Introducción a Modelos Psicométricos

Clases 2 y 3: Herramientas estadísticas para la psicometría

Iwin Leenen y Ramsés Vázquez-Lira

Facultad de Psicología, UNAM

Programa de Licenciatura y Posgrado en Psicología
Semestre 2019–1

Índice

Índice

Índice

Objetos, variables y observaciones

En la estadística aplicada, se dispone de los datos de varios objetos o unidades experimentales.

Por ejemplo:

- Hemos registrado la interacción social de 50 niños en el patio de la escuela.
- Tenemos las respuestas en un cuestionario de 60 participantes en un taller de *mindfulness*.
- Tenemos los resultados de 1,000 alumnos en el examen de entrada de la UNAM.

Para hacer referencia a las unidades experimentales, se suele asignarles un número. Por ejemplo, se habla del resultado del “alumno *i*” en el examen.

Objetos, variables y observaciones

En la estadística aplicada, se dispone de los datos de varios objetos o unidades experimentales.

Por ejemplo:

- Hemos registrado la interacción social de 50 niños en el patio de la escuela.
- Tenemos las respuestas en un cuestionario de 60 participantes en un taller de *mindfulness*.
- Tenemos los resultados de 1,000 alumnos en el examen de entrada de la UNAM.

Para hacer referencia a las unidades experimentales, se suele asignarles un número. Por ejemplo, se habla del resultado del “alumno i” en el examen.

Objetos, variables y observaciones

En la estadística aplicada, se dispone de los datos de varios objetos o unidades experimentales.

Por ejemplo:

- Hemos registrado la interacción social de 50 niños en el patio de la escuela.
- Tenemos las respuestas en un cuestionario de 60 participantes en un taller de *mindfulness*.
- Tenemos los resultados de 1,000 alumnos en el examen de entrada de la UNAM.

Para hacer referencia a las unidades experimentales, se suele asignarles un número. Por ejemplo, se habla del resultado del “alumno *i*” en el examen.

Objetos, variables y observaciones

Para poder sistematizar los datos, se definen una o más variables.
Es común representar las variables por letras mayúsculas X, Y, etc.

Por ejemplo:

- Para el registro de la interacción social, se define la variable X: "Número de otros niños con los que habla".
- Para evaluar el efecto del taller, se definen las variables:
  - X_{pre}: "Puntuación en la aplicación *pre*"
  - X_{post}: "Puntuación en la aplicación *post*"
  - Y: "Diferencia entre las puntuación *pre* y *post*".
- Para el examen de opción múltiple, se definen las variables X_1 a X_{120}: "Puntuación en la pregunta j del examen".

Objetos, variables y observaciones

Para poder sistematizar los datos, se definen una o más variables.
Es común representar las variables por letras mayúsculas X, Y, etc.

Por ejemplo:

- Para el registro de la interacción social, se define la variable X: “Número de otros niños con los que habla”.
- Para evaluar el efecto del taller, se definen las variables:
  - X_{pre}: “Puntuación en la aplicación *pre*”
  - X_{post}: “Puntuación en la aplicación *post*”
  - Y: “Diferencia entre las puntuación *pre* y *post*”.
- Para el examen de opción múltiple, se definen las variables X_1 a X_{120}: “Puntuación en la pregunta j del examen”.

Objetos, variables y observaciones

Se observan para los objetos valores en las variables.

Por ejemplo:

- El niño i tuvo el valor de 8 en la variable X “Número de de otros niños con los que habla”:
$$x_i = 8.$$
- El alumno i contestó correctamente en la pregunta j del examen de opción múltiple.
Es decir, tiene el valor 1 en la variable X_j:
$$x_{ij} = 1.$$

Objetos, variables y observaciones

Se observan para los objetos valores en las variables.

