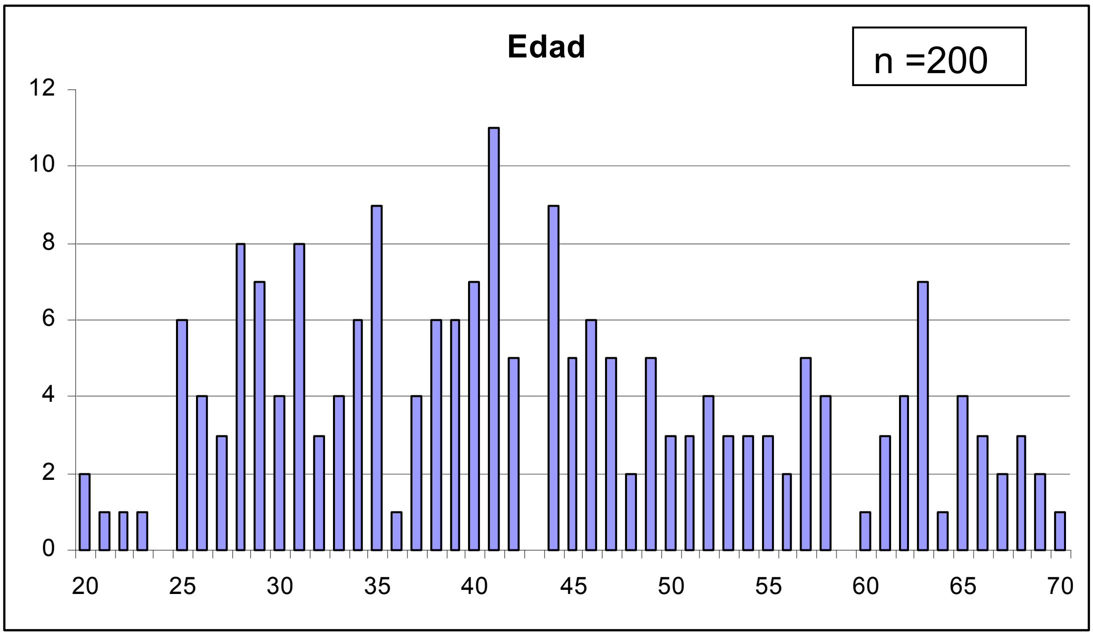
Error Medio = 10.68 años

Varianza = 162.75 $años^2$

Desviación Estándar = 12.76 años

Comente sobre la descripción, que sobre los datos de esta variable, permiten estos resúmenes.

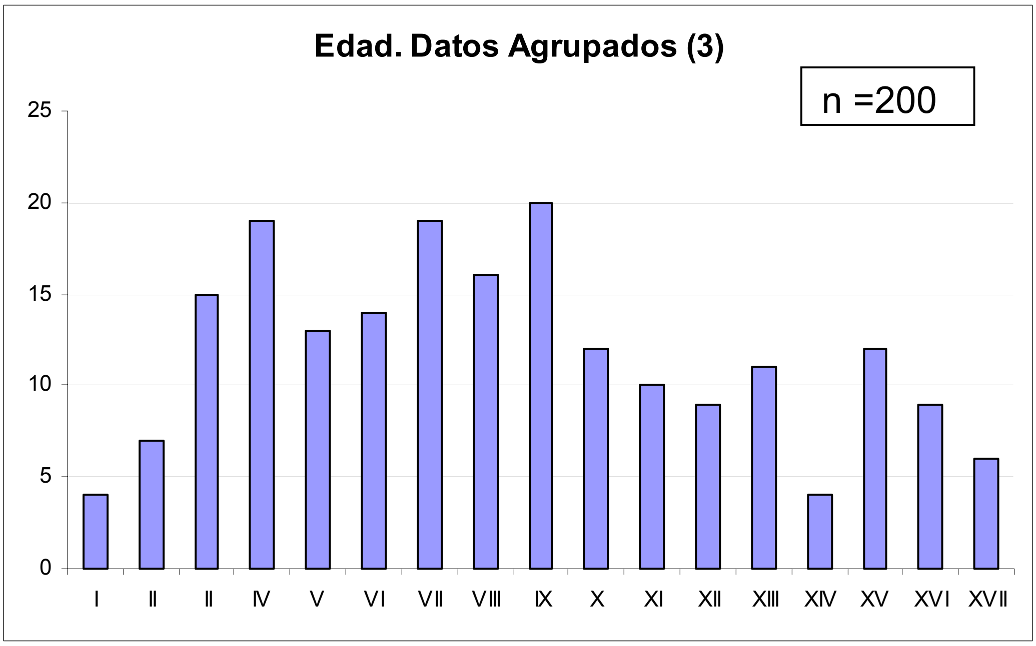
La descripción gráfica habitual en estos casos, es también el diagrama de barras.



Otro tratamiento que con frecuencia se aplica a los datos continuos, para obtener resúmenes más concisos, es el de agrupar los datos.

Tabla de Frecuencias de la variable Edad con intervalos de 3 años.

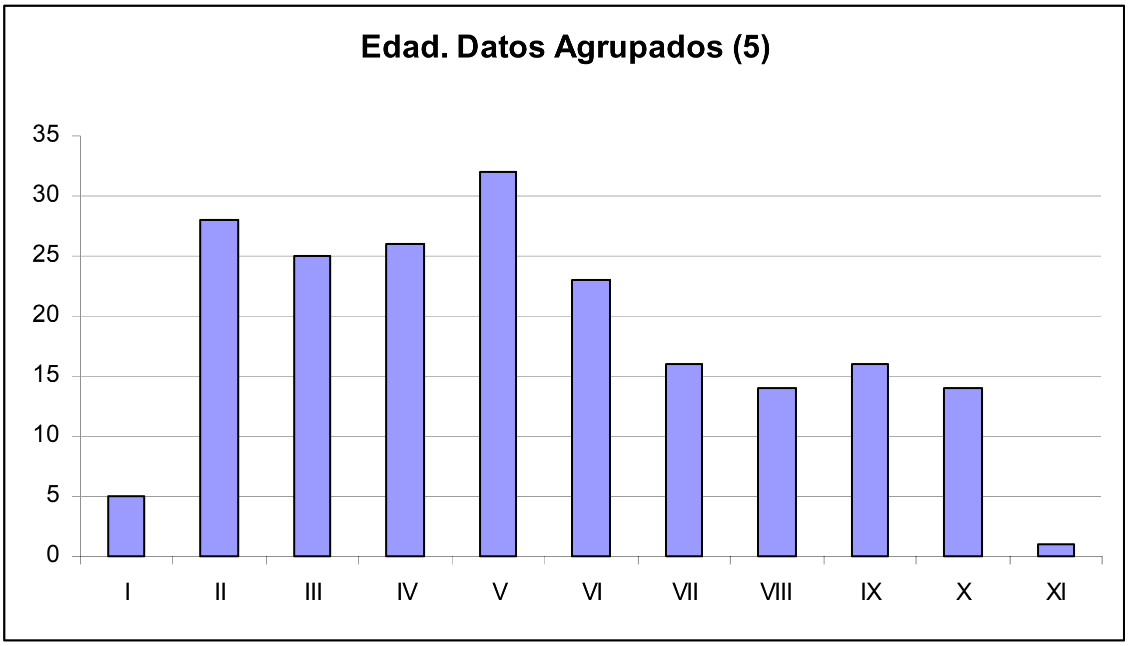
X	f	fr	%fr	F	Fr	%Fr
20-22	4	0.020	2.0	4	0.020	2.0
23-25	7	0.035	3.5	11	0.055	5.5
26-28	15	0.075	7.5	26	0.130	13.0
29-31	19	0.095	9.5	45	0.225	22.5
32-34	13	0.065	6.5	58	0.290	29.0
35-37	14	0.070	7.0	72	0.360	36.0
38-40	19	0.095	9.5	91	0.455	45.5
41-43	16	0.080	8.0	107	0.535	53.5
44-46	20	0.100	10.0	127	0.635	63.5
47-49	12	0.060	6.0	139	0.695	69.5
50-52	10	0.050	5.0	149	0.745	74.5
53-55	9	0.045	4.5	158	0.790	79.0
56-58	11	0.055	5.5	169	0.845	84.5
59-61	4	0.020	2.0	173	0.865	86.5
62-64	12	0.060	6.0	185	0.925	92.5
65-67	9	0.045	4.5	194	0.970	97.0
68-70	6	0.030	3.0	200	1.000	100.0
Suma	**200**	**1**	**100 %**			



Comente las características más evidentes en esta gráfica.

Tabla de Frecuencias de la variable Edad con intervalos de 5 años.

X	f	fr	%fr	F	Fr	%Fr
20-24	5	0.025	2.5	5	0.025	2.5
25-29	28	0.140	14.0	33	0.165	16.5
30-34	25	0.125	12.5	58	0.290	29.0
35-39	26	0.130	13.0	84	0.420	42.0
40-44	32	0.160	16.0	116	0.580	58.0
45-49	23	0.115	11.5	139	0.695	69.5
50-54	16	0.080	8.0	155	0.775	77.5
55-59	14	0.070	7.0	169	0.845	84.5
60-64	16	0.080	8.0	185	0.925	92.5
65-69	14	0.070	7.0	199	0.995	99.5
70-74	1	0.005	0.5	200	1.000	100.0
Suma	**200**	**1**	**100 %**			



Comente las características más evidentes en esta gráfica.

En relación con los resúmenes numéricos parciales, estos no tienen porque modificarse después de agrupar los datos. Por otra parte, si en el banco los datos aparecen ya agrupados, hay dos posibilidades:

1. Tratar los datos como ordinales.
2. Elegir un valor en cada intervalo y suponer que todos los datos ahí coinciden con ese representante.

Otra representación gráfica de uso frecuente en el caso de variables cuantitativas es el Diagrama de Caja o Box-Plot.

Este gráfico resume información tanto de localización como de dispersión y utiliza, para ello los datos del mínimo, los tres cuartiles y el máximo.

Diagrama de Caja para la variable Edad

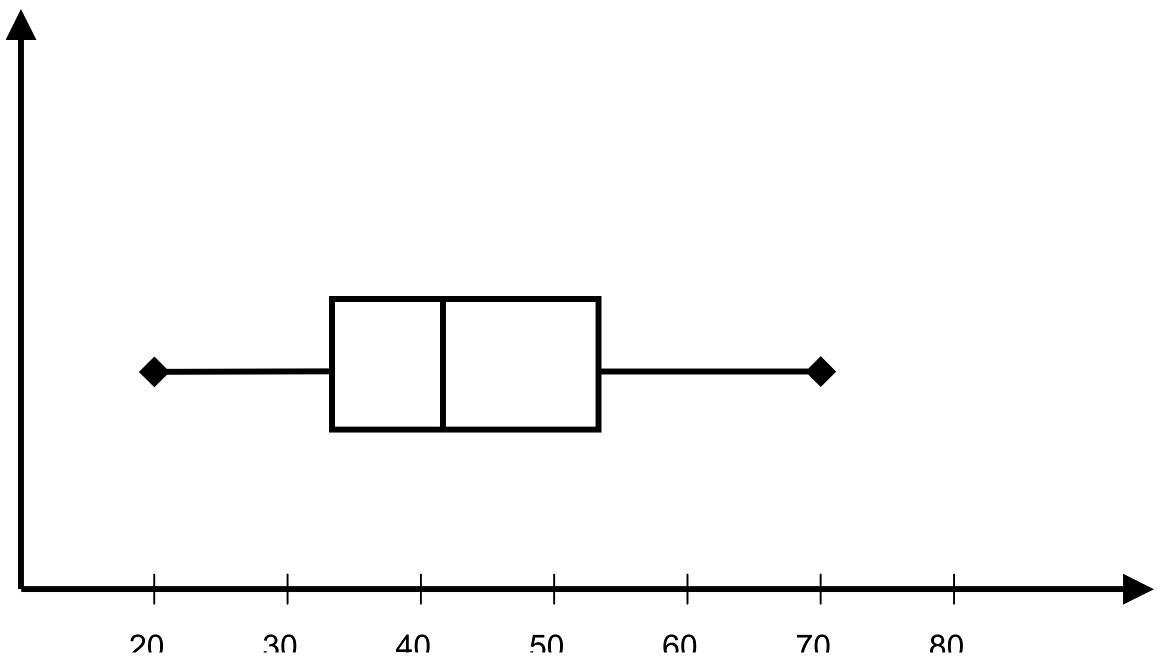
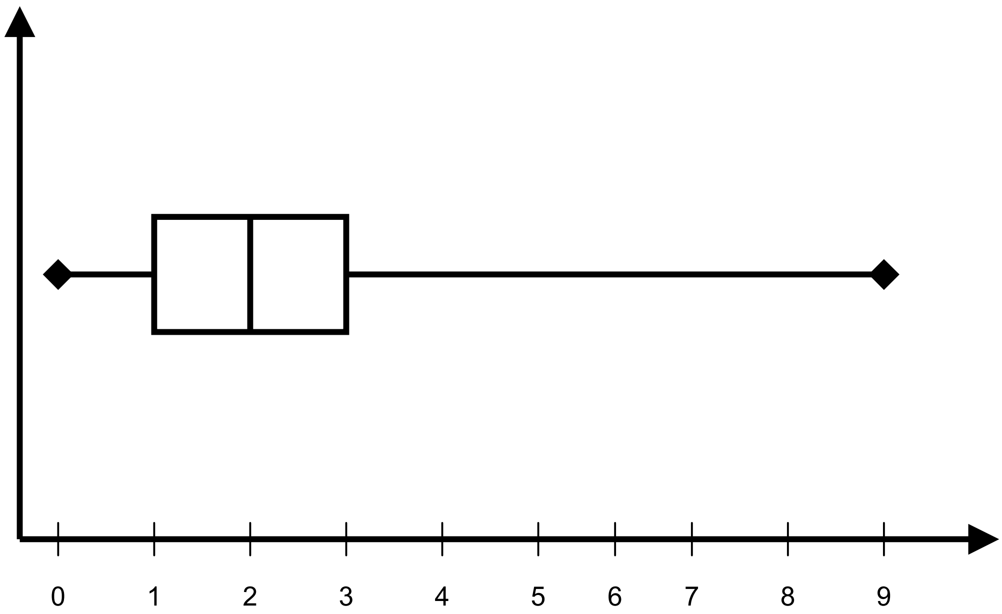


Diagrama de Caja para la variable Cursos



Comente sobre las conclusiones que, en general, se pueden extraer de un Diagrama de Caja.

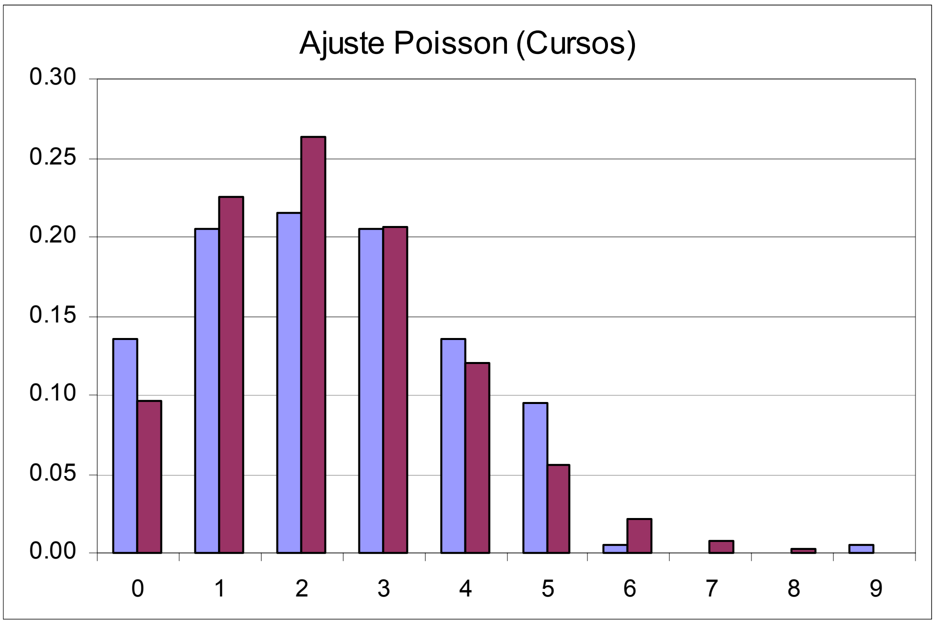
Otra manera de intentar una descripción concisa de la información en el banco es la adopción de modelos.

Un modelo es una representación aproximada de la realidad y pueden utilizarse modelos para representar la información contenida, por ejemplo, en una tabla de frecuencias.

La idea de un modelo, en este contexto, se refiere a una fórmula o ecuación que sea capaz de reproducir las frecuencias observadas en la tabla.

Los modelos que se emplean para este fin, son modelos de probabilidad. Existen distintas clases y gran cantidad de modelos de este tipo. Algunos ejemplos: Poisson, Weibull, Geométrico, Exponencial, Binomial, Normal entre otros.

Como ilustración considere la aproximación que produce un modelo Poisson a los datos de la variable Cursos.



Las barras que corresponden al modelo Poisson se calculan de acuerdo a la siguiente regla:

$$f_e = n \times P(X = x),$$

con f_e la frecuencia esperada, bajo el modelo, y

$$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

para $x = 0, 1, 2, \ldots$

En particular, las barras para este ejemplo fueron calculadas con un valor $\lambda = 2.34$.

X	fr	Poisson
0	0.1350	0.0963
1	0.2050	0.2254
2	0.2150	0.2637
3	0.2050	0.2057
4	0.1350	0.1203
5	0.0950	0.0563
6	0.0050	0.0220
7	0.0000	0.0073
8	0.0000	0.0021
9	0.0050	0.0006
10		0.0001
11		0.0000
12		0.0000
13		0.0000
14		0.0000
15		0.0000

Otra posibilidad es el modelo Binomial Negativo, en cuyo caso se tiene:

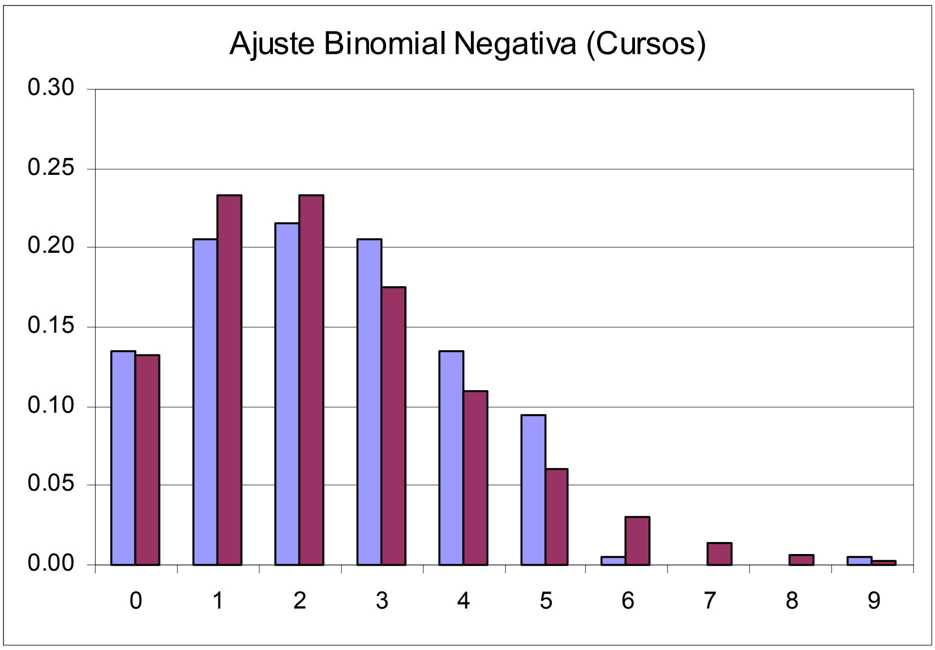
$$f_e = n \times P(X = x),$$

donde

$$P(X = x) = \binom{r + x - 1}{x} p^r (1 - p)^x$$

para $x = 0, 1, 2, \ldots$.

Las barras en este ejemplo fueron calculadas con valores $r = 7$ y $p = 0.75$.

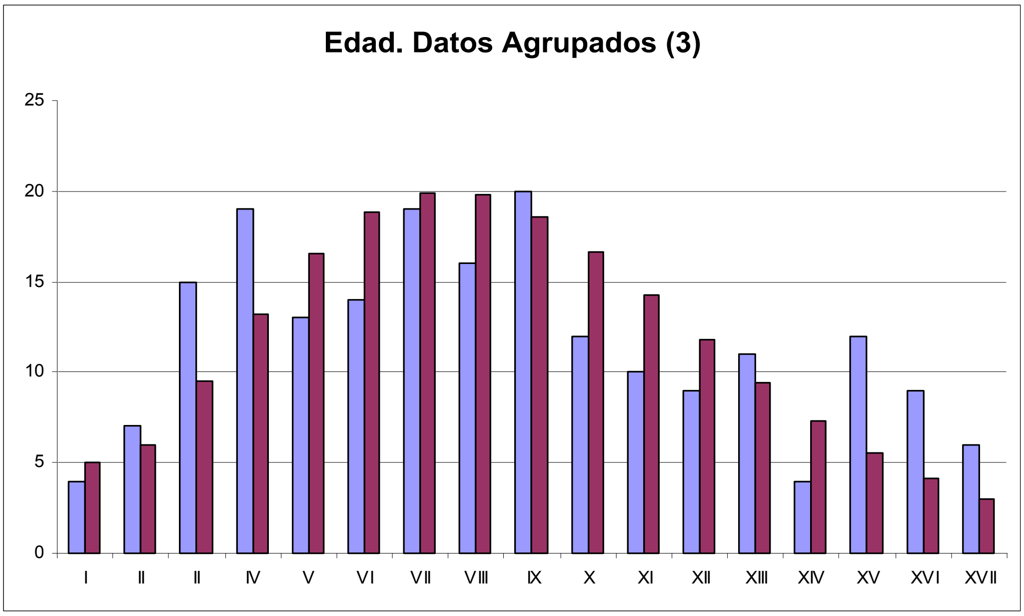


X	fr	Bin. Neg.
0	0.1350	0.1328
1	0.2050	0.2329
2	0.2150	0.2334
3	0.2050	0.1754
4	0.1350	0.1099
5	0.0950	0.0606
6	0.0050	0.0303
7	0.0000	0.0141
8	0.0000	0.0062
9	0.0050	0.0026
10		0.0010
11		0.0004
12		0.0002
13		0.0001
14		0.0000
15		0.0000

Un procedimiento similar se puede seguir para el caso de la variable Edad. La ilustración exhibe la aproximación que se obtiene si las frecuencias relativas se aproximan con un modelo Gama:

$$P(X = x) = \int_{x}^{x+1} \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t)dt$$

para x =1, 2, 3,…. En el ejemplo se utilizaron los valores α = 11.54 y β = 0.266.



Una de las características más atractivas del uso de modelos es que, una vez que se elige la clase modelos, todo el resumen se reduce a especificar los valores de los parámetros del modelo.

Análisis Exploratorio Comparativo

El AEC se refiere al empleo de las herramientas del Análisis Exploratorio para comparar la misma variable en dos o más bancos de datos.

Como ya se ha indicado, la descripción del comportamiento de una variable en un banco de datos se lleva a cabo a través de los resúmenes, numéricos o gráficos, apropiados.

Por otra parte, el tipo de resúmenes que son indicados para describir un conjunto de datos depende, como ya se discutió, del tipo de variable de que se trate.

En esos términos, el procedimiento para comparar los datos de una misma variable, provenientes de dos o más bancos resulta inmediato.

Los datos son del mismo tipo (es la misma variable) y por tanto la comparación se puede realizar a través de los resúmenes, numéricos o gráficos, que correspondan.

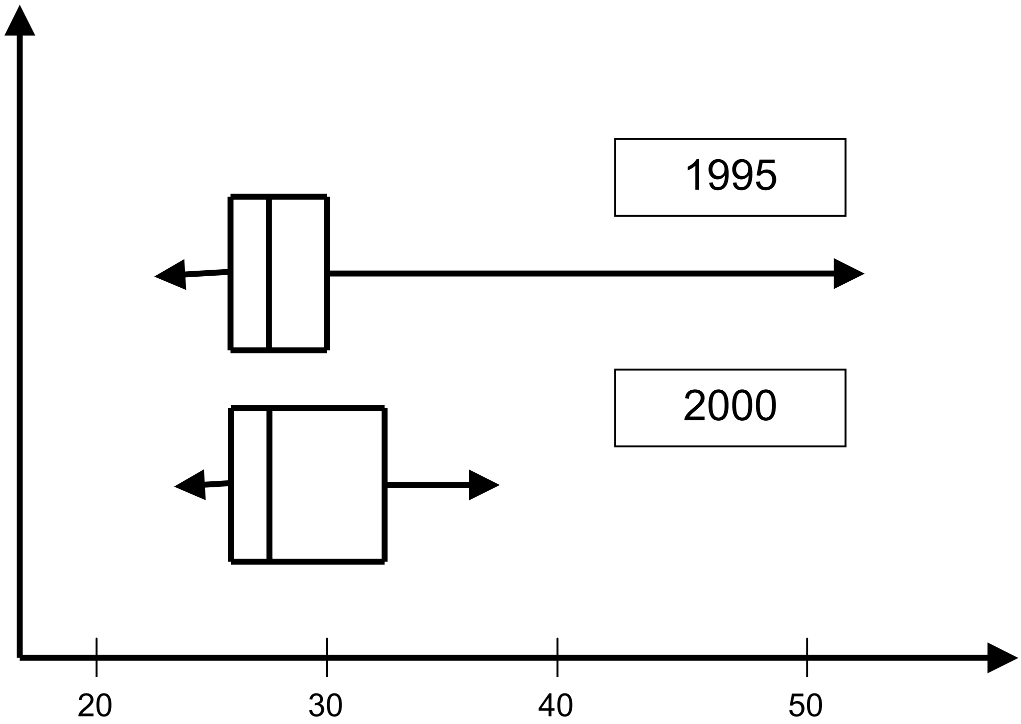
Como ejemplo considere los datos de la variable Edad del banco de datos del grupo y la información equivalente de un grupo que curso esta asignatura hace unos trimestres.

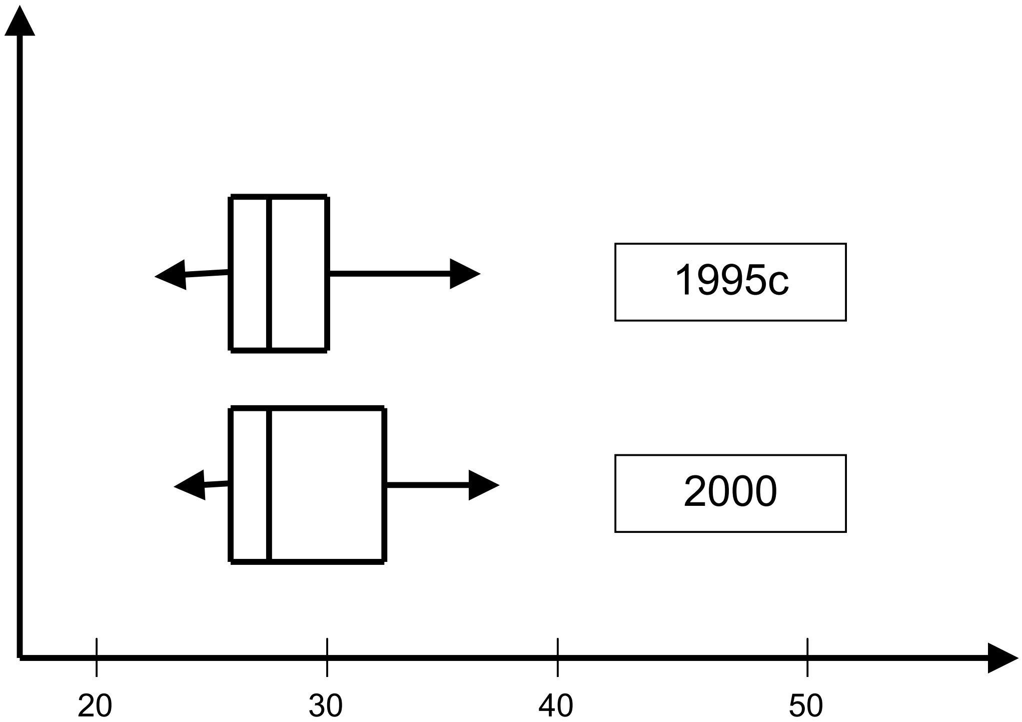
	2000		1995	
X	f	%fr	f	%fr
23	0	0.00	1	3.6
24	2	10.53	0	0.0
25	1	5.26	2	7.1
26	2	10.53	5	17.9
27	4	21.05	5	17.9
28	2	10.53	2	7.1
29	0	0.00	5	17.9
30	2	10.53	1	3.6
31	0	0.00	2	7.1
32	2	10.53	2	7.1
33	2	10.53	1	3.6
34	0	0.00	0	0.0
35	1	5.26	0	0.0
36	0	0.00	1	3.6
37	1	5.26	0	0.0
52	0	0.00	1	3.6
Suma	**19**	**100**	**28**	**100**

Resúmenes numéricos parciales:

	2000	**1995**
Moda	27 (21%)	26, 27 y 29 (18%)
Mediana	28	28
1er. Cuartil	26	26
3er. Cuartil	32	30
Mínimo	24	23
Máximo	37	52

Análisis Comparativo.
Diagramas de Caja 1995 y 2000





Existe una regla empírica que se utiliza para aislar posibles observaciones atípicas en los diagramas de caja. La idea básica es considerar que las observaciones más extremas son atípicas si la mitad 'exterior' de los datos ocupa un intervalo que equivale a 3 o más veces el rango entre cuartiles.

En una distribución simétrica la mitad exterior se dividiría por partes iguales a cada lado del primer y tercer cuartiles en un intervalo de longitud 1.5 veces el rango entre cuartiles respectivamente.

Como referencia, en el caso de un modelo Normal se tiene que:

1. El rango entre cuartiles es igual a 1.350 veces la desviación estándar,

2. La distancia entre el cuantil de orden 0.99 y el tercer cuartil es de 1.675 veces la desviación estándar,

3. Por tanto, la proporción entre estas dos diferencias es de 1.2407.

El modelo Normal es muy común; se utiliza para describir datos que provienen de variable continua con distribución que tiene una solo moda y es simétrica. La probabilidad que este modelo asigna al intervalo (a, b] está definida como:

$$\mathrm{P}(\mathrm{a} < \mathrm{X} \leq \mathrm{b}) = \int_{\mathrm{a}}^{\mathrm{b}} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{1}{2\sigma^2}(t-\mu)^2\}\, dt$$

para cualesquiera valores a < b.

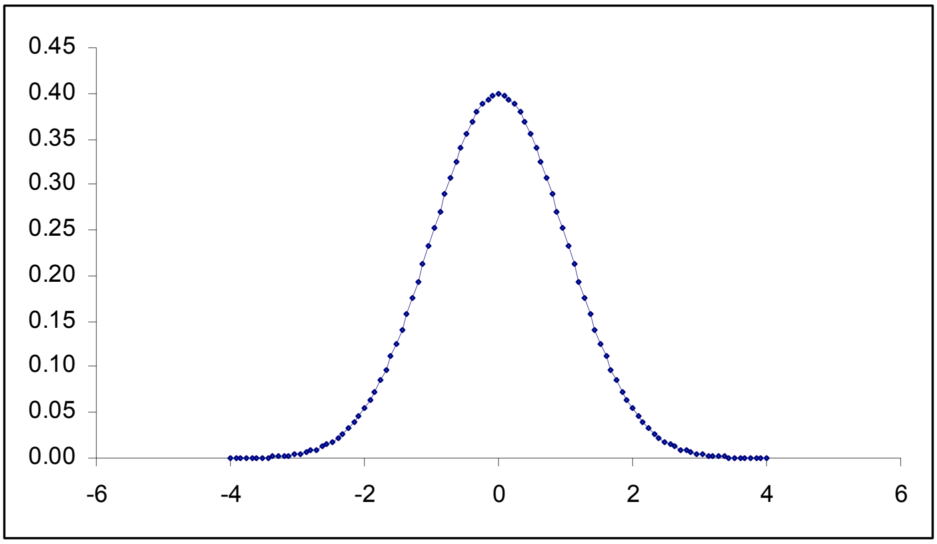
La función

$$\mathrm{f(x)} = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

está definida para toda -∞ < x < ∞, y se conoce como el nombre de función de densidad Normal con parámetros μ y σ^2.

Los parámetros μ y σ^2 se identifican con la media y la varianza respectivamente.

Densidad Normal con parámetros μ = 0, σ^2 = 1.
(Normal Estándar)



La probabilidad del evento $a < X \leq b$ se calcula como el área bajo la curva comprendida entre los límites a y b.

Si X tiene una distribución Normal con parámetros μ y σ^2, entonces $Z = (X - \mu) / \sigma$ sigue una distribución Normal estándar. Así, los cálculos de probabilidad de X se pueden efectuar a partir de Z.

De esta manera se tiene que

$$
\begin{aligned}
P[X \leq c] &= P[(X - \mu) \leq (c - \mu)] \\
&= P[(X - \mu) / \sigma \leq (c - \mu) / \sigma] \\
&= P[Z \leq (c - \mu) / \sigma]
\end{aligned}
$$

para cualquier valor de c.

Esta propiedad implica, en particular, que los cuantiles $X_{(q)}$ y $Z_{(q)}$ siguen la relación

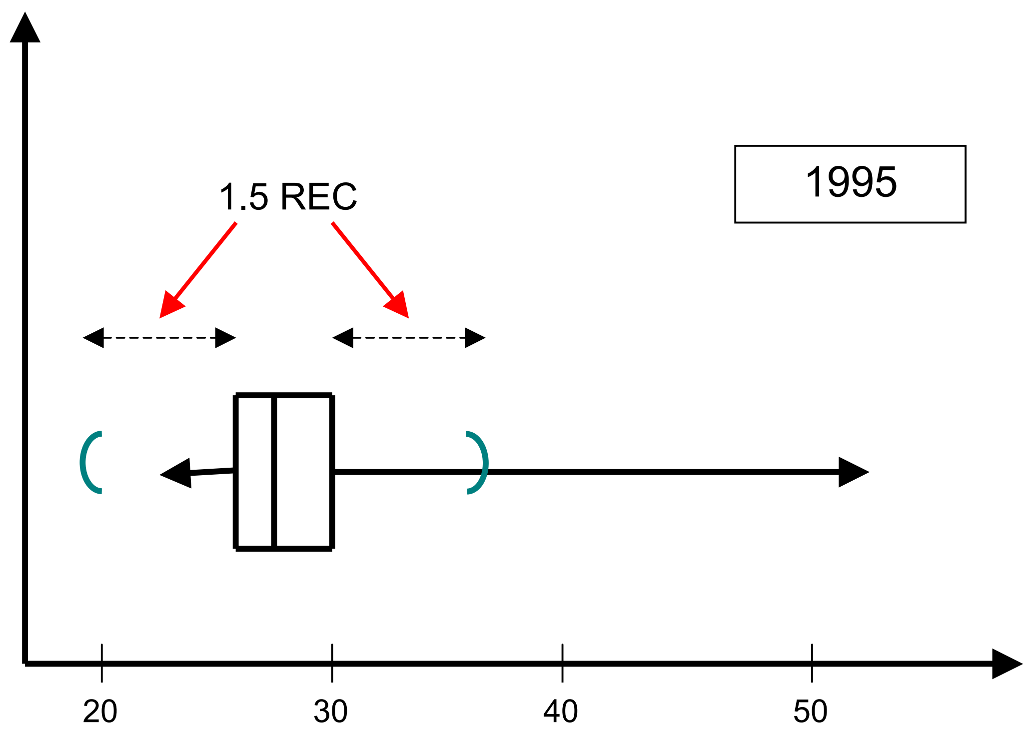
$$X_{(q)} = \sigma Z_{(q)} + \mu$$

Para cualquier $0 \leq q \leq 1$.

Haciendo uso de este resultado es muy fácil comprobar la relación entre el REC y la distancia entre el tercer cuartil y el cuantil de orden 0.99 en cualquier distribución Normal.

Volviendo al ejemplo con los datos de las edades, con el banco de 1995 ocurre que:

1. El rango entre cuartiles es 30 – 26 = 4,
2. La distancia entre el máximo y el tercer cuartil es 52 – 30 = 22,
3. La proporción entre estas dos diferencias es de 22 / 4 = 5.5.
4. Existen elementos, de acuerdo a la regla, para pensar que el máximo es atípico.



Diga usted, si se elimina de este conjunto de datos el máximo, las observaciones restantes, ¿ya no contienen valores atípicas?

El análisis comparativo, cuando se trata de una variable de algún otro tipo se lleva a cabo de forma similar, recurriendo a los resúmenes que correspondan, de acuerdo a la naturaleza de la variable y a los objetivos específicos del estudio (si existe alguno).

En un grupo de esta misma asignatura, hace algún tiempo, se aplicó un cuestionario similar y los datos que los 19 alumnos ofrecieron respecto al número de días de la semana en que leen algún periódico fueron los siguientes:

Caso	X	Caso	X
1	7	11	1
2	0	12	1
3	3	13	1
4	3	14	7
5	5	15	2
6	6	16	3
7	3	17	7
8	0	18	3
9	7	19	2
10	1		

Compare este banco de datos con el del grupo y comente los resultados.

Análisis Exploratorio de Asociación

El AEA se refiere al empleo de las herramientas del Análisis Exploratorio para establecer si dos o más variables están relacionadas entre sí.

Cuando se aborda el análisis conjunto de dos variables es necesario considerar el tipo de cada una de las dos. Para empezar, considere el caso de dos variables cualitativas.

Caso	Sexo	Facil.	Caso	Sexo	Facil.
1	1	1	11	1	1
2	1	0	12	1	1
3	1	2	13	0	0
4	1	0	14	1	1
5	1	1	15	1	1
6	1	1	16	1	1
7	1	1	17	0	2
8	1	1	18	1	1
9	0	2	19	1	0
10	1	1			

(Hombre = 1; Mujer = 0)
(Difícil = 0, Ni fácil ni difícil = 1; Difícil = 2)

Como siempre en el AE, lo primero es obtener un resumen eficiente. En este caso, de nuevo, se trata de la tabla de frecuencias. Con la variante de que, ahora, se trata de una tabla de frecuencias conjuntas.

Tabla de Frecuencias Conjuntas para las variables Sexo (Y) y Facilidad (X).

		Facilidad (X)			
		0	**1**	**2**	
Sexo (Y)	**1**	3	12	1	**16**
	0	1	0	2	**3**
		4	**12**	**3**	**19**

Tabla de Frecuencias Conjuntas Relativas Porcentuales para las variables Sexo y Facilidad.

		Facilidad (X)			
		0	**1**	**2**	
Sexo (Y)	**1**	15.79	63.16	5.26	**84.21**
	0	5.26	0.00	10.53	**15.79**
		21.05	**63.16**	**15.79**	**100%**

(n = 19)

Comente las principales conclusiones que se pueden obtener de esta tabla.

Tablas de Frecuencias Marginales Relativas Porcentuales para las variables Sexo y Facilidad.

		Facilidad (X)			
		0	1	2	
Sexo (Y)	1	15.79	63.16	5.26	**84.21**
	0	5.26	0.00	10.53	**15.79**
		21.05	**63.16**	**15.79**	**100%**

Comente las principales conclusiones que se pueden obtener de cada una de estas dos tablas.

Tablas de Frecuencias Condicionales Relativas Porcentuales para la variable Facilidad dada la variable Sexo.

		Facilidad (X)			
		0	1	2	
Sexo (Y)	1	18.75	75.00	6.25	**100%**
	0	33.33	0.00	66.67	**100%**

Compare contra la marginal correspondiente:

Facilidad (X)			
0	1	2	
21.05	**63.16**	**15.79**	**100%**

Comente las principales conclusiones que se pueden obtener de cada una de estas dos tablas.

Tablas de Frecuencias Condicionales Relativas Porcentuales para la variable Sexo dada la variable Facilidad.

		Facilidad (X)			
		0	**1**	**2**	
Sexo (Y)	**1**	75.00	100.00	33.33	
	0	25.00	0.00	66.67	
		100%	**100%**	**100%**	

Compare contra la marginal correspondiente:

Sexo (Y)	**1**	**84.21**
	0	**15.79**
		100%

Comente las principales conclusiones que se pueden obtener de cada una de estas dos tablas.

Estos cálculos pueden considerarse de frente a la producción de pronósticos.

¿Qué tabla utilizaría para pronosticar el grado de facilidad que le atribuiría a la Estadística una persona de esta población, formada por 19 casos?

¿Modificaría su pronóstico si antes de producirlo le fuese comunicado que la persona en cuestión es mujer? ¿Cómo lo modificaría? ¿Y si le dijesen que es hombre?

El caso es que si no se considera la información sobre la variable Sexo, la tabla de frecuencias marginal de la variable Facilidad sugiere que el pronóstico adecuado es

Facilidad = 1 (ni fácil ni difícil)

con una confiabilidad de 63.16%. En términos de probabilidad se tiene que

$$P(X = 1) = 0.6316.$$

Por otro lado, si la persona es mujer, ($Y = 0$) y el análisis incorpora esa restricción reduciendo, por tanto, la población bajo consideración, entonces de la tabla de frecuencias condicionales se tiene que el pronóstico apropiado es ahora

Facilidad = 2 (difícil)

con una confiabilidad de 66.67%. En términos de probabilidad se tiene

$$P(X = 2 \mid Y = 0) = 0.667$$

En el caso $Y = 1$ (hombre), de forma similar, se tiene que el pronóstico es ahora

Facilidad = 1 (ni fácil ni difícil)

con una confiabilidad de 75%. En términos de probabilidad se tiene

$$P(X = 1 \mid Y = 1) = 0.75$$

A partir de este ejemplo se puede introducir el siguiente concepto: dos variables en un banco de datos se dice que están relacionadas si la descripción de una de ellas se modifica cuando se conoce el valor de la otra.

En el lenguaje de los pronósticos, dos variables están relacionadas cuando el pronóstico sobre una de ellas se modifica cuando es revelado el valor de la otra.

Equivalentemente, dos variables en un banco se dice que son independientes si la descripción de una de ellas no se modifica por el conocimiento del valor de la otra.

También se puede decir que dos variables son independientes si el pronóstico sobre una de ellas no se modifica cuando es revelado el valor de la otra variable.

Para profundizar en la comprensión del concepto de independencia vale la pena considerar las siguientes definiciones. Sean

1. **n** el número total de casos en el banco,
2. **n(i, •)** el número de casos en la i-ésima categoría de la variable Y (i = 1, 2,…r),
3. **n(•, j)** el número de casos en la j-ésima categoría de la variable X (j = 1, 2,…c),
4. **n(i, j)** el número de casos en la celda (i, j) para i = 1, 2,…r y j = 1, 2,…c.

Como ilustración, considere los datos del ejemplo con las variables Sexo con r = 2 categorías y Facilidad con c = 3 categorías.

		Facilidad (X)			
		0	**1**	**2**	
Sexo (Y)	**0**	3	12	1	**16**
	1	1	0	2	**3**
		4	**12**	**3**	**19**

En este caso,

1. **n** el número total de casos en el banco es **19**.

2. **n(i, •)** el número de casos en la i-ésima categoría de la variable Y (i = 1, r).

   $$n(1, \bullet) = 16 \quad y \quad n(2, \bullet) = 3$$

3. **n(•, j)** el número de casos en la j-ésima categoría de la variable X (j = 1, 2, 3).

   $$n(\bullet, 1) = 4,\ n(\bullet, 2) = 12 \ \ y \ \ n(\bullet, 3) = 3,$$

4. **n(i, j)** el número de casos en la celda (i, j) para i = 1, 2,…r y j = 1, 2,…c.

   $n(1, 1) = 3,\ \ n(1, 2) = 12,\ \ n(1, 3) = 1,$
   $n(2, 1) = 1,\ \ n(2, 2) = 0,\ \ n(2, 3) = 2,$

Con esta notación, la frecuencia relativa conjunta de la celda (i, j) se calcula como

$$fr_{(i,j)} = n(i, j) / n .$$

Además, la frecuencia relativa condicional de la categoría j de la variable X (Facilidad) dado que la variable Y (Sexo) toma el valor de su categoría i está dada por

$$fr_{(j \mid i)} = n(i, j) / n(i, \bullet) .$$

Por otra parte, la frecuencia relativa marginal de la misma categoría j de la variable X se calcula como

$$fr_{(j)} = n(\bullet, j) / n .$$

En términos de estas frecuencias relativas, se dice que las variables X e Y son independientes si para toda celda ocurre que la frecuencia relativa condicional $fr_{(j \mid i)}$ coincide con la correspondiente frecuencia relativa marginal $fr_{(j)}$. Es decir, X e Y son independientes si

$$n(i, j) / n(i, \bullet) = n(\bullet, j) / n$$

sin importar el valor de i.