Por ejemplo:

- El niño i tuvo el valor de 8 en la variable X "Número de de otros niños con los que habla":
$$x_i = 8.$$
- El alumno i contestó correctamente en la pregunta j del examen de opción múltiple.
Es decir, tiene el valor 1 en la variable X_j:
$$x_{ij} = 1.$$

Índice

El signo sumatorio Σ (Sigma)

¿Qué significa la siguiente expresión?

$$\sum_{k=3}^{6} \frac{k^2}{2} = ?$$

El signo sumatorio Σ (Sigma)

¿Qué significa la siguiente expresión?

$$\sum_{k=3}^{6} \frac{k^2}{2} = \underbrace{\frac{3^2}{2}}_{k=3} + \underbrace{\frac{4^2}{2}}_{k=4} + \underbrace{\frac{5^2}{2}}_{k=5} + \underbrace{\frac{6^2}{2}}_{k=6} = 43$$

- índice sumatorio: k
- término genérico: $\frac{k^2}{2}$
- límites: 3 (límite inferior) y 6 (límite superior)

El signo sumatorio Σ (Sigma)

Otro ejemplo:

$$\sum_{i=1}^{n} x_i \; = ?$$

El signo sumatorio Σ (Sigma)

Otro ejemplo:

$$\sum_{i=1}^{n} x_i = \underbrace{x_1}_{i=1} + \underbrace{x_2}_{i=2} + \underbrace{x_3}_{i=3} + \ldots + \underbrace{x_n}_{i=n}$$

El signo sumatorio Σ (Sigma)

Sumas dobles

“Sumas dobles”:

$$\sum_{i=2}^{4}\sum_{j=3}^{5} x_{ij} = ?$$

El signo sumatorio Σ (Sigma)

Sumas dobles

“Sumas dobles”:

$$\sum_{i=2}^{4}\sum_{j=3}^{5} x_{ij} = \sum_{i=2}^{4}\left(\sum_{j=3}^{5} x_{ij}\right)$$

El signo sumatorio Σ (Sigma)

Sumas dobles

“Sumas dobles”:

$$\sum_{i=2}^{4}\sum_{j=3}^{5} x_{ij} = \sum_{i=2}^{4}\left(\sum_{j=3}^{5} x_{ij}\right)$$

$$= \underbrace{\sum_{j=3}^{5} x_{2j}}_{i=2} + \underbrace{\sum_{j=3}^{5} x_{3j}}_{i=3} + \underbrace{\sum_{j=3}^{5} x_{4j}}_{i=4}$$

El signo sumatorio Σ (Sigma)

Sumas dobles

“Sumas dobles”:

$$\begin{aligned}
\sum_{i=2}^{4}\sum_{j=3}^{5} x_{ij} &= \sum_{i=2}^{4}\left(\sum_{j=3}^{5} x_{ij}\right) \\
&= \underbrace{\sum_{j=3}^{5} x_{2j}}_{i=2} + \underbrace{\sum_{j=3}^{5} x_{3j}}_{i=3} + \underbrace{\sum_{j=3}^{5} x_{4j}}_{i=4} \\
&= (x_{23} + x_{24} + x_{25}) + (x_{33} + x_{34} + x_{35}) + (x_{43} + x_{44} + x_{45})
\end{aligned}$$

El signo sumatorio Σ (Sigma)

Sumas dobles

En una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio

$$\sum_{i=1}^{4}\sum_{j=i+1}^{4}(x_i - x_j)^2 = ?$$

El signo sumatorio Σ (Sigma)

Sumas dobles

En una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio

$$\sum_{i=1}^{4}\sum_{j=i+1}^{4}(x_i - x_j)^2 = \sum_{i=1}^{4}\left[\sum_{j=i+1}^{4}(x_i - x_j)^2\right]$$

El signo sumatorio Σ (Sigma)

Sumas dobles

En una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio

$$\sum_{i=1}^{4}\sum_{j=i+1}^{4}(x_i - x_j)^2 = \sum_{i=1}^{4}\left[\sum_{j=i+1}^{4}(x_i - x_j)^2\right]$$

$$= \underbrace{\sum_{j=1+1}^{4}(x_1 - x_j)^2}_{i=1} + \underbrace{\sum_{j=2+1}^{4}(x_2 - x_j)^2}_{i=2} + \underbrace{\sum_{j=3+1}^{4}(x_3 - x_j)^2}_{i=3} + \underbrace{\sum_{j=4+1}^{4}(x_4 - x_j)^2}_{i=4}$$

El signo sumatorio Σ (Sigma)