Es evidente que si se cumple la condición

$$n(i, j) / n(i, \bullet) = n(\bullet, j) / n$$

entonces se tiene que

$$n(i, j) / n = \{ n(\bullet, j) / n \} \times \{ n(i, \bullet) / n \}$$

o, en otros términos,

$$fr_{(i,j)} = fr_{(i)} \times fr_{(j)}$$

En otras palabras, dos variables se dice que son independientes si en cada celda se verifica que la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales respectivas.

En el ejemplo, evidentemente, las variables Sexo y Facilidad no son independientes.

En la práctica, es virtualmente imposible encontrar bancos de datos con variables independientes.

Sin embargo, en muchos casos la magnitud de la diferencia entre las tablas marginales y las tablas condicionales es suficientemente pequeña para que, para propósitos prácticos, las variables se consideren independientes.

En cualquier caso, existen algunos resúmenes numéricos diseñados con el objetivo específico de medir la asociación (falta de independencia) entre variables.

En el caso de las variables cualitativas, todas las medidas de asociación proceden de la siguiente manera:

1. Se evalúan las diferencias que existen entre las frecuencias relativas (marginales y condicionales, o conjuntas y el producto de las marginales) en cada una de las celdas.

2. Se calcula una medida resumen de esas diferencias.

3. Las variables se declaran independientes únicamente en el caso en que la medida resumen es suficientemente pequeña.

En el ejemplo, se puede comparar la tabla de frecuencias conjuntas relativas de los datos con la tabla de frecuencias relativas que se obtendría como resultado de multiplicar las frecuencias marginales correspondientes.

Tabla de Frecuencias Conjuntas Relativas Porcentuales para las variables Sexo y Facilidad.

Sexo (Y) \ Facilidad (X)	0	1	2	
1	15.79	63.16	5.26	**84.21**
0	5.26	0.00	10.53	**15.79**
	21.05	**63.16**	**15.79**	**100%**

Tabla de Frecuencias Relativas Porcentuales (producto de las marginales) para las variables Sexo y Facilidad.

Sexo (Y) \ Facilidad (X)	0	1	2	
1	17.73	53.19	13.30	**84.21**
0	3.32	9.97	2.49	**15.79**
	21.05	**63.16**	**15.79**	**100%**

Así, las diferencias relevantes son las siguientes:

$$15.79 - 17.73 = -1.94\ \%$$
$$63.16 - 53.19 = 9.97\ \%$$
$$05.26 - 13.30 = -8.04\ \%$$
$$05.26 - 03.32 = 1.94\ \%$$
$$00.00 - 09.97 = -9.97\ \%$$
$$10.53 - 02.49 = 8.04\ \%$$

que suman a cero puesto que las frecuencias en cada tabla suman 100%.

Algunas medidas refieren estos porcentajes al tamaño del banco original (n = 19)

$$-1.94\ \% \times 19 = -0.369$$
$$9.97\ \% \times 19 = 1.894$$
$$-8.04\ \% \times 19 = -1.528$$
$$1.94\ \% \times 19 = 0.369$$
$$-9.97\ \% \times 19 = -1.894$$
$$8.04\ \% \times 19 = 1.528$$

En esos términos, multiplicando por n, cada una de estas cantidades se puede interpretar como la diferencia entre el número de casos observados en la respectiva celda (o_{ij}) y el número de casos esperado (e_{ij}), bajo el supuesto de independencia.

Así, las diferentes medidas de independencia para estos casos, típicamente utilizan un resumen de las diferencias entre observados y esperados. Un par de ejemplos son los siguientes:

$$\Delta_1 = \sum (o_{ij} - e_{ij})^2 \,,$$

$$\Delta_2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \,.$$

En ambos casos, la suma se calcula sobre todas las celdas y se dice que no hay independencia si los valores de los resúmenes son grandes.

Como ilustración, calcule el valor de Δ_1 y Δ_2 para los datos del banco con 19 casos.

La asociación entre variables se puede analizar en pares de tres distintos tipos:

I. Cualitativa vs cualitativa,
II. Cualitativa vs cuantitativa y
III. Cuantitativa vs cuantitativa.

Observe que la asociación es una propiedad simétrica (X independiente de Y, es lo mismo que Y independiente de X).

El primer caso ya ha sido comentado; por su parte, el caso cualitativa vs cuantitativa se trata en una forma que semeja un Análisis Exploratorio de comparación.

La idea es que para cada una de las categorías de la variable cualitativa –que usualmente son pocas- se describe el comportamiento condicional de la variable cuantitativa, ya sea a través de la tabla de frecuencias respectiva o mediante cualquier otro tipo de resúmenes.

Como ejemplo, considere las variables Sexo y Peso para las 28 personas del grupo de Estadística y Pronósticos de 1995.

Base de Datos 1995

estatura	peso	edad	sexo
165	70	29	1
185	90	32	0
165	63	30	1
156	45	36	1
181	95	28	0
174	74	27	0
178	71	31	0
174	73	23	0
183	75	27	0
176	88	33	0
174	64	29	0
186	75	52	0
181	69	26	0
180	74	26	0
165	55	27	1
176	62	26	1
178	72	27	0
163	54	29	1
160	52	31	1
170	70	26	1
167	62	26	1
179	73	29	0
158	53	25	1
163	68	27	1
176	76	29	0
170	65	25	0
174	76	28	0
170	63	32	0

En este caso, por facilidad, es conveniente estudiar las (dos) distribuciones condicionales de la variable Peso que corresponden a los valores de la variable Sexo. Los datos específicos para el ejemplo son los siguientes:

Mujeres

sexo	peso
1	70
1	63
1	45
1	55
1	62
1	54
1	52
1	70
1	62
1	53
1	68
11	

Hombres

sexo	peso
0	90
0	95
0	74
0	71
0	73
0	75
0	88
0	64
0	75
0	69
0	74
0	72
0	73
0	76
0	65
0	76
0	63
17	

Las tablas de frecuencias relativas porcentuales condicionales para el Peso dado el Sexo son:

Nivel	Intervalo	Mujeres	Hombres
I	40-44	0.00	0.00
II	45-49	9.09	0.00
III	50-54	27.27	0.00
IV	55-59	9.09	0.00
V	60-64	27.27	11.76
VI	64-69	9.09	11.76
VII	70-74	18.18	35.29
VIII	75-79	0.00	23.53
IX	80-84	0.00	0.00
X	85-89	0.00	5.88
XI	90-94	0.00	5.88
XII	95-100	0.00	5.88
		100.00	**100.00**

El examen directo de las dos tablas sugiere una conclusión general inmediata. Se observa que los pesos de los hombres tienden a ser mayores que los pesos de las mujeres.

Esta apreciación introduce, por una parte, la noción de tendencia en el análisis de asociación. Esta idea se refiere a un comportamiento general que se observa en el banco de datos aun cuando no se aplique a pares individuales de casos.

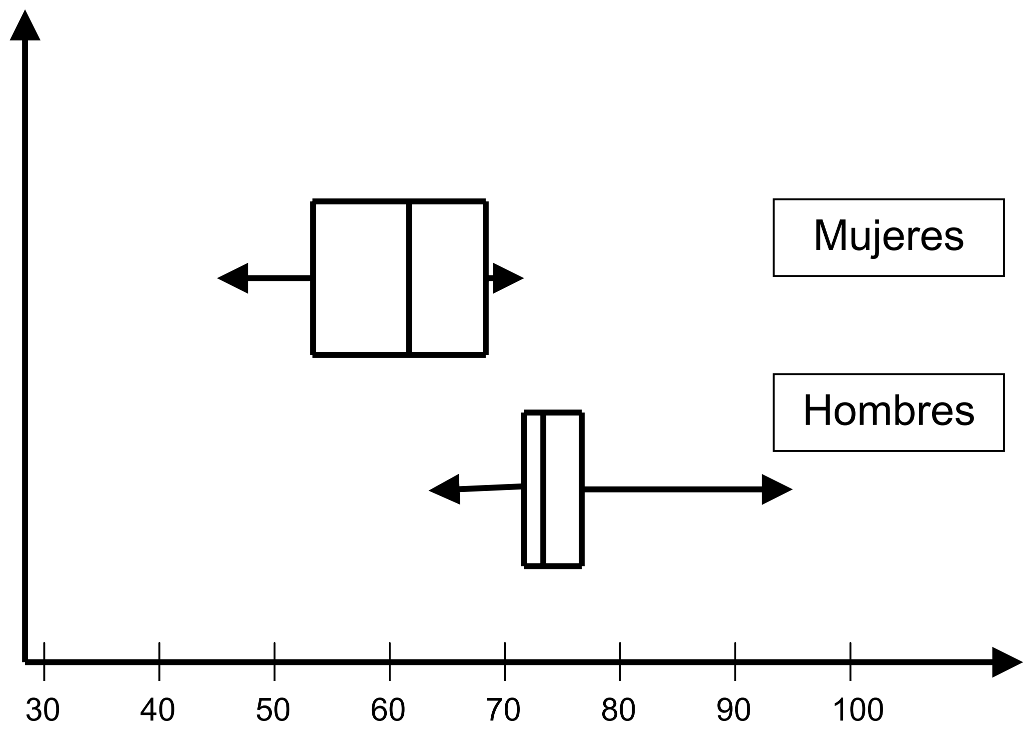
La primera impresión se puede precisar a través de los resúmenes numéricos parciales:

	Mujeres	**Hombres**
Moda	62 y 70 (18%)	73, 74, 75 y 76 (12%)
Mediana	62	74
1er. Cuartil	53	71
3er. Cuartil	68	76
Mínimo	45	63
Máximo	70	95

Como se puede observar, para todas las medidas de localización consideradas, las del estrato de las mujeres son menores que las correspondientes de los hombres. En lo que se refiere a dispersión, se tiene el siguiente panorama:

	Mujeres	**Hombres**
Rango	25	32
REC	15	5
Varianza	61.52	72.46
D. Estándar	7.84	8.51

En conjunto, esta información se puede desplegar convenientemente mediante un diagrama de caja.



Del diagrama no solamente se confirme que existe asociación entre las variables Peso y Sexo (no son independientes) sino que a demás se clarifica que la influencia no se limita a la localización sino también a la dispersión.

Así, se tiene que:

1. Los pesos de los hombres tienden a ser mayores que los de las mujeres.

2. En términos globales, los pesos de los hombres presentan una mayor dispersión que los de las mujeres.

3. En términos locales, la mitad central de los pesos de los hombres presenta menor dispersión que la mitad central de los pesos de la mujeres.

Como ejercicio complementario, determine la posible existencia de observaciones atípicas tanto en los pesos de mujeres como de hombres.

Por lo que toca al análisis de asociación con pares de variables del tercer tipo, es decir, cuantitativa vs cuantitativa, se presentan algunas peculiaridades que merecen atención.

En primer lugar, la comparación de las frecuencias conjuntas con el producto de las marginales (o de las marginales con las condicionales) implica el examen de las correspondientes tablas de frecuencias.

Ahora, como ya se ha discutido, en el caso de variables cuantitativas las correspondientes tablas de frecuencias pueden ser voluminosas y, por tanto, inconvenientes.

Considere, como ejemplo, el caso de las variables Peso y Estatura en el banco de datos de un grupo de personas.

Tabla de frecuencias conjuntas de las variables Peso y Estatura.

Estatura

Peso

	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	
150-154	1	0	0	0	0	0	0	0	0	0	0	0	**1**
155-159	0	0	0	0	0	0	0	0	0	0	0	0	**0**
160-164	0	0	0	0	0	0	0	0	0	0	0	0	**0**
165-169	0	1	1	3	0	2	0	0	0	0	0	0	**7**
170-174	0	0	0	0	1	3	2	0	0	0	0	0	**6**
175-179	0	0	0	2	0	0	0	0	0	0	0	0	**2**
180-184	0	0	0	0	0	0	0	0	0	0	0	1	**1**
185-189	0	0	0	0	0	0	0	0	1	0	0	0	**1**
190-194	0	0	0	0	0	0	0	0	0	0	0	0	**0**
195-199	0	0	0	0	0	0	0	0	0	0	1	0	**1**
	1	**1**	**1**	**5**	**1**	**5**	**2**	**0**	**1**	**0**	**1**	**1**	**19**

En esta tabla es evidente que las distribuciones marginales no son iguales entre sí y, por tanto, no son iguales a las correspondientes marginales.

Tablas de frecuencias condicionales relativas porcentuales para la variable Estatura dado el valor de la Peso.

	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104
150-154	100											
155-159												
160-164												
165-169		100	100	60		40						
170-174					100	60	100					
175-179				40								
180-184												100
185-189									100			
190-194												
195-199											100	
	100	**100**	**100**	**100**	**100**	**100**	**100**		**100**		**100**	**100**

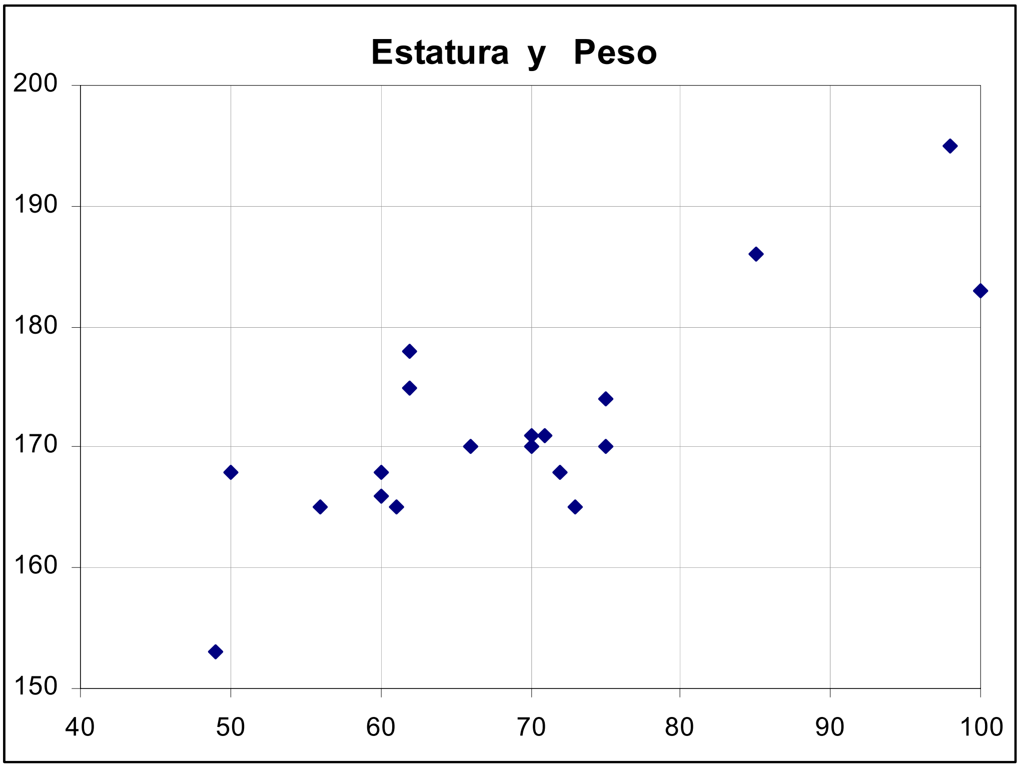
	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	
150-154	1												**1**
155-159													**0**
160-164													**0**
165-169		1	1	3		2							**7**
170-174					1	3	2						**6**
175-179				2									**2**
180-184												1	**1**
185-189									1				**1**
190-194													**0**
195-199											1		**1**
	1	**1**	**1**	**5**	**1**	**5**	**2**	**0**	**1**	**0**	**1**	**1**	**19**

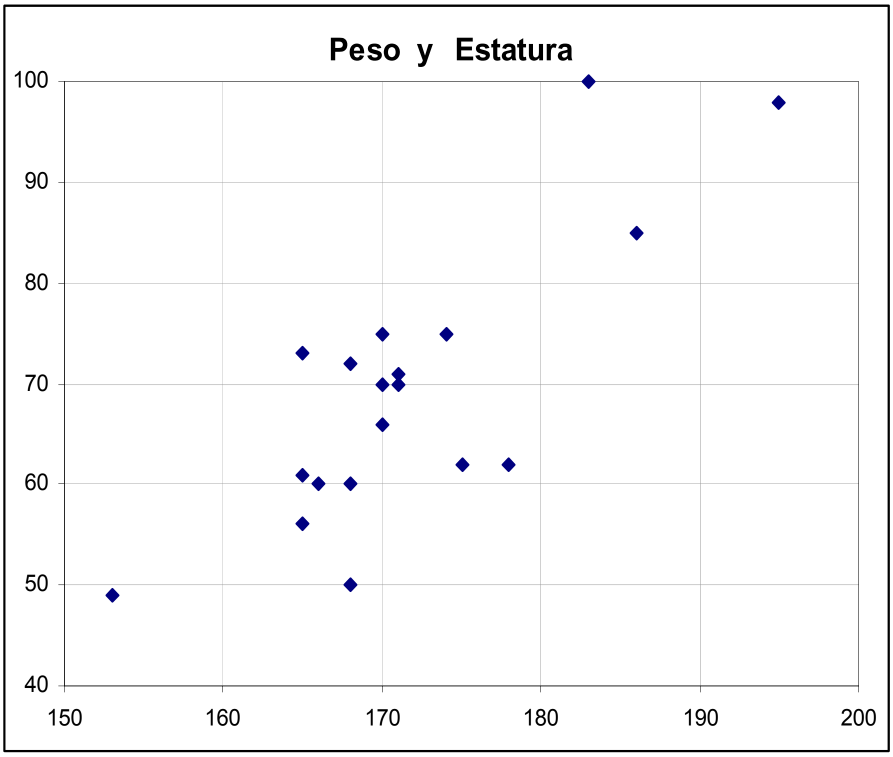
Si se aplica de nuevo con el fin de clarificar más la estrategia de agrupar, se obtienen tablas más reducidas.

	40-49	50-59	60-69	70-79	80-89	90-99	100-109	
150-159	1							**1**
160-169		2	3	2				**7**
170-179			3	5				**8**
180-189					1		1	**2**
190-199						1		**1**
	1	**2**	**6**	**7**	**1**	**1**	**1**	**19**

	40-59	60-79	89-99	100-119	
150-169	3	5			**8**
170-189		8	1	1	**10**
190-209			1		**1**
	3	**13**	**2**	**1**	**19**

Estas tablas sugieren, sin embargo, una estrategia en la dirección opuesta que conduce a un resumen que puede considerarse híbrido entre los gráficos y numéricos. Se trata del Diagrama de Dispersión.



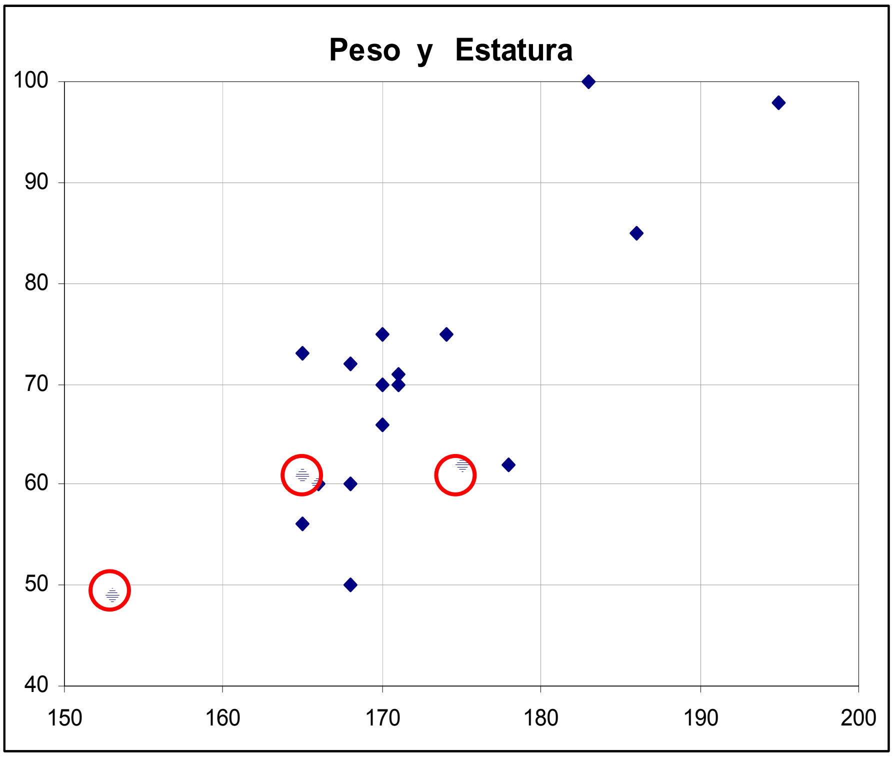


De estos resúmenes gráfico-numéricos se puede concluir que X e Y no son independientes pero además, se puede sugerir el tipo de relación que mantienen:

1. Valores bajos de una aparecen en conjunto con valores bajos de la otra y

2. Valores altos de una aparecen en conjunto con valores altos de la otra.

En otros términos, estas variables presentan una tendencia monótona creciente.



En este diagrama de dispersión se han marcado, como un elemento adicional las tres observaciones (casos) que corresponden a las personas de sexo femenino en el banco de datos.

De entre todas las posibles tendencias monótonas existe una que por su simplicidad resulta muy conveniente para describir el comportamiento de un conjunto de datos. Se trata de la Tendencia Lineal.

Si se cuenta con la información de dos variables cuantitativas para n casos de forma que el banco incluye los pares (x_1, y_1), (x_2, y_2), … (x_n, y_n) entonces se dice que (en el banco) las variables tienen una relación lineal si existen dos constantes a y b tales que la ecuación

$$y_i = a + bx_i$$

se cumple de manera exacta para todos y cada uno de los casos.

En la realidad es muy poco frecuente que dos variables en un banco de datos tengan una relación lineal. Es decir, es poco probable que la ecuación correspondiente se satisfaga en forma exacta para todos los casos en el banco.

Sin embargo, ocurre en muchas aplicaciones que la relación se presenta en forma aproximada. Esto es, la ecuación se cumple o casi se cumple para una fracción importante de los casos en el banco. En esas condiciones se dice que los datos presentan una Tendencia Lineal.

Precisamente debido al hecho de que la tendencia se refiere a una relación aproximada, es conveniente contar con resúmenes que cuantifiquen el grado en que esa tendencia reproduce una relación exacta. Para el caso lineal el resumen que se emplea se conoce con el nombre de Coeficiente de Correlación y se define como sigue:

$$R_{X,Y} = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$$

en donde

$$S_{XY} = \sum(x_i - \bar{x})(y_i - \bar{y}),$$
$$S_{XX} = \sum(x_i - \bar{x})^2 \quad y$$
$$S_{YY} = \sum(y_i - \bar{y})^2 .$$

Entre las propiedades más importantes del coeficiente de correlación se encuentran las siguientes:

1. $-1 \leq R_{X,Y} \leq 1$

2. $R_{X,Y} = 1$ sólo si la ecuación $y_i = a + bx_i$ se cumple para todos los casos en el banco y la constante b es positiva.