Sumas dobles

En una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio

$$\sum_{i=1}^{4}\sum_{j=i+1}^{4}(x_i - x_j)^2 = \sum_{i=1}^{4}\left[\sum_{j=i+1}^{4}(x_i - x_j)^2\right]$$

$$= \underbrace{\sum_{j=1+1}^{4}(x_1 - x_j)^2}_{i=1} + \underbrace{\sum_{j=2+1}^{4}(x_2 - x_j)^2}_{i=2} + \underbrace{\sum_{j=3+1}^{4}(x_3 - x_j)^2}_{i=3} + \underbrace{\sum_{j=4+1}^{4}(x_4 - x_j)^2}_{i=4}$$

$$= \sum_{j=2}^{4}(x_1 - x_j)^2 + \sum_{j=3}^{4}(x_2 - x_j)^2 + \sum_{j=4}^{4}(x_3 - x_j)^2 + \sum_{j=5}^{4}(x_4 - x_j)^2$$

El signo sumatorio Σ (Sigma)

Sumas dobles

En una suma doble, el límite inferior o superior (o ambos) del segundo signo sumatorio puede contener el índice del primer signo sumatorio

$$\sum_{i=1}^{4}\sum_{j=i+1}^{4}(x_i - x_j)^2 = \sum_{i=1}^{4}\left[\sum_{j=i+1}^{4}(x_i - x_j)^2\right]$$

$$= \underbrace{\sum_{j=1+1}^{4}(x_1 - x_j)^2}_{i=1} + \underbrace{\sum_{j=2+1}^{4}(x_2 - x_j)^2}_{i=2} + \underbrace{\sum_{j=3+1}^{4}(x_3 - x_j)^2}_{i=3} + \underbrace{\sum_{j=4+1}^{4}(x_4 - x_j)^2}_{i=4}$$

$$= \sum_{j=2}^{4}(x_1 - x_j)^2 + \sum_{j=3}^{4}(x_2 - x_j)^2 + \sum_{j=4}^{4}(x_3 - x_j)^2 + \sum_{j=5}^{4}(x_4 - x_j)^2$$

$$= \left[(x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_1 - x_4)^2\right] + \left[(x_2 - x_3)^2 + (x_2 - x_4)^2\right] + \left[(x_3 - x_4)^2\right]$$

Índice

Nivel de medición

Ejemplo

El sociólogo que investiga el comportamiento electoral en las últimas elecciones define una variable X númerica que asigna los siguientes valores a cada candidato para la presidencia:

- Ricardo Anaya Cortés $\rightarrow X = 1$
- José Antonio Meade Kuribreña $\rightarrow X = 2$
- Andrés Manuel López Obrador $\rightarrow X = 3$
- Jaime Rodríguez Calderón $\rightarrow X = 4$
- otro candidato $\rightarrow X = 5$.

¿Qué operaciones sobre la variable X tienen sentido?
¿Qué significa ¿$3 > 1$?, ¿$1 + 2 = 3$?, etc.

Nivel de medición

El nivel de medición (o el nivel de la escala):

- considera la admisibilidad de las operaciones aritméticas a los valores de la variable.
- depende de las transformaciones que se pueden aplicar a los valores sin que cambie la interpretación empírica (Más libertad que hay para transformar, menos información está presente en los valores.)

Escala nominal

Definición

Si una escala es de medida nominal, entonces los números asignados a las observaciones se pueden transformar a otros números cualesquiera, siempre y cuando:

- dos observaciones que tienen el mismo número asignado antes de la transformación, siguen teniendo el mismo número después de la transformación;
- dos observaciones que tienen asignados números diferentes, siguen con números diferentes después de la transformación.

Ejemplo

En el ejemplo de los partidos políticos, podemos aplicar sin problemas esta transformación:

		Valor anterior		Valor nuevo
Votó por:		X		$X^\star$
RAC	$\longrightarrow$	1	$\longrightarrow$	0
JAMK	$\longrightarrow$	2	$\longrightarrow$	2
AMLO	$\longrightarrow$	3	$\longrightarrow$	-1
JRC	$\longrightarrow$	4	$\longrightarrow$	3.8
Otro	$\longrightarrow$	5	$\longrightarrow$	$\sqrt{\pi}$

Escala nominal

Definición

Si una escala es de medida nominal, entonces los números asignados a las observaciones se pueden transformar a otros números cualesquiera, siempre y cuando:

- dos observaciones que tienen el mismo número asignado antes de la transformación, siguen teniendo el mismo número después de la transformación;
- dos observaciones que tienen asignados números diferentes, siguen con números diferentes después de la transformación.