3. $R_{X,Y} = -1$ sólo si la ecuación $y_i = a + bx_i$ se cumple para todos los casos en el banco y la constante b es negativa.

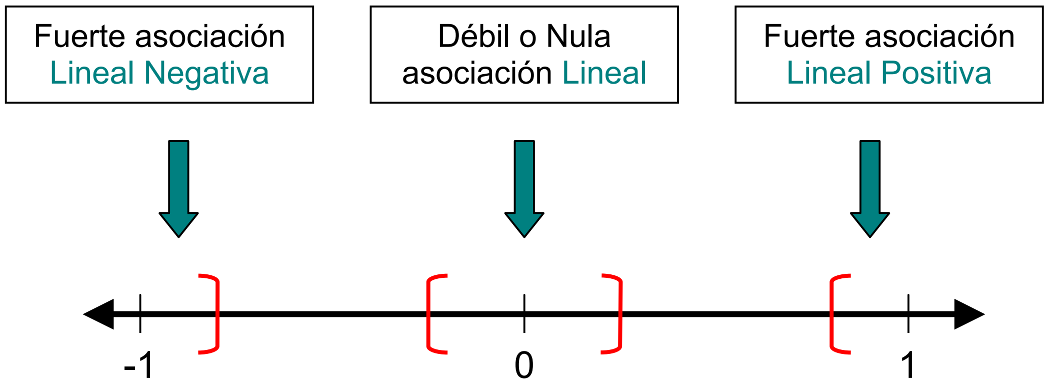
4. $R^2_{X,Y}$ se conoce como coeficiente de determinación y se interpreta como la proporción de la variabilidad de Y (X) que está explicada linealmente por X (Y).

Es necesario tener una idea perfectamente clara de la interpretación que se puede dar al valor que se obtiene cuando se calcula un coeficiente de correlación o de determinación.

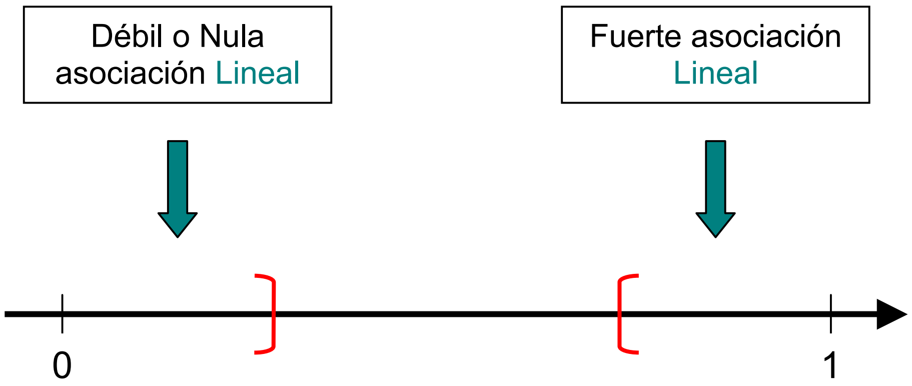
En primer lugar es conveniente tener presente que los coeficientes de correlación y determinación únicamente miden asociación lineal, sin embargo pueden existir casos en los que los datos presenten una tendencia de otro tipo (cuadrático, exponencial, periódico, etc.).

De esta forma, los valores de R y de R^2 sólo pueden emplearse para reportar que los datos presentan evidencia de una asociación lineal fuerte o débil (posiblemente nula). En ningún caso la ausencia de asociación lineal garantiza que no exista otro tipo de asociación.

El caso de R:



El caso de R^2:



Para los datos del ejemplo se tiene que

$S_{XY} = 1824.5$,

$S_{XX} = 1498.4$ y

$S_{YY} = 3423.2$;

Por tanto,

$$R_{X,Y} = 0.806 \quad y \quad R^2_{X,Y} = 0.649 .$$

De esta manera, se puede afirmar que existe evidencia de asociación lineal positiva en los datos. Más específicamente, el patrón de asociación lineal con el Peso (la Estatura) explica el 65% de la variabilidad de la Estatura (el Peso).

No existe una regla universal para juzgar cuando un coeficiente de correlación es suficientemente alto (o bajo). Los procedimientos estadísticos de inferencia que se han diseñado para el análisis de asociación lineal comúnmente se ocupan solo de establecer si la afirmación R = 0 tiene sustento en los datos. En términos generales si R no es cero depende del contexto particular del estudio y sus objetivos la consideración de si el nivel de asociación es útil o no.

Capítulo 3

Probabilidad

Probabilidad

La Probabilidad es una medida de la incertidumbre

La incertidumbre que mide es la asociada a la eventual ocurrencia de sucesos inciertos.

Toda medida requiere de un patrón de referencia que posibilite la interpretación de los resultados que produce.

En el caso de la Probabilidad se ha adoptado un patrón que se basa en las propiedades de las frecuencias relativas.

De esta manera, la probabilidad se caracteriza a través de la forma como opera, es decir a través de sus propiedades.

Sea A un evento incierto, es decir uno cuya ocurrencia no necesariamente es segura. Entonces, la probabilidad de que A ocurra se denota como P(A) y esta medida debe cumplir las siguientes propiedades básicas:

i) $0 \leq P(A) \leq 1$.

ii) $P(A) = 1$ sólo si A ocurre seguro.

iii) $P(A) = 0$ sólo si A no ocurre seguro.

Otras propiedades, que involucran dos eventos son las siguientes. Sean A y B dos eventos inciertos.

iv) Si A ocurre siempre que ocurre B ($B \subseteq A$) entonces, $P(B) \leq P(A)$.

v) Si sucede que A y B no pueden ocurrir simultáneamente ($B \cap A = \phi$) entonces, $P(A \cup B) = P(A) + P(B)$.

Además, es conveniente introducir la siguiente definición: Sean A y B dos eventos inciertos de forma que $P(A) \neq 0$. Entonces, la Probabilidad Condicional de B dado A se define como

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Con esta definición, se puede introducir una más. Sean A y B dos eventos. Entonces se dice que A y B son independientes si se tiene que

$$P(A \cap B) = P(A) \times P(B)$$

y en ese caso, $P(B \mid A) = P(B)$.

Todas estas propiedades han sido adoptadas con el propósito de reproducir, para las probabilidades, el comportamiento que tienen las frecuencias relativas. De hecho, en los casos más simples las probabilidades se calculan directamente como frecuencias relativas.

Existe, sin embargo, una inmensa variedad de situaciones en donde las frecuencias no están disponibles en la forma necesaria y es ahí donde las probabilidades cobran importancia ya que, en particular, se pueden calcular utilizando otros procedimientos.

Para empezar, no todos los fenómenos inciertos se manifiestan a través de eventos en donde sólo es relevante si estos ocurren o no. En muchas aplicaciones, el fenómeno aleatorio bajo estudio puede producir distintos resultados e interesa el análisis de todos los resultados posibles.

Aparece entonces, asociado a la observación de un fenómeno aleatorio, el concepto de variable aleatoria. Como en el caso del AE, una variable no es más que la codificación numérica de los posibles resultados que se derivan de la observación de un fenómeno.

De la misma forma que en el AE, las variables -ahora aleatorias- pueden clasificarse como cualitativas y cuantitativas, o con mas detalle en nominales, ordinales, discretas y continuas, de acuerdo con su naturaleza.

El interés se concentra en la descripción de la incertidumbre asociada al fenómeno aleatorio bajo estudio, es decir, la incertidumbre asociada a sus posibles resultados. Por su parte, la variable aleatoria asigna un valor distinto a cada posible resultado diferente.

Entonces, el problema equivale a describir la incertidumbre asociada a la variable aleatoria, es decir a la ocurrencia de sus distintos valores.

Ahora bien, si la ocurrencia de cada posible valor es incierta, la incertidumbre correspondiente se puede describir a través de la probabilidad respectiva.

Como conclusión, la incertidumbre asociada a una variable aleatoria queda descrita en cuanto se describe el conjunto de valores que puede producir y la probabilidad asociada con cada valor.

En este punto es conveniente introducir dos nuevos términos:

1. Al conjunto de los valores que puede producir una variable aleatoria X se le conoce como el Soporte de la variable y habitualmente se le denota como $\boldsymbol{\mathcal{X}}$.

2. Al soporte, junto con la relación que asocia a cada valor de la variable aleatoria su probabilidad de ocurrencia, se le conoce con el nombre de Función de Probabilidad y habitualmente se denota como P_X, $P(X)$ ó $P(X = x)$.

Con estos conceptos se tiene que la incertidumbre asociada a una variable aleatoria queda descrita en cuanto se conoce su función de probabilidad.

Imagine una variable aleatoria X que, al ser observada, produce uno de los siguientes valores: x_1, x_2, x_3, x_4, x_5 y x_6 .

Como ya se ha indicado, la incertidumbre asociada a X queda descrita en cuanto se informa de los valores x_1 a x_6 y de los valores de probabilidad $P(X = x_1)$, … $P(X = x_6)$.

Esta información puede reportarse a través de una tabla como sigue:

Función de Probabilidad de X

X	**P(X = x)**
x_1	$P(X = x_1)$
x_2	$P(X = x_2)$
x_3	$P(X = x_3)$
x_4	$P(X = x_4)$
x_5	$P(X = x_5)$
x_6	$P(X = x_6)$
	1.00

Equivalentemente, si se abrevia para a cada valor de X, $p_i = P(X = x_i)$ la tabla tiene el siguiente aspecto:

Función de Probabilidad de X

X	P(X = x)
x_1	p_1
x_2	p_2
x_3	p_3
x_4	p_4
x_5	p_5
x_6	p_6
	1.00

Esta representación establece una similitud evidente entre la función de probabilidad y la tabla de frecuencias relativas. Es muy importante percibir ahora las diferencias.

La diferencia más importante entre las frecuencias y las probabilidades estriba en el hecho de que las primeras describen la información observada (ya registrada) en un banco de datos, mientras que las probabilidades describen la manera en como podría registrarse la información por observarse (futura) de una variable aleatoria.

X	fr / p
0	0.10
1	0.30
2	0.20
3	0.20
4	0.15
5	0.05
	1.00

En la tabla que antecede, si los valores de la segunda columna fuesen frecuencias relativas, estos se podrían interpretar, por ejemplo, como sigue:

1. Un 15% de los casos en el banco presentó un valor de 4.

2. El valor que se presentó con más frecuencia (la moda) fue el 1 (30%).

Por otra parte, esta tabla debiese incluir el tamaño del banco (n).

X	fr / p
0	0.10
1	0.30
2	0.20
3	0.20
4	0.15
5	0.05
	1.00

Si, por el contrario, la tabla se refiriese a una función de probabilidad, la interpretación seguiría otra línea. Por ejemplo:

1. Con una probabilidad de 0.15 se espera que, al observar la variable, se presente un valor de 4.

2. El valor que se espera con mayor probabilidad (la moda) es el 1 (0.30).

En este caso la noción de tamaño del banco simplemente no existe.

Las características más notables de este ejemplo se pueden generalizar para concluir que, así como la tabla de frecuencias es un resumen eficiente (suficiente y minimal) de la información contenida en un banco de datos, la función de probabilidad constituye el resumen correspondiente si el propósito es describir la incertidumbre asociada a una variable aleatoria.

Por otra parte, una vez que ha sido establecida esta correspondencia entre tabla de frecuencias y función de probabilidad, surge de manera natural la idea de transportar también al terreno de la probabilidad otros resúmenes propios del Análisis Exploratorio.

El objetivo es el mismo: proveer una descripción, en este caso del comportamiento de la variable aleatoria. El concepto también coincide: emplear resúmenes parciales que si bien no capturan toda la información relevante, destacan algún aspecto que puede ser de especial interés.

De esta manera es posible referirse a las medidas de localización y dispersión para variables aleatorias introduciendo algunas modificaciones en las definiciones correspondientes.

Medidas de Localización
(para variables aleatorias)

Moda. Es el valor más probable de X.

Mediana. Es el valor $X_{(0.5)}$, en el soporte de X, que satisface simultáneamente las dos siguientes condiciones:

$$P(X \leq X_{(0.5)}) \geq 0.5$$
$$P(X \geq X_{(0.5)}) \geq 0.5 .$$

Cuantil (de orden q)**.** Es el valor $X_{(q)}$, en el soporte de X, que satisface simultáneamente las dos siguientes condiciones:

$$P(X \leq X_{(q)}) \geq q$$
$$P(X \geq X_{(q)}) \geq 1-q .$$

Mínimo. Es el valor más pequeño, $X_{[-]}$, en el soporte de X.

Máximo. Es el valor más grande, $X_{[+]}$, en el soporte de X.

Media (Valor Esperado ó Esperanza)**.** Es un promedio ponderado de los valores en el soporte de X donde las probabilidades respectivas sirven como pesos. Por ejemplo, si la variable X produce los valores x_1, x_2, x_3, x_4, x_5 y x_6 con probabilidades p_1, p_2, p_3, p_4, p_5 y p_6 respectivamente, entonces la media de X se calcula como:

$$E(X) = \sum_{i=1}^{6} p_i x_i$$

Como puede observarse, todas las medidas de localización originalmente diseñadas para servir propósitos descriptivos en el AE tienen una versión equivalente en Probabilidad que se obtiene, para todo propósito práctico, sustituyendo frecuencias relativas por probabilidades.

Posiblemente el caso en el que esta identificación resulta menos clara es la media. Aparentemente se tiene una diferencia relevante cuando aquí se define como un promedio ponderado mientras que en AE se definió como un promedio aritmético simple. Esta diferencia es sólo aparente.

Recuerde que en un banco de datos puede presentarse la repetición de valores. De hecho, el propósito del cálculo de frecuencias es precisamente dar cuenta de esas repeticiones.

Por otra parte, en el soporte de una variable solamente se incluyen los valores distintos de X.

Así pues si se observa que la definición de la media en AE se refiere al promedio sobre todos los casos en el banco, mientras que en Probabilidad se consideran los valores en el soporte, es inmediato comprobar que las dos definiciones coinciden.

Considere el siguiente conjunto de datos que describe los años transcurridos desde la obtención del título profesional para el grupo.

Años
4
3
5
5
3
9
0
5
4
4
9
3
15
3
3
5
5
6
2

Compruebe que si calcula la media utilizando, primero, la fórmula

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

en donde n es el número de casos y después construye la tabla de frecuencias relativas, sin agrupar, y calcula ahora

$$E(X) = \sum_{i=1}^{k} fr_i x_i$$

donde k es el número de valores distintos en el banco (de renglones en la tabla de frecuencias), obtiene el mismo resultado.

De la misma manera se puede proceder con las medidas de dispersión.

Medidas de Dispersión
(para variables aleatorias)

Rango. Es el valor R que se calcula como la diferencia entre los valores máximo y el mínimo de la variable:

$$\mathbf{R} = X_{[+]} - X_{[-]}.$$

Rango entre cuartiles. Es el valor REC, que se calcula como la diferencia entre los cuantiles de orden 0.75 (tercer cuartil) y 0.25 (primer cuartil):

$$\mathbf{REC} = X_{(0.75)} - X_{(0.25)}.$$

Error Medio. Es un promedio ponderado de las diferencias que guardan, en valor absoluto, los valores en el soporte de X respecto a una medida de localización. Las probabilidades respectivas sirven como pesos y por ejemplo, si la variable X produce los valores x_1, x_2, x_3, x_4, x_5 y x_6 con probabilidades p_1, p_2, p_3, p_4, p_5 y p_6 respectivamente, y $X_{(0.5)}$ es la mediana de X, entonces el **EM**, respecto a la mediana, se calcula como:

$$EM(X) = \sum_{i=1}^{6} p_i \left| x_i - X_{(0.5)} \right|$$

Varianza. Es el valor σ^2 que se calcula como un promedio ponderado de las diferencias que guardan, elevadas al cuadrado, los valores en el soporte de X respecto a la media de la variable. Las probabilidades respectivas sirven como pesos y por ejemplo, si la variable X produce los valores x_1, x_2, x_3, x_4, x_5 y x_6 con probabilidades p_1, p_2, p_3, p_4, p_5 y p_6 respectivamente, entonces la varianza de X se calcula como:

$$\sigma^2(X) = \mathrm{Var}(X) = \sum_{i=1}^{6} p_i (x_i - \overline{X})^2 .$$

Desviación estándar. Es el valor σ que se calcula como la raíz cuadrada de la varianza. De nuevo, por ejemplo, si la variable X produce los valores x_1, x_2, x_3, x_4, x_5 y x_6 con probabilidades p_1, p_2, p_3, p_4, p_5 y p_6, entonces la desviación estándar de X resulta:

$$\sigma(X) = [\mathrm{Var}(X)]^{1/2} = \sqrt{\sum_{i=1}^{6} p_i (x_i - \overline{X})^2} .$$

Tanto las medidas de localización como las medidas de dispersión que se han presentado para las variables aleatorias parten del supuesto de que el soporte $\mathcal{X}$ de la variable X esta formado por un número finito de posibles valores.

Estrictamente en la práctica, este supuesto es correcto. Sin embargo, existen modelos de probabilidad que son muy útiles y convenientes y que tienen un soporte con un número infinito de posibles valores.

En tal caso es necesario modificar, en cierto sentido, las definiciones de algunos de estos resúmenes. Considere el caso en que el soporte es infinito pero discreto. Es decir, el caso en que $\mathcal{X} = \{ x_1, x_2, x_3, \ldots \}$ donde los valores constituyen una secuencia de puntos aislados sin fin.

La primera consecuencia de esta estructura es que en correspondencia con los valores en ***X***, debe existir una secuencia, también infinita, de probabilidades $\{ p_1, p_2, p_3,\ldots \}$. Cada una de estos valores debe cumplir con las propiedades que caracterizan a las probabilidades y en conjunto deben satisfacer, la condición:

$$\sum_{i=1}^{\infty} p_i = 1.$$

En términos técnicos, se dice que la sucesión definida por la suma de las probabilidades debe converger a 1.

Además, todos los resúmenes que involucran promedios ponderados deben incorporar la modificación correspondiente. Esto significa que en el caso de variables aleatorias discretas con soporte infinito se tiene que:

$$E(X) = \sum_{i=1}^{\infty} p_i x_i \,,$$

$$EM(X) = \sum_{i=1}^{\infty} p_i \left| x_i - X_{(0.5)} \right| ,$$

$$\sigma^2 = \sum_{i=1}^{\infty} p_i (x_i - \overline{X})^2 \text{ y}$$

$$\sigma = \sqrt{\sum_{i=1}^{\infty} p_i (x_i - \overline{X})^2} \,.$$

Naturalmente, ahora cada una de estas medidas define una sucesión que, al menos en teoría, podría no converger. Por supuesto, los modelos más comunes y, en particular, los que se consideran en este texto no presentan esa anomalía.

Un caso ligeramente más sofisticado es el de las variables aleatorias continuas. Como en AE, se dice que una variable es continua si puede producir cualquier valor en un intervalo. Es decir, si el soporte $\boldsymbol{X}$ es un intervalo.

Conceptualmente, la dificultad más importante cuando se consideran variables continuas radica en el hecho de que si $\boldsymbol{X}$ es un intervalo, no existe forma de recorrer puntualmente, uno a la vez, los valores en el soporte para calcular los promedios requeridos.

Suponga, por ejemplo, que tiene una variable continua con soporte $\boldsymbol{\mathcal{X}}$ igual al intervalo [0, 10]. Evidentemente, el mínimo valor en el soporte es cero pero no es posible establecer cual es el que le sigue. Cualquier candidato, digamos A, queda descartado automáticamente si se reconoce que entre cero y el número A existe una infinidad de valores más.

La solución a este dilema está inspirada en la idea, ya bien conocida, de agrupar. Suponga que el soporte $\boldsymbol{\mathcal{X}}$ se divide en un número finito, digamos k, de subintervalos o clases. Suponga, además, que las probabilidades asociadas a todas y cada una de las clases son $p_1^A, p_2^A, \ldots, p_k^A$.

Si ahora se elige, en cada clase un valor de X como representante y se forma la colección $x_1^A, x_2^A, \ldots, x_k^A$ entonces una aproximación al valor de, por ejemplo, la media está dado por:

$$E^*(X) = \sum_{i=1}^{k} p_i^A x_i^A$$

Es intuitivamente claro que la calidad de esta aproximación será mejor en la medida en que el número de clases aumente y la longitud de todos y cada uno de los subintervalos sea cada vez menor. De esta manera, y desde un punto de vista técnico de nuevo, se tiene que la Media se puede obtener como resultado de un proceso límite:

$$\sum_{i=1}^{k} p_i^A x_i^A \xrightarrow{k \to \infty} E(X)$$

Ahora bien, es un resultado bien conocido del cálculo integral que una suma como la que se ha construido, cuando converge, lo hace a una integral. En este caso,

$$\sum_{i=1}^{k} p_i^A x_i^A \xrightarrow{k \to \infty} \int_{\mathcal{X}} x\, f(x)\, dx$$

en donde f(x) es una función con las siguientes características:

1. $f(x) \geq 0$ para toda x en $\mathcal{X}$.

2. $\int_{\mathcal{X}} f(x)\, dx = 1.$

A la función f(x) se le conoce como la función de densidad de probabilidad de X.

La otra propiedad importante de f(x) es la siguiente: Para cualquier intervalo [a, b] que se encuentre contenido en el soporte de X, la probabilidad del evento A = $\{a \leq X \leq b\}$ se calcula como:

$$P(A) = \int_a^b f(x)\,dx\,.$$

Esta propiedad implica que, en particular, un peculiar resultado según el cual si x es un valor fijo cualquiera de una variable aleatoria X continua entonces

$$P(\ X = x\ ) = 0.$$

Es decir, cualquier modelo de probabilidad para variables continuas:

1. Asigna a un intervalo una probabilidad que se calcula como el área bajo la curva definida por la función de densidad.

2. A todo valor aislado de la variable le asigna probabilidad cero.

Es importante observar que las variables aleatorias continuas no cuentan con función de probabilidad; en su lugar se tiene la función de densidad de probabilidad.

Comente cuales son las similitudes y diferencias entre la función de probabilidad (para variables discretas) y la función de densidad (para variables continuas).

Existe otra función definida para ambos tipos de variables: La Función de Distribución. Si X es una variable aleatoria y x es un valor cualquiera, entonces, la función de distribución de X evaluada en x se define como

$$F_X(x) = P(X \leq x).$$

A la Función de Distribución también se le llama función de probabilidad acumulada y la forma de cálculo es distinta para las variables discretas y continuas.

Caso discreto:

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} P(X = y).$$

Caso Continuo:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx.$$

Volviendo al problema de describir una variable aleatoria, ahora es posible plantear la descripción en términos más generales que incluyan tanto a las variables discretas como a las continuas:

Una variable aleatoria está completamente descrita en cuanto se define su soporte y su función de distribución.

En relación con este tema vale la pena observar que existe una cantidad incontable de variables aleatorias.

De hecho, cualquier colección de números positivos cuya suma sea finita puede dar lugar a una función de probabilidad y por tanto a una variable discreta.

Por su parte cualquier función positiva cuya área bajo la curva sea, también, finita puede dar lugar a una función de densidad y por tanto a una variable continua.

Una forma muy conveniente de trabajar con variables aleatorias es a través de modelos.

Un modelo (paramétrico) para una variable aleatoria es una ecuación que permite el cálculo de una función de distribución (o de probabilidad o de densidad) una vez que se fijan los valores de un conjunto de índices o parámetros.

Alguno de los modelos parámetricos más comunes son los siguientes:

Variables discretas.

1. Uniforme.
2. Triangular.
3. Bernoulli
4. Binomial.
5. Poisson.
6. Geométrica
7. Binomial Negativa.

Variables Continuas.

1. Uniforme
2. Triangular.
3. Exponencial,
4. Normal,
5. Ji cuadrada
6. **t** de Student
7. F.

Algunos de estos modelos, o mejor dicho, algunas de estas familias de modelos ya han sido presentadas. Sin embargo y para tener un panorama conjunto, aquí se revisan nuevamente.

Modelo Uniforme Discreto.

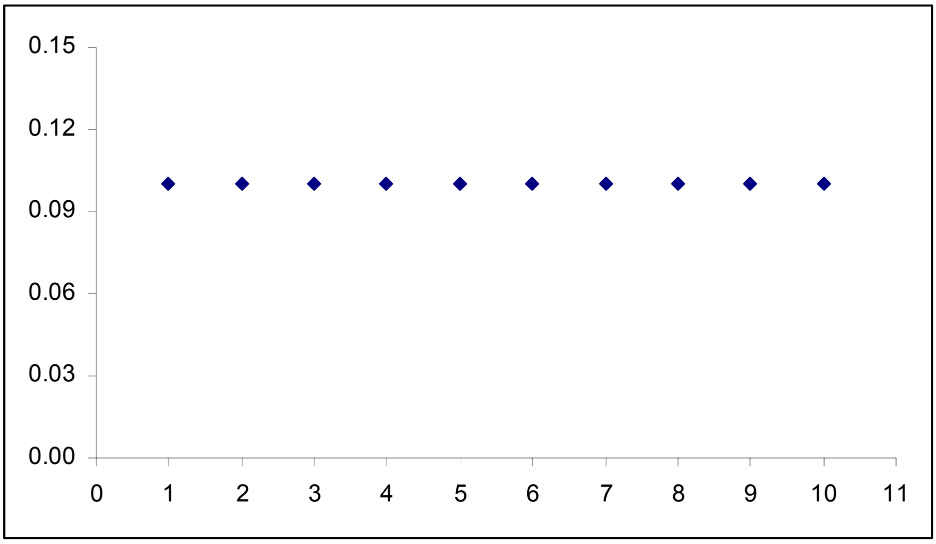
Soporte:

$$\mathcal{X} = \{1, 2, 3, \ldots N\}.$$

Función de Probabilidad:

$$P(X = x) = 1/N \quad \text{para todo } x \text{ en } \mathcal{X}.$$

Función de Probabilidad
Uniforme Discreta (N=10)



Modelo Uniforme Discreto.

Esperanza:

$$E(X) = (N+1)/2$$

Varianza:

$$Var(x) = (N^2 - 1)/12$$

Esta familia tiene tantos elementos como valores existen de N enteros mayores que 1.

Modelo Triangular Discreto.

Soporte:

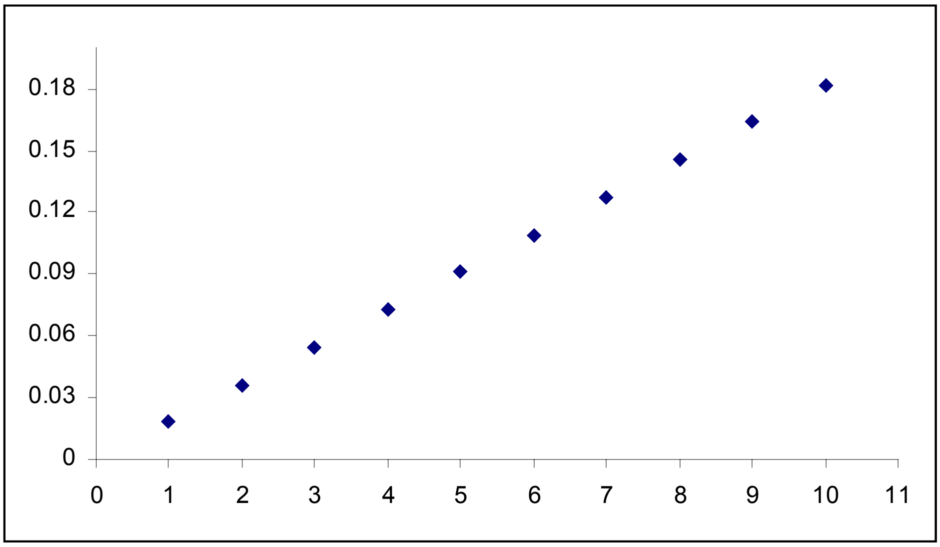
$$\boldsymbol{X} = \{1, 2, 3, \ldots N\}.$$

Función de Probabilidad:

$$P(X = x) = bx \quad \text{para todo } x \text{ en } \boldsymbol{X}.$$

(b = 2/{(N(N+1)}).

Función de Probabilidad
Triangular Discreta (N=10)



Modelo Triangular Discreto.

Esperanza:

$$E(X) = (2N+1)/3$$

Varianza:

$$Var(x) = \{(N-1)(N+1)\}/18$$

Esta familia de modelos tiene tantos elementos como valores N existen enteros y mayores que 1.

Modelo Bernoulli.

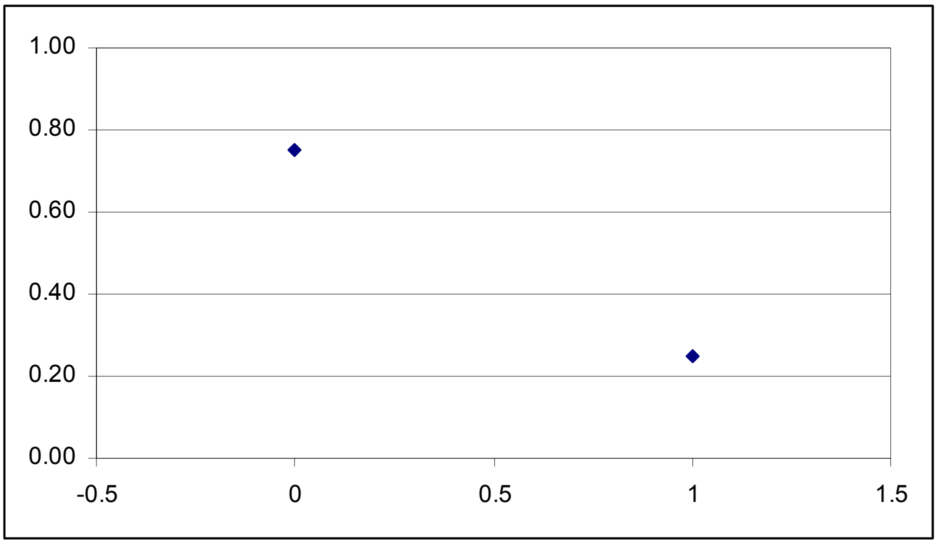
Soporte:

$$\boldsymbol{\mathcal{X}} = \{0, 1\}.$$

Función de Probabilidad:

$$P(X = x) = p^x (1-p)^{(1-x)} \quad \text{para todo } x \text{ en } \boldsymbol{\mathcal{X}}.$$

Función de Probabilidad
Bernoulli (p = 0.25)



Modelo Bernoulli.

Esperanza:

$$E(X) = p$$

Varianza:

$$Var(x) = p(1-p)$$

Esta familia de modelos tiene tantos elementos como valores existen de p en el intervalo [0,1].

Modelo Binomial.

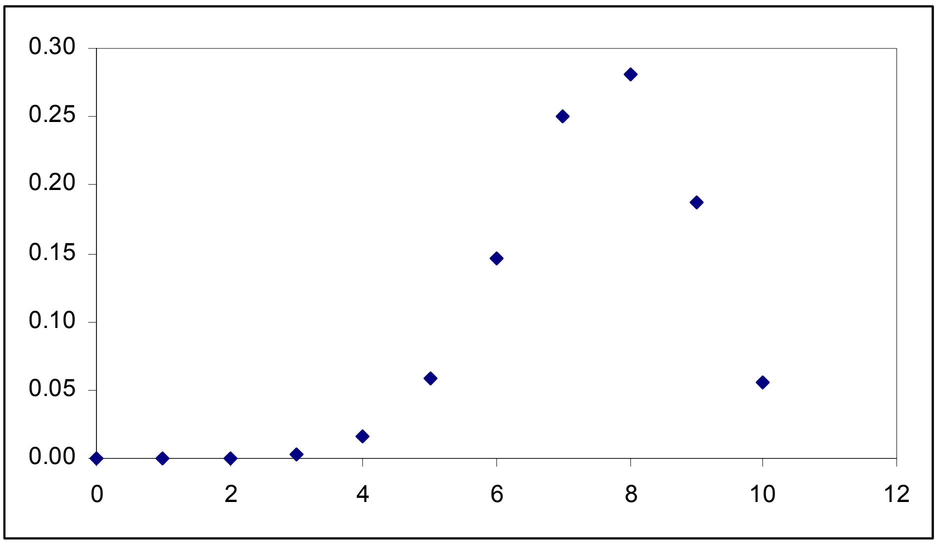
Soporte:

$$\mathcal{X} = \{0, 1, \ldots, N\}.$$

Función de Probabilidad:

$$P(X = x) = \binom{N}{x} p^x (1-p)^{N-x} \qquad \text{para todo } x \text{ en } \mathcal{X}.$$

Función de Probabilidad
Binomial (p = 0.75, N = 10)



Modelo Binomial.

Esperanza:

$$E(X) = Np$$

Varianza:

$$Var(x) = Np(1-p)$$

Esta familia de modelos tiene tantos elementos como parejas existen de valores de N enteros positivos y de p en el intervalo [0,1].

Modelo Poisson.

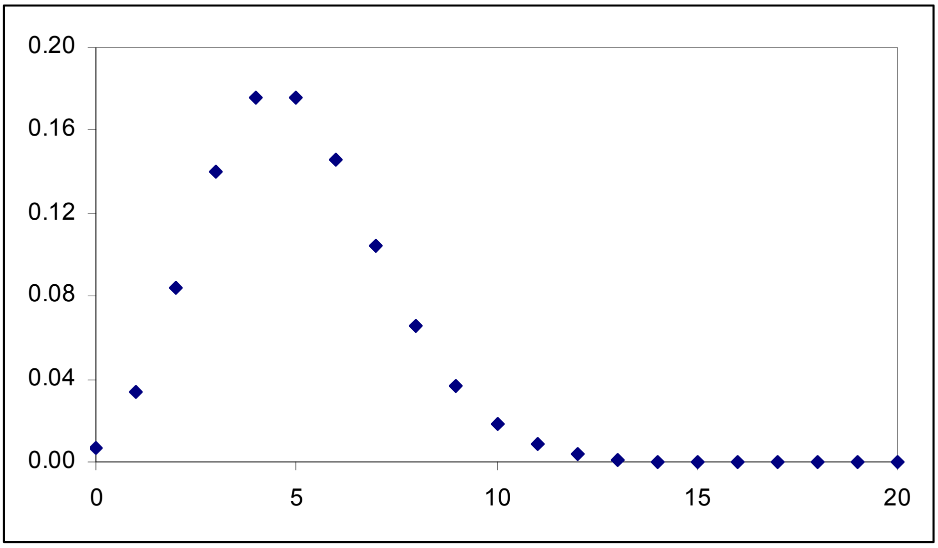
Soporte:

$$\boldsymbol{\mathcal{X}} = \{0, 1, 2, \ldots \}.$$

Función de Probabilidad:

$$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!} \qquad \text{para todo } x \text{ en } \boldsymbol{\mathcal{X}}.$$

Función de Probabilidad
Poisson (λ = 5)



Modelo Poisson.

Esperanza:

$$E(X) = \lambda$$

Varianza:

$$Var(x) = \lambda$$

Esta familia de modelos tiene tantos elementos como valores de λ existen positivos.

Modelo Geométrico.

Soporte:

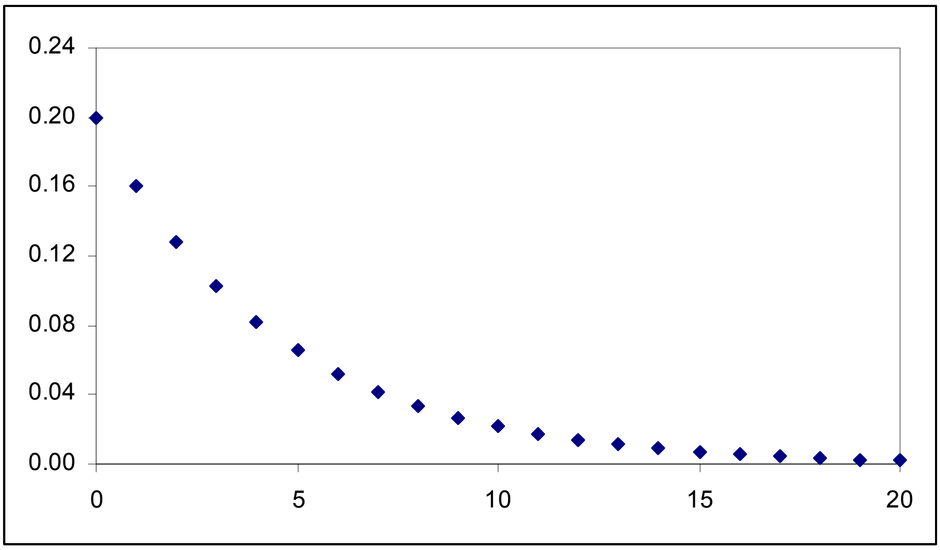
$$\boldsymbol{X} = \{0, 1, 2, \ldots\}.$$

Función de Probabilidad:

$$P(X = x) = pq^{x} \qquad \text{para todo } x \text{ en } \boldsymbol{X}.$$

($q = (1-p)$).

Función de Probabilidad Geométrica (p = 0.2)



Modelo Geométrico.

Esperanza:

$$E(X) = q / p$$

Varianza:

$$Var(x) = q / p^2$$

Esta familia de modelos tiene tantos elementos como valores de p existen en [0, 1].

Modelo Binomial Negativa.

Soporte:

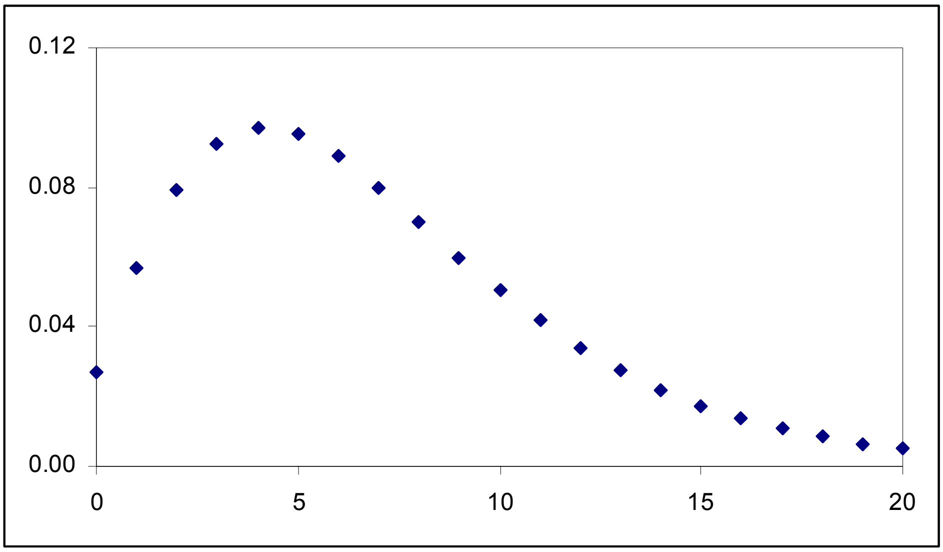
$$\boldsymbol{\mathcal{X}} = \{0, 1, 2, \ldots \}.$$

Función de Probabilidad:

$$P(X = x) = \binom{r + x - 1}{x} p^r (1 - p)^x$$

para todo x en $\boldsymbol{\mathcal{X}}$.

Función de Probabilidad
Binomial Negativa (r = 3, p = 0.3)



Modelo Binomial Negativa.

Esperanza:

$$E(X) = r\ (q\ /\ p)$$

Varianza:

$$Var(x) = r\ (q\ /\ p^2)$$

Esta familia de modelos tiene tantos elementos como existen parejas de valores r entero y positivo p en [0, 1].

Modelo Uniforme Continua.

Soporte:

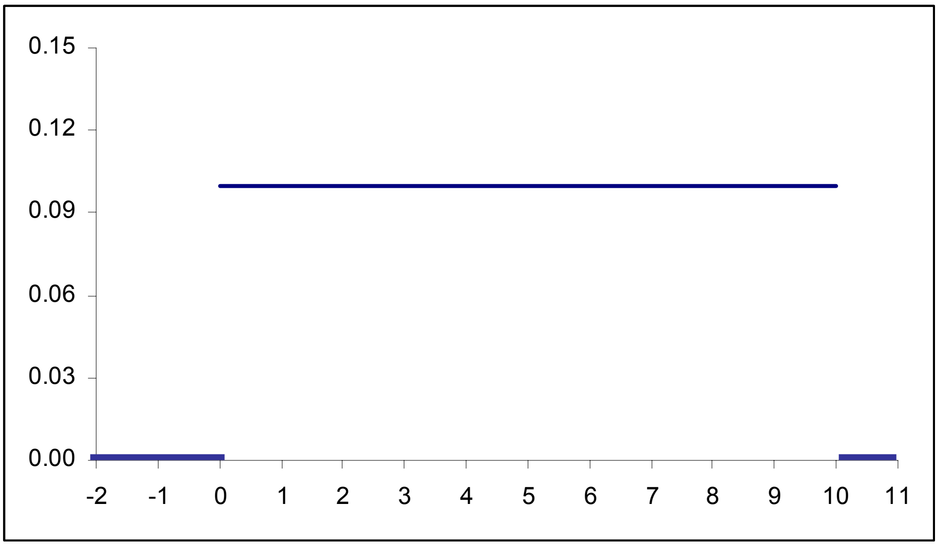
$$\boldsymbol{\mathcal{X}} = [a, b]. \qquad (a < b)$$

Función de Densidad:

$$f_X(x) = \frac{1}{(b-a)}$$

para todo x en $\boldsymbol{\mathcal{X}}$ y cero en otro caso.

Función de Densidad Uniforme
(a = 0, b = 10))



Modelo Uniforme Continua.

Esperanza:

$$E(X) = (a + b) / 2$$

Varianza:

$$Var(x) = (b - a)^2 / 12$$

Esta familia de modelos tiene tantos elementos como existen parejas de valores a y b con a < b.

Modelo Exponencial.

Soporte:

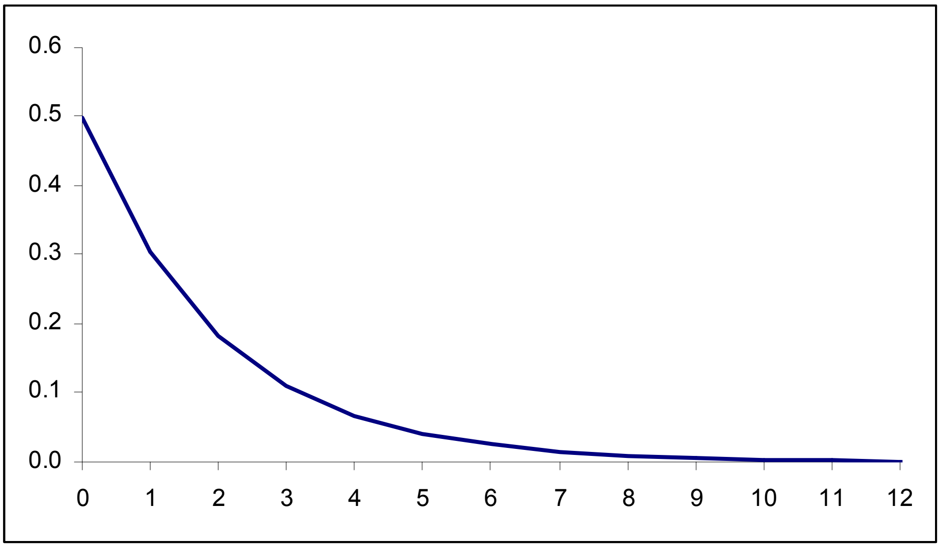
$$\mathcal{X} = (0, \infty).$$

Función de Densidad:

$$f_X(x) = \lambda \exp(-\lambda x)$$

para todo x en $\mathcal{X}$ y cero en otro caso.

Función de Densidad Exponencial
(λ= 0.5)



Modelo Exponencial.

Esperanza:

$$E(X) = \lambda^{-1}$$

Varianza:

$$Var(x) = \lambda^{-2}$$

Esta familia de modelos tiene tantos elementos como valores de λ positivos.

Modelo Normal.