Ejemplo

En el ejemplo de los partidos políticos, podemos aplicar sin problemas esta transformación:

Votó por:		Valor anterior X		Valor nuevo $X^\star$
RAC	$\longrightarrow$	1	$\longrightarrow$	0
JAMK	$\longrightarrow$	2	$\longrightarrow$	2
AMLO	$\longrightarrow$	3	$\longrightarrow$	-1
JRC	$\longrightarrow$	4	$\longrightarrow$	3.8
Otro	$\longrightarrow$	5	$\longrightarrow$	$\sqrt{\pi}$

Escala ordinal

Definición

Si una escala es de medida ordinal, entonces los números asignados a las observaciones se pueden transformar a otros números cualesquiera, siempre y cuando la transformación respeta el orden en los valores originales.

Ejemplo

Por ejemplo, si pedimos a un estudiante evaluar el desempeño de su profesor utilizando cuatro categorías

		Valor anterior X		Valor nuevo X^*
Totalmente insatisfactorio	$\longrightarrow$	1	$\longrightarrow$	-4
Poco satisfactorio	$\longrightarrow$	2	$\longrightarrow$	8
Neutro	$\longrightarrow$	3	$\longrightarrow$	9
Satisfactorio	$\longrightarrow$	4	$\longrightarrow$	10
Muy satisfactorio	$\longrightarrow$	5	$\longrightarrow$	10.1

Escala ordinal

Definición

Si una escala es de medida ordinal, entonces los números asignados a las observaciones se pueden transformar a otros números cualesquiera, siempre y cuando la transformación respeta el orden en los valores originales.

Ejemplo

Por ejemplo, si pedimos a un estudiante evaluar el desempeño de su profesor utilizando cuatro categorías

		Valor anterior X		Valor nuevo $X^\star$
Totalmente insatisfactorio	$\longrightarrow$	1	$\longrightarrow$	-4
Poco satisfactorio	$\longrightarrow$	2	$\longrightarrow$	8
Neutro	$\longrightarrow$	3	$\longrightarrow$	9
Satisfactorio	$\longrightarrow$	4	$\longrightarrow$	10
Muy satisfactorio	$\longrightarrow$	5	$\longrightarrow$	10.1

Escala de intervalo

Definición

Si una escala es de nivel intervalo, entonces la transformación debe no solo respeta el orden, sino también las diferencias relativas entre los valores originales. Es decir, solo transformaciones lineales están permitidas.

Ejemplo

Por ejemplo, si medimos la temperatura corporal de un bebé:

Valor anterior C°		Valor nuevo F°
36	$\longrightarrow$	96.8
37	$\longrightarrow$	98.6
38	$\longrightarrow$	100.4
39	$\longrightarrow$	102.2
40	$\longrightarrow$	104.0

Escala de intervalo

Definición

Si una escala es de nivel intervalo, entonces la transformación debe no solo respeta el orden, sino también las diferencias relativas entre los valores originales. Es decir, solo transformaciones lineales están permitidas.

Ejemplo

Por ejemplo, si medimos la temperatura corporal de un bebé:

Valor anterior C°		Valor nuevo F°
36	$\longrightarrow$	96.8
37	$\longrightarrow$	98.6
38	$\longrightarrow$	100.4
39	$\longrightarrow$	102.2
40	$\longrightarrow$	104.0

Escala de razón

Definición

Si una escala es de nivel razón, entonces la transformación debe respeta las razones entre los valores originales. Es decir, solo se permiten transformaciones de multiplicar los valores originales con una constante

Ejemplo

Por ejemplo, si medimos la distancia de los capitales de los estados hacia el zocalo del Distrito Federal:

	Valor anterior X (km.)		Valor nuevo $X^{\star}$ (millas)
Chihuahua	1009	$\longrightarrow$	627
Merida	1240	$\longrightarrow$	771
Oaxaca	365	$\longrightarrow$	227
Puebla	110	$\longrightarrow$	68
Veracruz	313	$\longrightarrow$	194

Escala de razón

Definición

Si una escala es de nivel razón, entonces la transformación debe respeta las razones entre los valores originales. Es decir, solo se permiten transformaciones de multiplicar los valores originales con una constante

Ejemplo

Por ejemplo, si medimos la distancia de los capitales de los estados hacia el zocalo del Distrito Federal:

	Valor anterior X (km.)		Valor nuevo $X^\star$ (millas)
Chihuahua	1009	$\longrightarrow$	627
Merida	1240	$\longrightarrow$	771
Oaxaca	365	$\longrightarrow$	227
Puebla	110	$\longrightarrow$	68
Veracruz	313	$\longrightarrow$	194

Escala absoluto

Definición

Si una escala es de nivel absoluto, no se permite ninguna transformación.