Soporte:

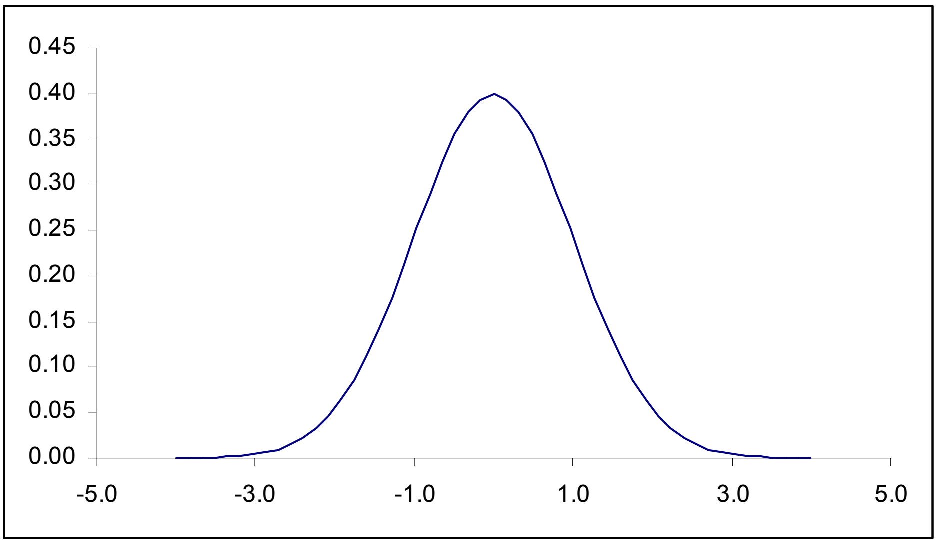
$$\mathcal{X} = (-\infty, \infty).$$

Función de Densidad:

$$\text{f(x)} = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

para todo x en $\mathcal{X}$.

Función de Densidad Normal
Estándar (μ= 0, σ^2 = 1)



Modelo Normal.

Esperanza:

$$E(X) = \mu$$

Varianza:

$$Var(x) = \sigma^2$$

Esta familia de modelos tiene tantos elementos como parejas existen de valores μ en los reales y σ^2 positivos.

Modelo Ji cuadrada (χ^2).

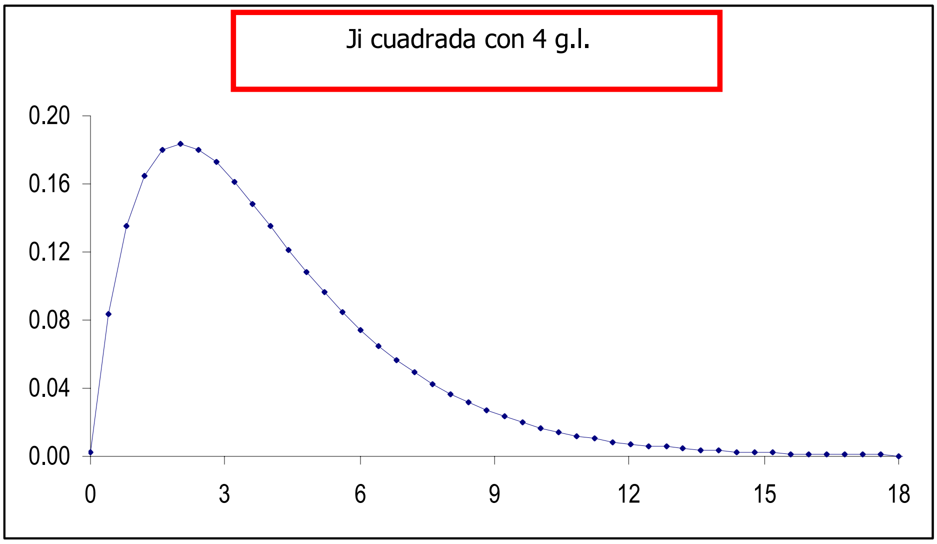
Soporte:

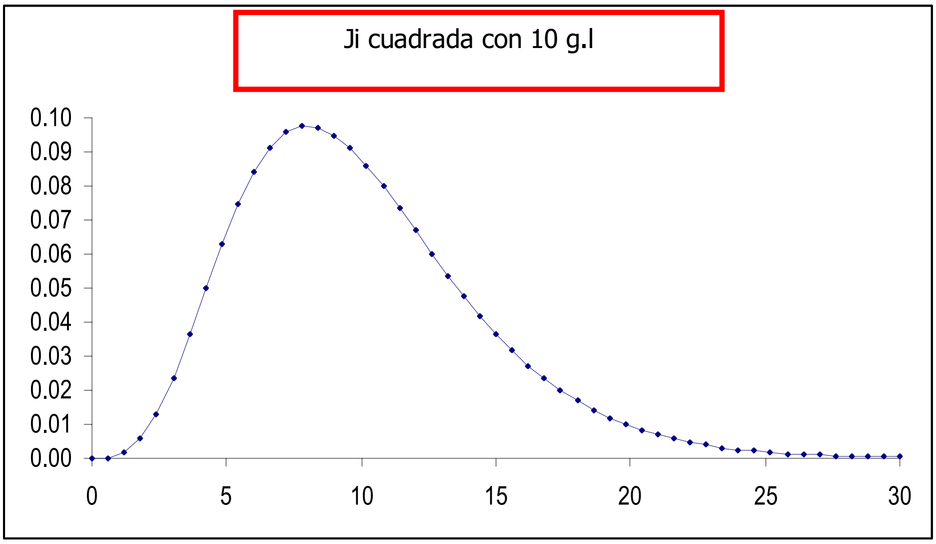
$$\boldsymbol{X} = (0, \infty).$$

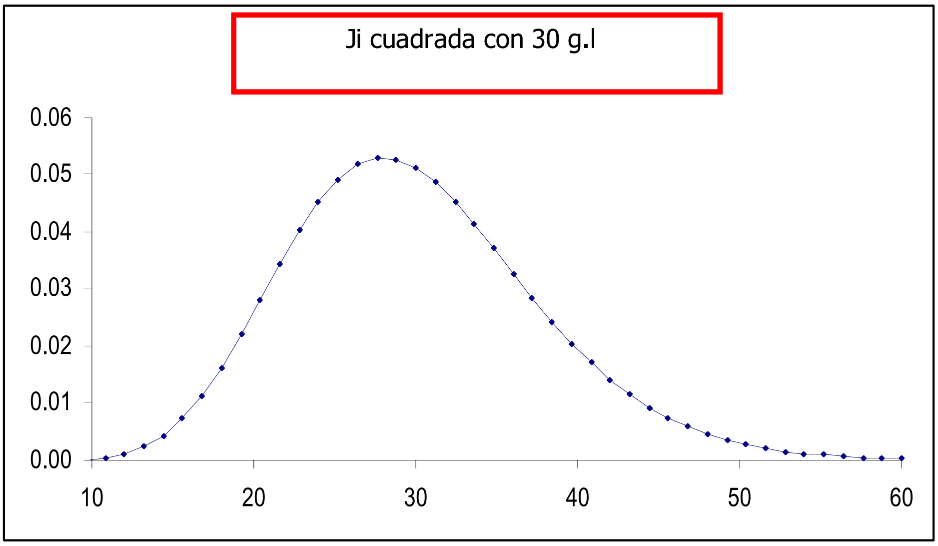
Función de Densidad:

$$f(x) = \frac{1}{\Gamma(r/2)} (2)^{-r/2}\, x^{(r/2-1)} \exp(-x/2)$$

para todo x en $\boldsymbol{X}$.


Ji cuadrada con 4 g.l.
0.20
0.16
0.12
0.08
0.04
0.00
0
3
6
9
12
15
18


Ji cuadrada con 10 g.l
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
30


Ji cuadrada con 30 g.l
0.06
0.05
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60

Modelo Ji cuadrada (χ^2).

Esperanza:

$$E(X) = r$$

Varianza:

$$Var(x) = 2\ r$$

Esta familia de modelos tiene tantos elementos como valores de r existen enteros positivos. Al parámetro r se le conoce como el número de grados de libertad o, mas brevemente, grados de libertad.

El modelo Ji cuadrada está relacionado con el modelo Normal en virtud de que si Z tiene una distribución Normal estándar, entonces Z^2 sigue una distribución Ji cuadrada con r = 1 (un grado de libertad).

Por otra parte, si U es Ji cuadrada con r grados de libertad y V es Ji cuadrada con s grados de libertad y, además U y V son independientes, entonces la variable W = U + V tiene una distribución Ji cuadrada con r + s grados de libertad.

Modelo **t** de Student.

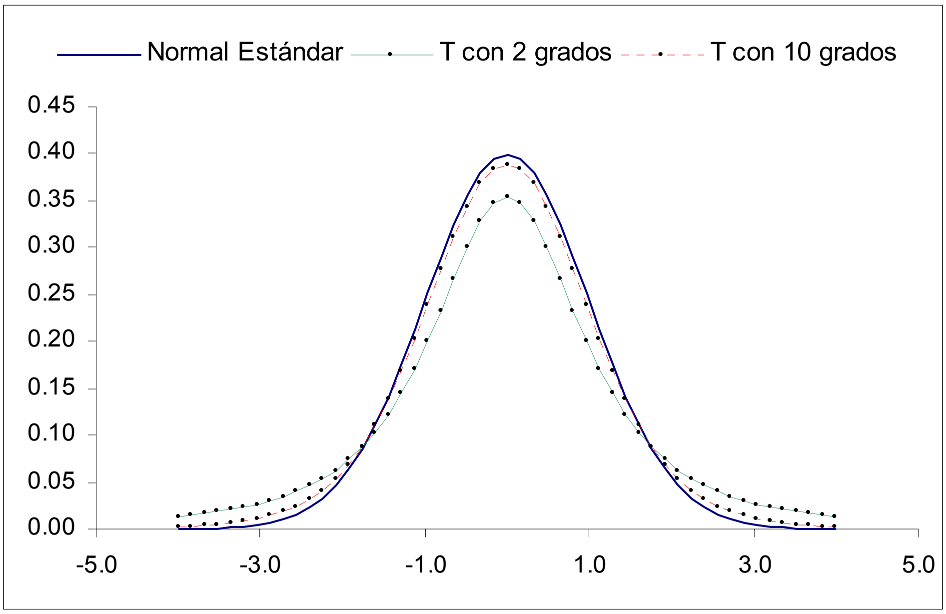
Soporte:

$$\mathcal{X} = (-\infty, \infty).$$

Función de Densidad:

$$f(x) = \frac{\Gamma[(r+1)/2]}{\Gamma(r/2)} (r\pi)^{-1/2} (1 + x^2/r)^{-(r+1)/2}$$

para todo x en $\mathcal{X}$.



Modelo **t** de Student.

Esperanza:

$$E(X) = 0 \qquad \text{(si } r > 1\text{)}$$

Varianza:

$$Var(x) = r / (r-2) \qquad \text{(si } r > 2\text{)}$$

Esta familia de modelos tiene tantos elementos como existen valores de r enteros positivos. A r se le conoce como los grados de libertad de la distribución.

El modelo **t** de Student se aproxima al modelo Normal estándar a medida que el número de grados de libertad aumenta.

Por otra parte, es interesante tener presente el siguiente resultado. Si Z es Normal estándar, V es Ji cuadrada con r grados de libertad y estas dos variables son independientes entonces

$$W = \frac{Z}{\sqrt{V / r}}$$

tiene una distribución **t** de Student con r grados de libertad.

Modelo F.

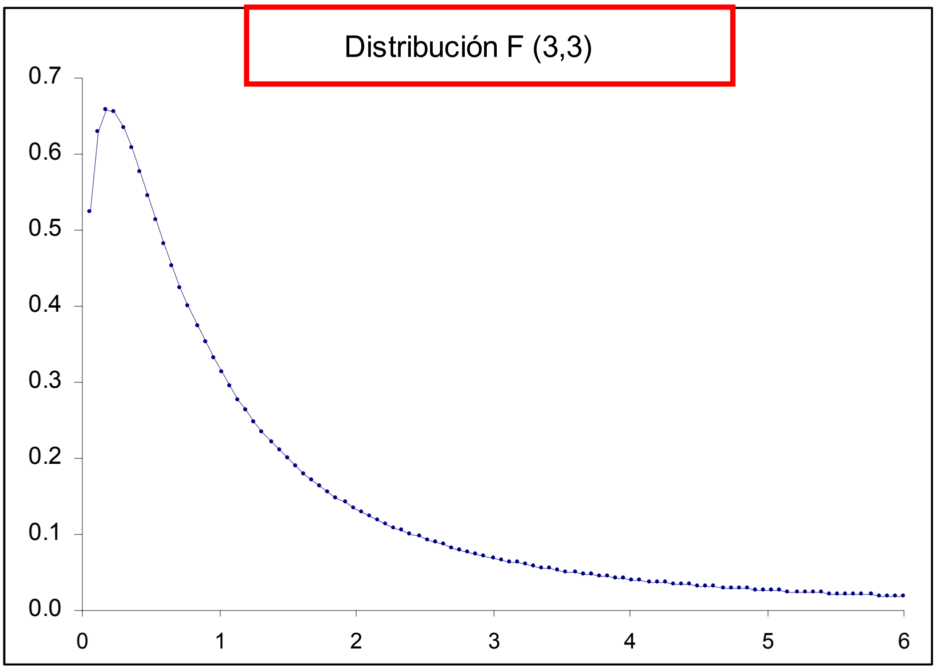
Soporte:

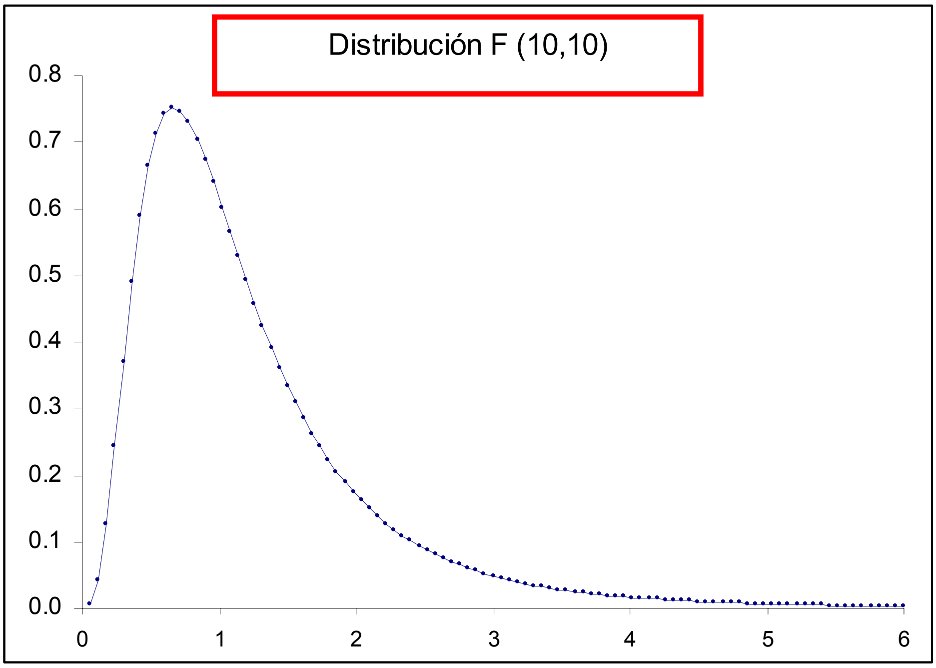
$$\mathcal{X} = (0, \infty).$$

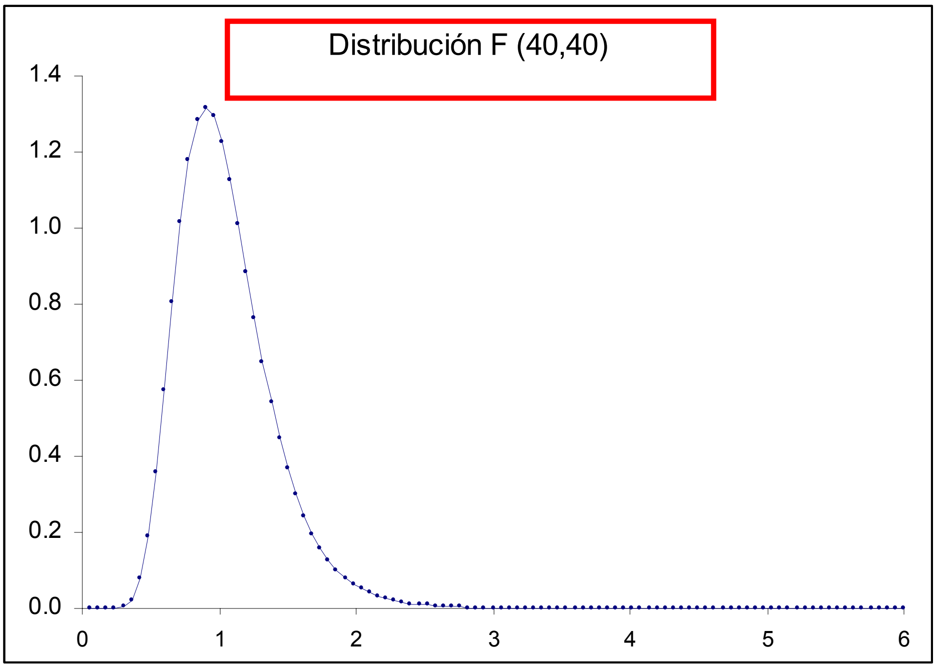
Función de Densidad:

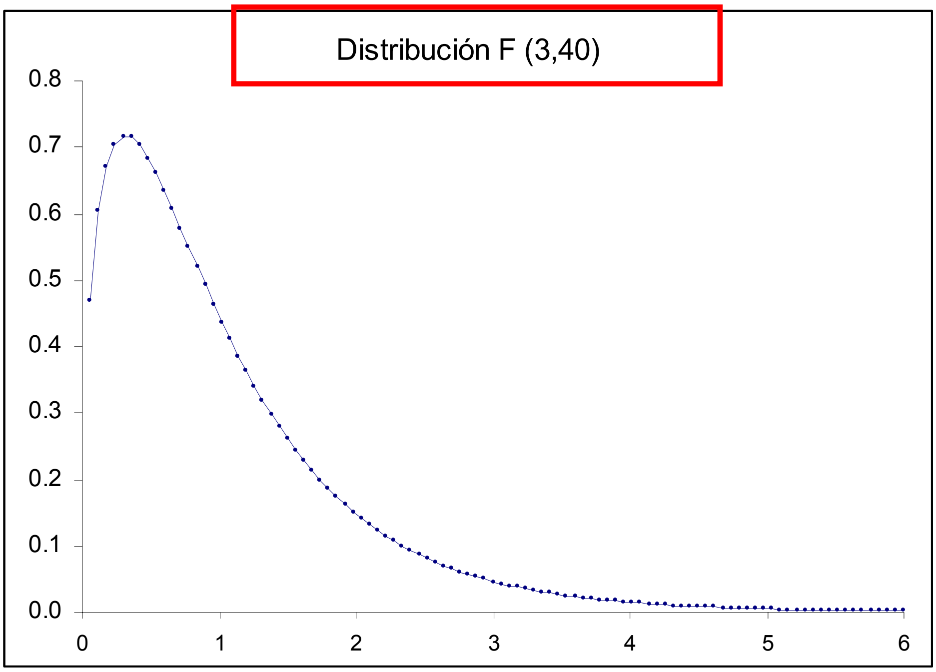
$$f(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}}$$

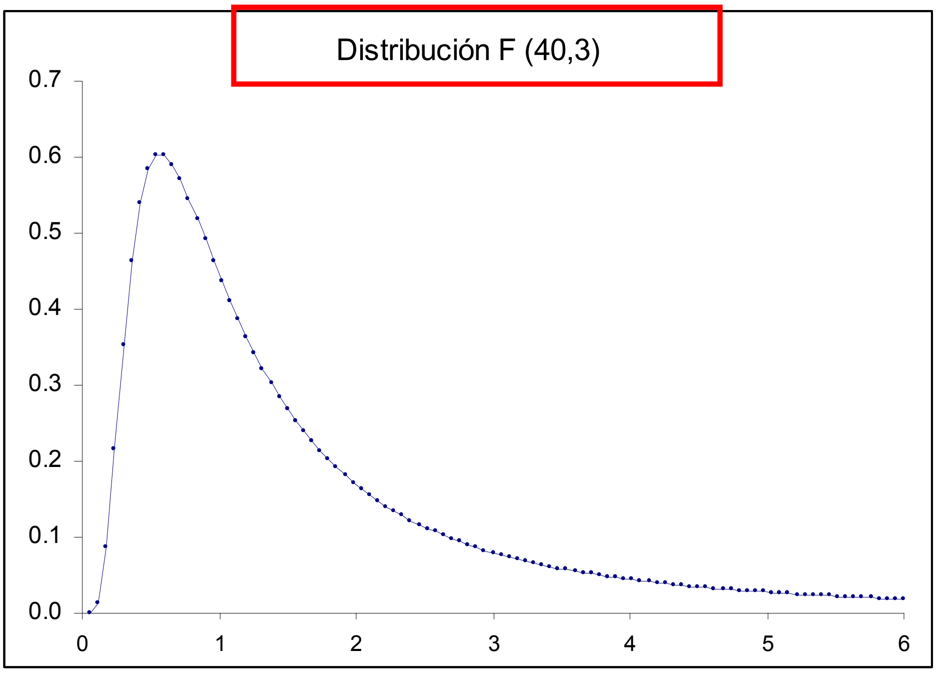
para todo x en $\mathcal{X}$.




Distribución F (10,10)


Distribución F (40,40)


Distribución F (3,40)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6


Distribución F (40,3)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6

Modelo F.

Esperanza:

$$E(X) = n / (n-2) \qquad \text{(si } n > 2\text{)}$$

Varianza:

$$Var(x) = [2n^2 (m+n-2)]/ [m(n-2)^2 (n-4)]$$

$$\text{(si } n > 4\text{)}$$

Esta familia de modelos tiene tantos elementos como parejas existen de enteros positivos m y n. Al parámetro m se le conoce como grados de libertad del numerador mientras que a n se le llama grados de libertad del denominador.

El modelo F está relacionado con el modelo Ji cuadrada y por tanto con el modelo Normal. Específicamente, si U y V son dos variables aleatorias independientes y U es Ji cuadrada con m grados de libertad mientras que V es Ji cuadrada con n grados de libertad, entonces la variable

$$W = \frac{U/m}{V/n}$$

sigue una distribución F con m grados de libertad en el numerador y n grados de libertad en el denominador.

Pronósticos Probabilísticos

La idea básica que da origen a la producción de pronósticos probabilísticos es la identificación de las frecuencias relativas con las probabilidades.

Como ya se ha indicado, las frecuencias tienen el propósito de describir lo que ya ocurrió, a partir de la información contenida en un banco de datos.

Por su parte, las probabilidades pretenden la descripción de lo que puede ocurrir en el futuro.

Identificar frecuencias con probabilidades, tiene sentido sólo si, en un nivel más primario, se acepta que lo que va a ocurrir es similar con lo que ya ha ocurrido.

Esta idea se traduce, cuando no se utilizan modelos, a suponer que el valor futuro se habrá de producir como si fuese resultado de un sorteo entre los valores que se han presentado en el pasado.

En ese caso, además, la probabilidad con la que se supone aparecerá un valor futuro es igual a la frecuencia relativa con la que se presentó en el banco de datos.

En otras palabras, el valor futuro se considera una observación de una variable aleatoria.

El soporte de la variable aleatoria está determinado por valores que aparecieron en el banco (y sólo esos) y su función de probabilidad por las frecuencias relativas correspondientes.

Este procedimiento tiene ventajas:

1. Es simple,

2. Hace uso del resumen eficiente,

3. No pone en duda la representatividad de la información en el banco.

Pero también implica desventajas:

1. No contempla la posibilidad de valores que no hayan ocurrido previamente,

2. No permite el contraste de pronósticos que proceden de bancos similares pero diferentes.

3. No toma en cuenta el procedimiento de construcción del banco.

En tales circunstancias es conveniente y resulta natural, el empleo de modelos para la producción de pronósticos probabilísticos.

Si el pronóstico se considera el producto de la observación de una variable aleatoria, es posible y razonable utilizar un modelo para describir el comportamiento de esa variable aleatoria.

Si, además, se considera conveniente utilizar un modelo relativamente simple y con propiedades que sean interpretables con facilidad, entonces una posibilidad son las familias paramétricas de modelos (Binomial, Poisson, Normal, Uniforme y Exponencial entre otras).

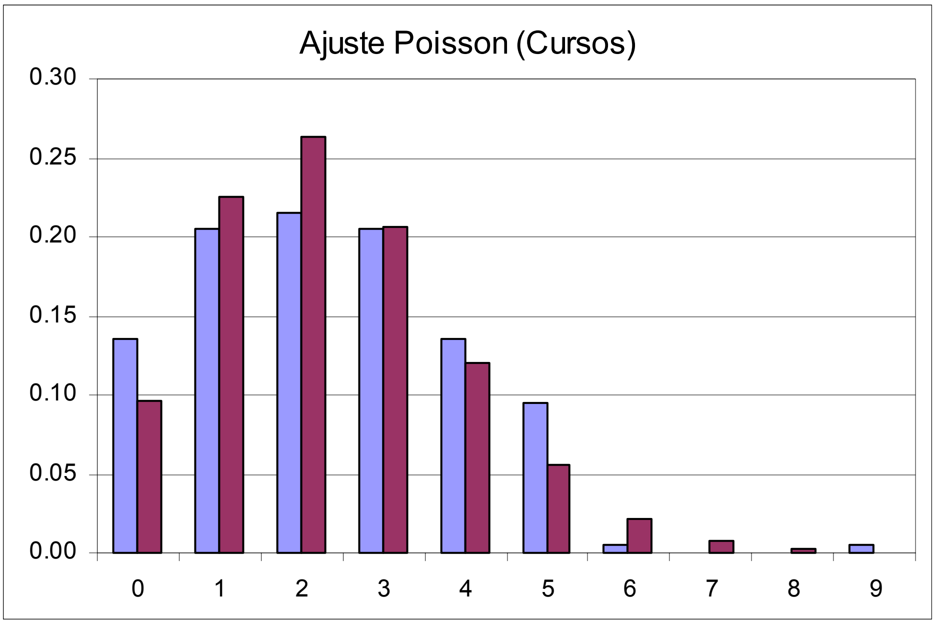
Ahora bien, es importante reconocer que los pronósticos que se obtienen a partir de modelos prácticamente nunca coinciden con los que se obtienen en forma directa de las frecuencias relativas. No tendría por que ocurrir de otro modo, estos pronósticos se fundan en supuestos diferentes y tienen distintos alcances.

Considere el ejemplo de la variable cursos que se discutió en la lección 9. Ahí, se tenía por una parte las frecuencias observadas que se muestran en la segunda columna de la siguiente tabla.

X	fr	Poisson
0	0.1350	0.0963
1	0.2050	0.2254
2	0.2150	0.2637
3	0.2050	0.2057
4	0.1350	0.1203
5	0.0950	0.0563
6	0.0050	0.0220
7	0.0000	0.0073
8	0.0000	0.0021
9	0.0050	0.0006
10		0.0001
11		0.0000
12		0.0000
13		0.0000
14		0.0000
15		0.0000

Con esa información, recurriendo únicamente a las frecuencias relativas, se podría pronosticar, por ejemplo, entre cero y cuatro cursos con una probabilidad de 0.895. Por supuesto, un pronóstico puntual podría ser $X = 2$ con una probabilidad de 0.215. También se puede afirmar que $X > 6$ sólo ocurrirá con probabilidad 0.005.

Si, alternativamente, se emplea un modelo Poisson (en este caso con $\lambda = 2.34$) se obtienen las probabilidades en la tercera columna de la tabla y se puede pronosticar que se observarán entre cero y cuatro cursos con una probabilidad de 0.9114. De la misma manera, un pronóstico puntual podría ser $X = 2$ pero ahora con una probabilidad de 0.2637. También en forma análoga, se tiene que la probabilidad de observar un valor de $X > 6$ es de 0.0102.



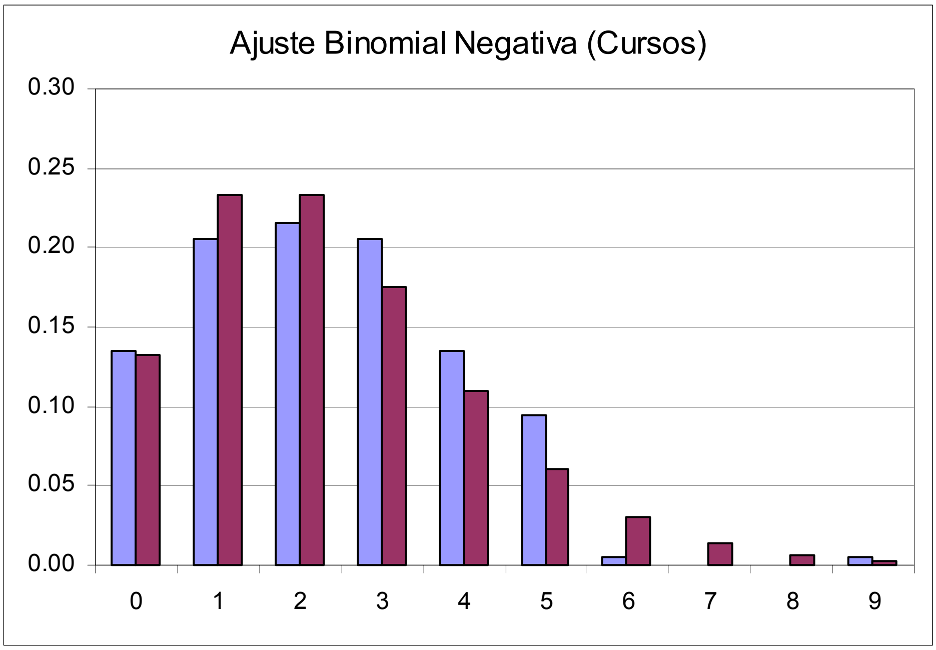
Como puede constatarse, los resultados son parecidos pero no iguales. En particular, con las frecuencias relativas no tiene sentido plantearse la ocurrencia de valores mayores que 9 mientras que con el modelo Poisson, los valores de 10 y más pueden producirse con probabilidad 0.00017.

Es importante notar que, en términos generales, los pronósticos del modelo Poisson son compatibles con los obtenidos directamente a partir de las frecuencias. La diferencia cualitativa más notable es que con el modelo algunos valores que no ocurrieron en el banco tienen una probabilidad positiva.

Por otra parte, si se adopta un modelo distinto (como el Binomial Negativo para los datos del ejemplo) también se pueden registrar cambios en los pronósticos.

X	fr	Bin. Neg.
0	0.1350	0.1328
1	0.2050	0.2329
2	0.2150	0.2334
3	0.2050	0.1754
4	0.1350	0.1099
5	0.0950	0.0606
6	0.0050	0.0303
7	0.0000	0.0141
8	0.0000	0.0062
9	0.0050	0.0026
10		0.0010
11		0.0004
12		0.0002
13		0.0001
14		0.0000
15		0.0000

Calcule los pronósticos con el modelo Binomial Negativo (con parámetros r = 7 y p = 0.75) equivalentes a los que se han presentado hasta aquí y compárelos.



En este punto vale la pena recordar que los modelos Poisson y Binomial Negativo, al igual que los demás que se han considerado en este texto, constituyen en realidad familias de modelos y que existen tantos modelos en cada familia como valores posibles puedan asignarse a los parámetros respectivos.

Anexo 1

Bases de Datos

Banco de Datos (n =19) con información de las variables Estatura y Peso

Caso	Estatura	Peso
1	153	49
2	165	73
3	165	61
4	165	56
5	166	60
6	168	60
7	168	72
8	168	50
9	170	66
10	170	70
11	170	75
12	171	71
13	171	70
14	174	75
15	175	62
16	178	62
17	183	100
18	186	85
19	195	98

Anexo 2

Tablas de Distribuciones

Tabla A de la Distribución Normal Estándar

$\Phi(x) = P(X \leq x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.51	0.9940
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.52	0.9941
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.53	0.9943
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.54	0.9945
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.55	0.9946
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.56	0.9948
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.57	0.9949
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.58	0.9951
0.19	0.5753	0.79	0.7852	1.39	0.9177	1.99	0.9767	2.59	0.9952
0.20	0.5793	0.80	0.7881	1.40	0.9192	2.00	0.9772	2.60	0.9953
0.21	0.5832	0.81	0.7910	1.41	0.9207	2.01	0.9778	2.61	0.9955
0.22	0.5871	0.82	0.7939	1.42	0.9222	2.02	0.9783	2.62	0.9956
0.23	0.5910	0.83	0.7967	1.43	0.9236	2.03	0.9788	2.63	0.9957
0.24	0.5948	0.84	0.7995	1.44	0.9251	2.04	0.9793	2.64	0.9959
0.25	0.5987	0.85	0.8023	1.45	0.9265	2.05	0.9798	2.65	0.9960
0.26	0.6026	0.86	0.8051	1.46	0.9279	2.06	0.9803	2.66	0.9961
0.27	0.6064	0.87	0.8078	1.47	0.9292	2.07	0.9808	2.67	0.9962
0.28	0.6103	0.88	0.8106	1.48	0.9306	2.08	0.9812	2.68	0.9963
0.29	0.6141	0.89	0.8133	1.49	0.9319	2.09	0.9817	2.69	0.9964
0.30	0.6179	0.90	0.8159	1.50	0.9332	2.10	0.9821	2.70	0.9965
0.31	0.6217	0.91	0.8186	1.51	0.9345	2.11	0.9826	2.71	0.9966
0.32	0.6255	0.92	0.8212	1.52	0.9357	2.12	0.9830	2.72	0.9967
0.33	0.6293	0.93	0.8238	1.53	0.9370	2.13	0.9834	2.73	0.9968
0.34	0.6331	0.94	0.8264	1.54	0.9382	2.14	0.9838	2.74	0.9969
0.35	0.6368	0.95	0.8289	1.55	0.9394	2.15	0.9842	2.75	0.9970
0.36	0.6406	0.96	0.8315	1.56	0.9406	2.16	0.9846	2.76	0.9971
0.37	0.6443	0.97	0.8340	1.57	0.9418	2.17	0.9850	2.77	0.9972
0.38	0.6480	0.98	0.8365	1.58	0.9429	2.18	0.9854	2.78	0.9973
0.39	0.6517	0.99	0.8389	1.59	0.9441	2.19	0.9857	2.79	0.9974
0.40	0.6554	1.00	0.8413	1.60	0.9452	2.20	0.9861	2.80	0.9974
0.41	0.6591	1.01	0.8438	1.61	0.9463	2.21	0.9864	2.81	0.9975
0.42	0.6628	1.02	0.8461	1.62	0.9474	2.22	0.9868	2.82	0.9976
0.43	0.6664	1.03	0.8485	1.63	0.9484	2.23	0.9871	2.83	0.9977
0.44	0.6700	1.04	0.8508	1.64	0.9495	2.24	0.9875	2.84	0.9977
0.45	0.6736	1.05	0.8531	1.65	0.9505	2.25	0.9878	2.85	0.9978
0.46	0.6772	1.06	0.8554	1.66	0.9515	2.26	0.9881	2.86	0.9979
0.47	0.6808	1.07	0.8577	1.67	0.9525	2.27	0.9884	2.87	0.9979
0.48	0.6844	1.08	0.8599	1.68	0.9535	2.28	0.9887	2.88	0.9980
0.49	0.6879	1.09	0.8621	1.69	0.9545	2.29	0.9890	2.89	0.9981
0.50	0.6915	1.10	0.8643	1.70	0.9554	2.30	0.9893	2.90	0.9981
0.51	0.6950	1.11	0.8665	1.71	0.9564	2.31	0.9896	2.91	0.9982
0.52	0.6985	1.12	0.8686	1.72	0.9573	2.32	0.9898	2.92	0.9982
0.53	0.7019	1.13	0.8708	1.73	0.9582	2.33	0.9901	2.93	0.9983
0.54	0.7054	1.14	0.8729	1.74	0.9591	2.34	0.9904	2.94	0.9984
0.55	0.7088	1.15	0.8749	1.75	0.9599	2.35	0.9906	2.95	0.9984
0.56	0.7123	1.16	0.8770	1.76	0.9608	2.36	0.9909	2.96	0.9985
0.57	0.7157	1.17	0.8790	1.77	0.9616	2.37	0.9911	2.97	0.9985
0.58	0.7190	1.18	0.8810	1.78	0.9625	2.38	0.9913	2.98	0.9986
0.59	0.7224	1.19	0.8830	1.79	0.9633	2.39	0.9916	2.99	0.9986

Tabla B de la Distribución Normal Estándar

$\Phi(x) = P(X \leq x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	1.00	0.8413	2.00	0.9772	3.00	0.9987
0.02	0.5080	1.02	0.8461	2.02	0.9783	3.02	0.9987
0.04	0.5160	1.04	0.8508	2.04	0.9793	3.04	0.9988
0.06	0.5239	1.06	0.8554	2.06	0.9803	3.06	0.9989
0.08	0.5319	1.08	0.8599	2.08	0.9812	3.08	0.9990
0.10	0.5398	1.10	0.8643	2.10	0.9821	3.10	0.9990
0.12	0.5478	1.12	0.8686	2.12	0.9830	3.12	0.9991
0.14	0.5557	1.14	0.8729	2.14	0.9838	3.14	0.9992
0.16	0.5636	1.16	0.8770	2.16	0.9846	3.16	0.9992
0.18	0.5714	1.18	0.8810	2.18	0.9854	3.18	0.9993
0.20	0.5793	1.20	0.8849	2.20	0.9861	3.20	0.9993
0.22	0.5871	1.22	0.8888	2.22	0.9868	3.22	0.9994
0.24	0.5948	1.24	0.8925	2.24	0.9875	3.24	0.9994
0.26	0.6026	1.26	0.8962	2.26	0.9881	3.26	0.9994
0.28	0.6103	1.28	0.8997	2.28	0.9887	3.28	0.9995
0.30	0.6179	1.30	0.9032	2.30	0.9893	3.30	0.9995
0.32	0.6255	1.32	0.9066	2.32	0.9898	3.32	0.9995
0.34	0.6331	1.34	0.9099	2.34	0.9904	3.34	0.9996
0.36	0.6406	1.36	0.9131	2.36	0.9909	3.36	0.9996
0.38	0.6480	1.38	0.9162	2.38	0.9913	3.38	0.9996
0.40	0.6554	1.40	0.9192	2.40	0.9918	3.40	0.9997
0.42	0.6628	1.42	0.9222	2.42	0.9922	3.42	0.9997
0.44	0.6700	1.44	0.9251	2.44	0.9927	3.44	0.9997
0.46	0.6772	1.46	0.9279	2.46	0.9931	3.46	0.9997
0.48	0.6844	1.48	0.9306	2.48	0.9934	3.48	0.9997
0.50	0.6915	1.50	0.9332	2.50	0.9938	3.50	0.9998
0.52	0.6985	1.52	0.9357	2.52	0.9941	3.52	0.9998
0.54	0.7054	1.54	0.9382	2.54	0.9945	3.54	0.9998
0.56	0.7123	1.56	0.9406	2.56	0.9948	3.56	0.9998
0.58	0.7190	1.58	0.9429	2.58	0.9951	3.58	0.9998
0.60	0.7257	1.60	0.9452	2.60	0.9953	3.60	0.9998
0.62	0.7324	1.62	0.9474	2.62	0.9956	3.62	0.9999
0.64	0.7389	1.64	0.9495	2.64	0.9959	3.64	0.9999
0.66	0.7454	1.66	0.9515	2.66	0.9961	3.66	0.9999
0.68	0.7517	1.68	0.9535	2.68	0.9963	3.68	0.9999
0.70	0.7580	1.70	0.9554	2.70	0.9965	3.70	0.9999
0.72	0.7642	1.72	0.9573	2.72	0.9967	3.72	0.9999
0.74	0.7704	1.74	0.9591	2.74	0.9969	3.74	0.9999
0.76	0.7764	1.76	0.9608	2.76	0.9971	3.76	0.9999
0.78	0.7823	1.78	0.9625	2.78	0.9973	3.78	0.9999
0.80	0.7881	1.80	0.9641	2.80	0.9974	3.80	0.9999
0.82	0.7939	1.82	0.9656	2.82	0.9976	3.82	0.9999
0.84	0.7995	1.84	0.9671	2.84	0.9977	3.84	0.9999
0.86	0.8051	1.86	0.9686	2.86	0.9979	3.86	0.9999
0.88	0.8106	1.88	0.9699	2.88	0.9980	3.88	0.9999
0.90	0.8159	1.90	0.9713	2.90	0.9981	3.90	1.0000
0.92	0.8212	1.92	0.9726	2.92	0.9982	3.92	1.0000
0.94	0.8264	1.94	0.9738	2.94	0.9984	3.94	1.0000
0.96	0.8315	1.96	0.9750	2.96	0.9985	3.96	1.0000
0.98	0.8365	1.98	0.9761	2.98	0.9986	3.98	1.0000

Cuantiles de la Distribución Ji Cuadrada

G.L.	0.700	0.800	0.900	0.950	0.975	0.990	0.999
1	1.07	1.64	2.71	3.84	5.02	6.63	10.83
2	2.41	3.22	4.61	5.99	7.38	9.21	13.82
3	3.66	4.64	6.25	7.81	9.35	11.34	16.27
4	4.88	5.99	7.78	9.49	11.14	13.28	18.47
5	6.06	7.29	9.24	11.07	12.83	15.09	20.51
6	7.23	8.56	10.64	12.59	14.45	16.81	22.46
7	8.38	9.80	12.02	14.07	16.01	18.48	24.32
8	9.52	11.03	13.36	15.51	17.53	20.09	26.12
9	10.66	12.24	14.68	16.92	19.02	21.67	27.88
10	11.78	13.44	15.99	18.31	20.48	23.21	29.59
11	12.90	14.63	17.28	19.68	21.92	24.73	31.26
12	14.01	15.81	18.55	21.03	23.34	26.22	32.91
13	15.12	16.98	19.81	22.36	24.74	27.69	34.53
14	16.22	18.15	21.06	23.68	26.12	29.14	36.12
15	17.32	19.31	22.31	25.00	27.49	30.58	37.70
16	18.42	20.47	23.54	26.30	28.85	32.00	39.25
17	19.51	21.61	24.77	27.59	30.19	33.41	40.79
18	20.60	22.76	25.99	28.87	31.53	34.81	42.31
19	21.69	23.90	27.20	30.14	32.85	36.19	43.82
20	22.77	25.04	28.41	31.41	34.17	37.57	45.31
21	23.86	26.17	29.62	32.67	35.48	38.93	46.80
22	24.94	27.30	30.81	33.92	36.78	40.29	48.27
23	26.02	28.43	32.01	35.17	38.08	41.64	49.73
24	27.10	29.55	33.20	36.42	39.36	42.98	51.18
25	28.17	30.68	34.38	37.65	40.65	44.31	52.62
26	29.25	31.79	35.56	38.89	41.92	45.64	54.05
27	30.32	32.91	36.74	40.11	43.19	46.96	55.48
28	31.39	34.03	37.92	41.34	44.46	48.28	56.89
29	32.46	35.14	39.09	42.56	45.72	49.59	58.30
30	33.53	36.25	40.26	43.77	46.98	50.89	59.70
31	34.60	37.36	41.42	44.99	48.23	52.19	61.10
32	35.66	38.47	42.58	46.19	49.48	53.49	62.49
33	36.73	39.57	43.75	47.40	50.73	54.78	63.87
34	37.80	40.68	44.90	48.60	51.97	56.06	65.25
35	38.86	41.78	46.06	49.80	53.20	57.34	66.62
36	39.92	42.88	47.21	51.00	54.44	58.62	67.98
37	40.98	43.98	48.36	52.19	55.67	59.89	69.35
38	42.05	45.08	49.51	53.38	56.90	61.16	70.70
39	43.11	46.17	50.66	54.57	58.12	62.43	72.05
40	44.16	47.27	51.81	55.76	59.34	63.69	73.40
50	54.72	58.16	63.17	67.50	71.42	76.15	86.66
60	65.23	68.97	74.40	79.08	83.30	88.38	99.61
70	75.69	79.71	85.53	90.53	95.02	100.43	112.32
80	86.12	90.41	96.58	101.88	106.63	112.33	124.84
90	96.52	101.05	107.56	113.15	118.14	124.12	137.21
100	106.91	111.67	118.50	124.34	129.56	135.81	149.45

Cuantiles de la Distribución t de Student

G.L.	0.700	0.800	0.900	0.950	0.975	0.990	0.999
1	0.727	1.376	3.078	6.314	12.706	31.821	318.289
2	0.617	1.061	1.886	2.920	4.303	6.965	22.328
3	0.584	0.978	1.638	2.353	3.182	4.541	10.214
4	0.569	0.941	1.533	2.132	2.776	3.747	7.173
5	0.559	0.920	1.476	2.015	2.571	3.365	5.894
6	0.553	0.906	1.440	1.943	2.447	3.143	5.208
7	0.549	0.896	1.415	1.895	2.365	2.998	4.785
8	0.546	0.889	1.397	1.860	2.306	2.896	4.501
9	0.543	0.883	1.383	1.833	2.262	2.821	4.297
10	0.542	0.879	1.372	1.812	2.228	2.764	4.144
11	0.540	0.876	1.363	1.796	2.201	2.718	4.025
12	0.539	0.873	1.356	1.782	2.179	2.681	3.930
13	0.538	0.870	1.350	1.771	2.160	2.650	3.852
14	0.537	0.868	1.345	1.761	2.145	2.624	3.787
15	0.536	0.866	1.341	1.753	2.131	2.602	3.733
16	0.535	0.865	1.337	1.746	2.120	2.583	3.686
17	0.534	0.863	1.333	1.740	2.110	2.567	3.646
18	0.534	0.862	1.330	1.734	2.101	2.552	3.610
19	0.533	0.861	1.328	1.729	2.093	2.539	3.579
20	0.533	0.860	1.325	1.725	2.086	2.528	3.552
21	0.532	0.859	1.323	1.721	2.080	2.518	3.527
22	0.532	0.858	1.321	1.717	2.074	2.508	3.505
23	0.532	0.858	1.319	1.714	2.069	2.500	3.485
24	0.531	0.857	1.318	1.711	2.064	2.492	3.467
25	0.531	0.856	1.316	1.708	2.060	2.485	3.450
26	0.531	0.856	1.315	1.706	2.056	2.479	3.435
27	0.531	0.855	1.314	1.703	2.052	2.473	3.421
28	0.530	0.855	1.313	1.701	2.048	2.467	3.408
29	0.530	0.854	1.311	1.699	2.045	2.462	3.396
30	0.530	0.854	1.310	1.697	2.042	2.457	3.385
31	0.530	0.853	1.309	1.696	2.040	2.453	3.375
32	0.530	0.853	1.309	1.694	2.037	2.449	3.365
33	0.530	0.853	1.308	1.692	2.035	2.445	3.356
34	0.529	0.852	1.307	1.691	2.032	2.441	3.348
35	0.529	0.852	1.306	1.690	2.030	2.438	3.340
36	0.529	0.852	1.306	1.688	2.028	2.434	3.333
37	0.529	0.851	1.305	1.687	2.026	2.431	3.326
38	0.529	0.851	1.304	1.686	2.024	2.429	3.319
39	0.529	0.851	1.304	1.685	2.023	2.426	3.313
40	0.529	0.851	1.303	1.684	2.021	2.423	3.307
50	0.528	0.849	1.299	1.676	2.009	2.403	3.261
60	0.527	0.848	1.296	1.671	2.000	2.390	3.232
70	0.527	0.847	1.294	1.667	1.994	2.381	3.211
80	0.526	0.846	1.292	1.664	1.990	2.374	3.195
90	0.526	0.846	1.291	1.662	1.987	2.368	3.183
100	0.526	0.845	1.290	1.660	1.984	2.364	3.174

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

b \ **a**	1	2	3	4	5	6	7	8	9	10
1	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857	60.195
	161.446	199.499	215.707	224.583	230.160	233.988	236.767	238.884	240.543	241.882
	4052.185	4999.340	5403.534	5624.257	5763.955	5858.950	5928.334	5980.954	6022.397	6055.925
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392
	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396
	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230
	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785
	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920
	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297
	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937
	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703
	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538
	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347
	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416
	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323
	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248
	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188
	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138
	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095
	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059
	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

b	a: 12	14	16	18	20	22	24	26	28	30
1	60.705	61.073	61.350	61.566	61.740	61.883	62.002	62.103	62.189	62.265
	243.905	245.363	246.466	247.324	248.016	248.579	249.052	249.453	249.798	250.096
	6106.682	6143.004	6170.012	6191.432	6208.662	6223.097	6234.273	6244.518	6252.900	6260.350
2	9.408	9.420	9.429	9.436	9.441	9.446	9.450	9.453	9.456	9.458
	19.412	19.424	19.433	19.440	19.446	19.450	19.454	19.457	19.460	19.463
	99.419	99.426	99.437	99.444	99.448	99.455	99.455	99.462	99.462	99.466
3	5.216	5.205	5.196	5.190	5.184	5.180	5.176	5.173	5.170	5.168
	8.745	8.715	8.692	8.675	8.660	8.648	8.638	8.630	8.623	8.617
	27.052	26.924	26.826	26.751	26.690	26.639	26.597	26.562	26.531	26.504
4	3.896	3.878	3.864	3.853	3.844	3.837	3.831	3.826	3.821	3.817
	5.912	5.873	5.844	5.821	5.803	5.787	5.774	5.763	5.754	5.746
	14.374	14.249	14.154	14.079	14.019	13.970	13.929	13.894	13.864	13.838
5	3.268	3.247	3.230	3.217	3.207	3.198	3.191	3.184	3.179	3.174
	4.678	4.636	4.604	4.579	4.558	4.541	4.527	4.515	4.505	4.496
	9.888	9.770	9.680	9.609	9.553	9.506	9.466	9.433	9.404	9.379
6	2.905	2.881	2.863	2.848	2.836	2.827	2.818	2.811	2.805	2.800
	4.000	3.956	3.922	3.896	3.874	3.856	3.841	3.829	3.818	3.808
	7.718	7.605	7.519	7.451	7.396	7.351	7.313	7.281	7.253	7.229
7	2.668	2.643	2.623	2.607	2.595	2.584	2.575	2.568	2.561	2.555
	3.575	3.529	3.494	3.467	3.445	3.426	3.410	3.397	3.386	3.376
	6.469	6.359	6.275	6.209	6.155	6.111	6.074	6.043	6.016	5.992
8	2.502	2.475	2.454	2.438	2.425	2.414	2.404	2.396	2.389	2.383
	3.284	3.237	3.202	3.173	3.150	3.131	3.115	3.102	3.090	3.079
	5.667	5.559	5.477	5.412	5.359	5.316	5.279	5.248	5.221	5.198
9	2.379	2.351	2.330	2.312	2.298	2.287	2.277	2.268	2.261	2.255
	3.073	3.025	2.989	2.960	2.936	2.917	2.900	2.886	2.874	2.864
	5.111	5.005	4.924	4.860	4.808	4.765	4.729	4.698	4.672	4.649
10	2.284	2.255	2.233	2.215	2.201	2.189	2.178	2.170	2.162	2.155
	2.913	2.865	2.828	2.798	2.774	2.754	2.737	2.723	2.710	2.700
	4.706	4.601	4.520	4.457	4.405	4.363	4.327	4.296	4.270	4.247
11	2.209	2.179	2.156	2.138	2.123	2.111	2.100	2.091	2.083	2.076
	2.788	2.739	2.701	2.671	2.646	2.626	2.609	2.594	2.582	2.570
	4.397	4.293	4.213	4.150	4.099	4.057	4.021	3.990	3.964	3.941
12	2.147	2.117	2.094	2.075	2.060	2.047	2.036	2.027	2.019	2.011
	2.687	2.637	2.599	2.568	2.544	2.523	2.505	2.491	2.478	2.466
	4.155	4.052	3.972	3.910	3.858	3.816	3.780	3.750	3.724	3.701
13	2.097	2.066	2.042	2.023	2.007	1.994	1.983	1.973	1.965	1.958
	2.604	2.554	2.515	2.484	2.459	2.438	2.420	2.405	2.392	2.380
	3.960	3.857	3.778	3.716	3.665	3.622	3.587	3.556	3.530	3.507
14	2.054	2.022	1.998	1.978	1.962	1.949	1.938	1.928	1.919	1.912
	2.534	2.484	2.445	2.413	2.388	2.367	2.349	2.333	2.320	2.308
	3.800	3.698	3.619	3.556	3.505	3.463	3.427	3.397	3.371	3.348
15	2.017	1.985	1.961	1.941	1.924	1.911	1.899	1.889	1.880	1.873
	2.475	2.424	2.385	2.353	2.328	2.306	2.288	2.272	2.259	2.247
	3.666	3.564	3.485	3.423	3.372	3.330	3.294	3.264	3.237	3.214

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

a

b	35	40	45	50	55	60	65	70	75	80
1	62.416	62.529	62.617	62.688	62.746	62.794	62.835	62.871	62.901	62.927
	250.693	251.144	251.493	251.774	252.003	252.196	252.359	252.498	252.618	252.723
	6275.252	6286.427	6295.741	6302.260	6308.313	6312.970	6317.161	6320.886	6323.680	6326.474
2	9.463	9.466	9.469	9.471	9.473	9.475	9.476	9.477	9.478	9.479
	19.467	19.471	19.473	19.476	19.478	19.479	19.480	19.481	19.482	19.483
	99.470	99.477	99.477	99.477	99.481	99.484	99.484	99.484	99.484	99.484
3	5.163	5.160	5.157	5.155	5.153	5.151	5.150	5.149	5.148	5.147
	8.604	8.594	8.587	8.581	8.576	8.572	8.569	8.566	8.563	8.561
	26.451	26.411	26.379	26.354	26.334	26.316	26.302	26.289	26.278	26.269
4	3.810	3.804	3.799	3.795	3.792	3.790	3.787	3.786	3.784	3.782
	5.729	5.717	5.707	5.699	5.693	5.688	5.683	5.679	5.676	5.673
	13.785	13.745	13.714	13.690	13.669	13.652	13.638	13.626	13.615	13.605
5	3.165	3.157	3.152	3.147	3.143	3.140	3.138	3.135	3.133	3.132
	4.478	4.464	4.453	4.444	4.437	4.431	4.426	4.422	4.418	4.415
	9.329	9.291	9.262	9.238	9.218	9.202	9.188	9.176	9.166	9.157
6	2.789	2.781	2.775	2.770	2.765	2.762	2.759	2.756	2.754	2.752
	3.789	3.774	3.763	3.754	3.746	3.740	3.734	3.730	3.726	3.722
	7.180	7.143	7.115	7.091	7.073	7.057	7.043	7.032	7.022	7.013
7	2.544	2.535	2.528	2.523	2.518	2.514	2.511	2.508	2.506	2.504
	3.356	3.340	3.328	3.319	3.311	3.304	3.299	3.294	3.290	3.286
	5.944	5.908	5.880	5.858	5.839	5.824	5.810	5.799	5.789	5.781
8	2.371	2.361	2.354	2.348	2.343	2.339	2.336	2.333	2.330	2.328
	3.059	3.043	3.030	3.020	3.012	3.005	2.999	2.994	2.990	2.986
	5.151	5.116	5.088	5.065	5.047	5.032	5.019	5.007	4.998	4.989
9	2.242	2.232	2.224	2.218	2.213	2.208	2.205	2.202	2.199	2.196
	2.842	2.826	2.813	2.803	2.794	2.787	2.781	2.776	2.771	2.768
	4.602	4.567	4.539	4.517	4.498	4.483	4.470	4.459	4.449	4.441
10	2.142	2.132	2.124	2.117	2.112	2.107	2.103	2.100	2.097	2.095
	2.678	2.661	2.648	2.637	2.628	2.621	2.615	2.609	2.605	2.601
	4.201	4.165	4.138	4.115	4.097	4.082	4.069	4.058	4.048	4.039
11	2.062	2.052	2.043	2.036	2.031	2.026	2.022	2.019	2.016	2.013
	2.548	2.531	2.517	2.507	2.498	2.490	2.484	2.478	2.473	2.469
	3.895	3.860	3.832	3.810	3.791	3.776	3.763	3.752	3.742	3.734
12	1.997	1.986	1.977	1.970	1.965	1.960	1.956	1.952	1.949	1.946
	2.443	2.426	2.412	2.401	2.392	2.384	2.378	2.372	2.367	2.363
	3.654	3.619	3.592	3.569	3.551	3.535	3.522	3.511	3.501	3.493
13	1.943	1.931	1.923	1.915	1.909	1.904	1.900	1.896	1.893	1.890
	2.357	2.339	2.325	2.314	2.304	2.297	2.290	2.284	2.279	2.275
	3.461	3.425	3.398	3.375	3.357	3.341	3.328	3.317	3.307	3.298
14	1.897	1.885	1.876	1.869	1.862	1.857	1.853	1.849	1.846	1.843
	2.284	2.266	2.252	2.241	2.231	2.223	2.216	2.210	2.205	2.201
	3.301	3.266	3.238	3.215	3.197	3.181	3.168	3.157	3.147	3.138
15	1.857	1.845	1.836	1.828	1.822	1.817	1.812	1.808	1.805	1.802
	2.223	2.204	2.190	2.178	2.168	2.160	2.153	2.147	2.142	2.137
	3.167	3.132	3.104	3.081	3.063	3.047	3.034	3.022	3.012	3.004

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

	a									
b	1	2	3	4	5	6	7	8	9	10
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028
	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977
	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937
	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866
	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819
	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979
35	2.855	2.461	2.247	2.113	2.019	1.950	1.896	1.852	1.817	1.787
	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114
	7.419	5.268	4.396	3.908	3.592	3.368	3.200	3.069	2.963	2.876
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763
	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801
45	2.820	2.425	2.210	2.074	1.980	1.909	1.855	1.811	1.774	1.744
	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152	2.096	2.049
	7.234	5.110	4.249	3.767	3.454	3.232	3.066	2.935	2.830	2.743
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729
	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698
55	2.799	2.402	2.186	2.050	1.955	1.884	1.829	1.785	1.748	1.717
	4.016	3.165	2.773	2.540	2.383	2.269	2.181	2.112	2.055	2.008
	7.119	5.013	4.159	3.681	3.370	3.149	2.983	2.853	2.748	2.662
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707
	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632
65	2.784	2.386	2.170	2.033	1.938	1.867	1.811	1.767	1.730	1.699
	3.989	3.138	2.746	2.513	2.356	2.242	2.154	2.084	2.027	1.980
	7.042	4.947	4.098	3.622	3.313	3.093	2.928	2.798	2.693	2.607
70	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691
	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969
	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585
75	2.774	2.375	2.158	2.021	1.926	1.854	1.798	1.754	1.716	1.685
	3.968	3.119	2.727	2.494	2.337	2.222	2.134	2.064	2.007	1.959
	6.985	4.900	4.054	3.580	3.272	3.052	2.887	2.758	2.653	2.567
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680
	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951
	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

b \ a	12	14	16	18	20	22	24	26	28	30
16	1.985	1.953	1.928	1.908	1.891	1.877	1.866	1.855	1.847	1.839
	2.425	2.373	2.333	2.302	2.276	2.254	2.235	2.220	2.206	2.194
	3.553	3.451	3.372	3.310	3.259	3.216	3.181	3.150	3.124	3.101
18	1.933	1.900	1.875	1.854	1.837	1.823	1.810	1.800	1.791	1.783
	2.342	2.290	2.250	2.217	2.191	2.168	2.150	2.134	2.119	2.107
	3.371	3.269	3.190	3.128	3.077	3.035	2.999	2.968	2.942	2.919
20	1.892	1.859	1.833	1.811	1.794	1.779	1.767	1.756	1.746	1.738
	2.278	2.225	2.184	2.151	2.124	2.102	2.082	2.066	2.052	2.039
	3.231	3.130	3.051	2.989	2.938	2.895	2.859	2.829	2.802	2.778
25	1.820	1.785	1.758	1.736	1.718	1.702	1.689	1.678	1.668	1.659
	2.165	2.111	2.069	2.035	2.007	1.984	1.964	1.947	1.932	1.919
	2.993	2.892	2.813	2.751	2.699	2.657	2.620	2.589	2.562	2.538
30	1.773	1.737	1.709	1.686	1.667	1.651	1.638	1.626	1.616	1.606
	2.092	2.037	1.995	1.960	1.932	1.908	1.887	1.870	1.854	1.841
	2.843	2.742	2.663	2.600	2.549	2.506	2.469	2.437	2.410	2.386
35	1.739	1.703	1.674	1.651	1.632	1.615	1.601	1.589	1.579	1.569
	2.041	1.986	1.942	1.907	1.878	1.854	1.833	1.815	1.799	1.786
	2.740	2.639	2.560	2.497	2.445	2.401	2.364	2.333	2.305	2.281
40	1.715	1.678	1.649	1.625	1.605	1.588	1.574	1.562	1.551	1.541
	2.003	1.948	1.904	1.868	1.839	1.814	1.793	1.775	1.759	1.744
	2.665	2.563	2.484	2.421	2.369	2.325	2.288	2.256	2.228	2.203
45	1.695	1.658	1.629	1.605	1.585	1.568	1.553	1.540	1.529	1.519
	1.974	1.918	1.874	1.838	1.808	1.783	1.762	1.743	1.727	1.713
	2.608	2.506	2.427	2.363	2.311	2.267	2.230	2.197	2.169	2.144
50	1.680	1.643	1.613	1.588	1.568	1.551	1.536	1.523	1.512	1.502
	1.952	1.895	1.850	1.814	1.784	1.759	1.737	1.718	1.702	1.687
	2.563	2.461	2.382	2.318	2.265	2.221	2.183	2.151	2.123	2.098
55	1.668	1.630	1.600	1.575	1.555	1.537	1.522	1.509	1.498	1.487
	1.933	1.876	1.831	1.795	1.764	1.739	1.717	1.698	1.681	1.666
	2.526	2.424	2.345	2.281	2.228	2.184	2.146	2.113	2.085	2.060
60	1.657	1.619	1.589	1.564	1.543	1.526	1.511	1.498	1.486	1.476
	1.917	1.860	1.815	1.778	1.748	1.722	1.700	1.681	1.664	1.649
	2.496	2.394	2.315	2.251	2.198	2.153	2.115	2.083	2.054	2.028
65	1.649	1.610	1.580	1.555	1.534	1.516	1.501	1.488	1.476	1.465
	1.904	1.847	1.802	1.765	1.734	1.708	1.686	1.667	1.650	1.635
	2.471	2.369	2.289	2.225	2.172	2.128	2.089	2.056	2.028	2.002
70	1.641	1.603	1.572	1.547	1.526	1.508	1.493	1.479	1.467	1.457
	1.893	1.836	1.790	1.753	1.722	1.696	1.674	1.654	1.637	1.622
	2.450	2.348	2.268	2.204	2.150	2.106	2.067	2.034	2.005	1.980
75	1.635	1.596	1.565	1.540	1.519	1.501	1.485	1.472	1.460	1.449
	1.884	1.826	1.780	1.743	1.712	1.686	1.663	1.644	1.626	1.611
	2.431	2.329	2.249	2.185	2.132	2.087	2.048	2.015	1.986	1.960
80	1.629	1.590	1.559	1.534	1.513	1.495	1.479	1.465	1.453	1.443
	1.875	1.817	1.772	1.734	1.703	1.677	1.654	1.634	1.617	1.602
	2.415	2.313	2.233	2.169	2.115	2.070	2.032	1.999	1.969	1.944

Cuantiles (0.90, 0.95 y 0.99) de la distribución F

Grados de libertad = a (numerador) y b (denominador)

a

b	35	40	45	50	55	60	65	70	75	80
16	1.823	1.811	1.801	1.793	1.787	1.782	1.777	1.773	1.769	1.766
	2.169	2.151	2.136	2.124	2.114	2.106	2.099	2.093	2.087	2.083
	3.054	3.018	2.990	2.967	2.949	2.933	2.920	2.908	2.898	2.889
18	1.766	1.754	1.744	1.736	1.729	1.723	1.718	1.714	1.711	1.707
	2.082	2.063	2.048	2.035	2.025	2.017	2.009	2.003	1.998	1.993
	2.871	2.835	2.807	2.784	2.765	2.749	2.736	2.724	2.714	2.705
20	1.721	1.708	1.698	1.690	1.683	1.677	1.672	1.667	1.664	1.660
	2.013	1.994	1.978	1.966	1.955	1.946	1.939	1.932	1.927	1.922
	2.731	2.695	2.666	2.643	2.624	2.608	2.594	2.582	2.572	2.563
25	1.641	1.627	1.616	1.607	1.600	1.593	1.588	1.583	1.579	1.576
	1.892	1.872	1.855	1.842	1.831	1.822	1.814	1.807	1.801	1.796
	2.490	2.453	2.424	2.400	2.380	2.364	2.350	2.337	2.327	2.317
30	1.588	1.573	1.562	1.552	1.544	1.538	1.532	1.527	1.523	1.519
	1.813	1.792	1.775	1.761	1.749	1.740	1.731	1.724	1.718	1.712
	2.337	2.299	2.269	2.245	2.225	2.208	2.193	2.181	2.170	2.160
35	1.550	1.535	1.523	1.513	1.504	1.497	1.491	1.486	1.482	1.478
	1.757	1.735	1.718	1.703	1.691	1.681	1.672	1.665	1.658	1.652
	2.231	2.193	2.162	2.137	2.117	2.099	2.085	2.072	2.060	2.050
40	1.521	1.506	1.493	1.483	1.474	1.467	1.461	1.455	1.451	1.447
	1.715	1.693	1.675	1.660	1.648	1.637	1.628	1.621	1.614	1.608
	2.153	2.114	2.083	2.058	2.037	2.019	2.004	1.991	1.980	1.969
45	1.499	1.483	1.470	1.460	1.451	1.443	1.437	1.431	1.426	1.422
	1.683	1.660	1.642	1.626	1.614	1.603	1.594	1.586	1.579	1.573
	2.093	2.054	2.023	1.997	1.976	1.958	1.942	1.929	1.917	1.907
50	1.481	1.465	1.452	1.441	1.432	1.424	1.418	1.412	1.407	1.402
	1.657	1.634	1.615	1.599	1.587	1.576	1.566	1.558	1.551	1.544
	2.046	2.007	1.975	1.949	1.927	1.909	1.893	1.880	1.868	1.857
55	1.466	1.450	1.437	1.426	1.416	1.408	1.402	1.396	1.391	1.386
	1.636	1.612	1.593	1.577	1.564	1.553	1.544	1.535	1.528	1.521
	2.008	1.968	1.936	1.910	1.888	1.869	1.853	1.839	1.827	1.817
60	1.454	1.437	1.424	1.413	1.403	1.395	1.388	1.382	1.377	1.372
	1.618	1.594	1.575	1.559	1.546	1.534	1.525	1.516	1.508	1.502
	1.976	1.936	1.904	1.877	1.855	1.836	1.820	1.806	1.794	1.783
65	1.444	1.427	1.413	1.402	1.392	1.384	1.377	1.371	1.365	1.360
	1.603	1.579	1.560	1.543	1.530	1.518	1.508	1.500	1.492	1.485
	1.950	1.909	1.877	1.850	1.827	1.808	1.792	1.778	1.765	1.754
70	1.435	1.418	1.404	1.392	1.382	1.374	1.367	1.361	1.355	1.350
	1.591	1.566	1.546	1.530	1.516	1.505	1.494	1.486	1.478	1.471
	1.927	1.886	1.853	1.826	1.804	1.785	1.768	1.754	1.741	1.730
75	1.427	1.410	1.396	1.384	1.374	1.366	1.358	1.352	1.346	1.341
	1.580	1.555	1.535	1.518	1.504	1.493	1.482	1.473	1.466	1.459
	1.907	1.866	1.833	1.806	1.783	1.764	1.747	1.733	1.720	1.709
80	1.420	1.403	1.388	1.377	1.367	1.358	1.351	1.344	1.339	1.334
	1.570	1.545	1.525	1.508	1.494	1.482	1.472	1.463	1.455	1.448
	1.890	1.849	1.816	1.788	1.765	1.746	1.729	1.714	1.702	1.690