Índice

Índice

Distribuciones univariadas empíricas

Datos observados de una muestra

Supongamos que hemos registrado de 50 pacientes el índice de masa corporal (IMC):

No.	IMC	No.	IMC	No.	IMC	No.	IMC	No.	IMC
1	25	11	25	21	42	31	26	41	29
2	25	12	21	22	27	32	27	42	28
3	24	13	22	23	22	33	24	43	23
4	20	14	30	24	25	34	33	44	27
5	22	15	26	25	34	35	24	45	27
6	23	16	28	26	22	36	22	46	19
7	24	17	23	27	22	37	23	47	24
8	26	18	20	28	24	38	29	48	23
9	20	19	19	29	26	39	24	49	26
10	25	20	28	30	22	40	23	50	27

Distribuciones univariadas empíricas

Datos observados de una muestra

Supongamos que hemos registrado de 50 pacientes el índice de masa corporal (IMC):

No.	IMC	No.	IMC	No.	IMC	No.	IMC	No.	IMC
1	25	11	25	21	42	31	26	41	29
2	25	12	21	22	27	32	27	42	28
3	24	13	22	23	22	33	24	43	23
4	20	14	30	24	25	34	33	44	27
5	22	15	26	25	34	35	24	45	27
6	23	16	28	26	22	36	22	46	19
7	24	17	23	27	22	37	23	47	24
8	26	18	20	28	24	38	29	48	23
9	20	19	19	29	26	39	24	49	26
10	25	20	28	30	22	40	23	50	27

$frec(25) = 5$

Distribuciones univariadas empíricas

Datos observados de una muestra

Supongamos que hemos registrado de 50 pacientes el índice de masa corporal (IMC):

No.	IMC	No.	IMC	No.	IMC	No.	IMC	No.	IMC
1	25	11	25	21	42	31	26	41	29
2	25	12	21	22	27	32	27	42	28
3	24	13	22	23	22	33	24	43	23
4	20	14	30	24	25	34	33	44	27
5	22	15	26	25	34	35	24	45	27
6	23	16	28	26	22	36	22	46	19
7	24	17	23	27	22	37	23	47	24
8	26	18	20	28	24	38	29	48	23
9	20	19	19	29	26	39	24	49	26
10	25	20	28	30	22	40	23	50	27

$frec(25) = 5$

$p(29) = \frac{2}{50} = .04$

Distribuciones univariadas empíricas

Función de frecuencia y proporción

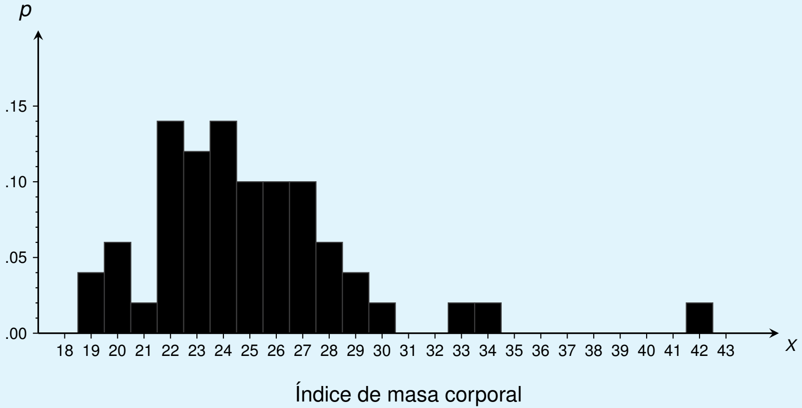
Se puede resumir la distribución de la variable X (IMC) en una tabla, que muestra la función de frecuencia y la función de proporción:

j	x_j	$frec(x_j)$	$p(x_j)$
1	19	2	.04
2	20	3	.06
3	21	1	.02
4	22	7	.14
5	23	6	.12
6	24	7	.14
7	25	5	.10
8	26	5	.10
9	27	5	.10
10	28	3	.06
11	29	2	.04
12	30	1	.02
13	33	1	.02
14	34	1	.02
15	42	1	.02
Total		50	1.00

Distribuciones univariadas empíricas

Función de frecuencia y proporción

Se puede representar gráficamente la función de frecuencia o proporción:



Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.
- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.
- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.
- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.

- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.

- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.

- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.
- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.
- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.
- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.
- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.
- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.
- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.
- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.
- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.
- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.

- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.

- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.

- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Modelo estadístico

Supongamos que:

- Se administra un examen de física cuántica que consiste en 10 preguntas de opción múltiple con tres opciones de respuesta.
  La puntuación X en el examen es el número de respuestas correctas.
- Una persona que no sabe nada de física cuántica lo responde.
  Elige en cada pregunta aleatoriamente una respuesta.
- ¿Cuál es la probabilidad de que esta persona obtenga un puntaje de x ($x = 0, 1, 2, \ldots, 10$) en el examen?

Para contestar esta pregunta, se especifica un modelo estadístico, donde, por ejemplo, se supone que:

- La probabilidad de elegir la respuesta correcta en cada pregunta es $\frac{1}{3}$.
- Las respuestas son independientes.

Lo anterior, estadísticamente, se traduce diciendo que X tiene una distribución binomial:

$$X \sim \text{Binomial}\left(10, \frac{1}{3}\right).$$

Distribuciones univariadas teóricas para variables discretas

Función de probabilidad

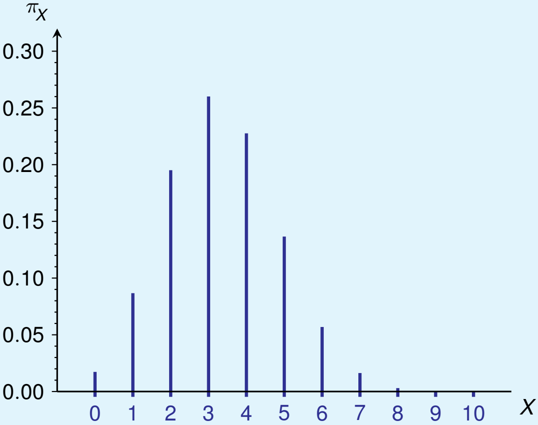
Esta distribución teórica de la variable X (puntaje en el examen) se puede resumir en una tabla, mostrando la función de probabilidad de X:

j	x_j	$\pi_X(x_j)$
1	0	.017
2	1	.087
3	2	.195
4	3	.260
5	4	.228
6	5	.137
7	6	.057
8	7	.016
9	8	.003
10	9	.000
11	10	.000
Total		1.000

Distribuciones univariadas teóricas para variables discretas

Función de probabilidad

Y se puede representar gráficamente:



Distribuciones univariadas teóricas para variables continuas

Modelo estadístico y función de densidad

Supongamos que:

- Se aplica un examen en la computadora y, para una pregunta concreta, se registra el tiempo X (en segundos) entre el momento en que se muestra la pregunta en la pantalla y el momento de la selección de la respuesta;
- La variable X tiene la siguiente distribución:

Distribuciones univariadas teóricas para variables continuas

Modelo estadístico y función de densidad

Supongamos que:

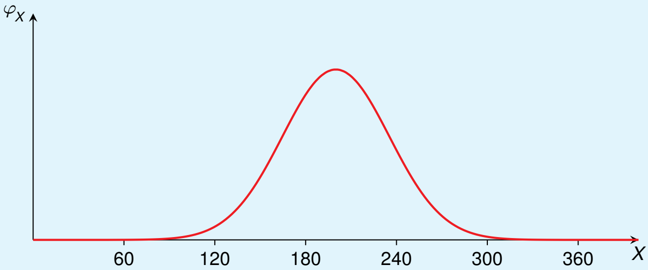
- Se aplica un examen en la computadora y, para una pregunta concreta, se registra el tiempo X (en segundos) entre el momento en que se muestra la pregunta en la pantalla y el momento de la selección de la respuesta;
- La variable X tiene la siguiente distribución:

Distribuciones univariadas teóricas para variables continuas

Modelo estadístico y función de densidad

Supongamos que:

- Se aplica un examen en la computadora y, para una pregunta concreta, se registra el tiempo X (en segundos) entre el momento en que se muestra la pregunta en la pantalla y el momento de la selección de la respuesta;
- La variable X tiene la siguiente distribución:

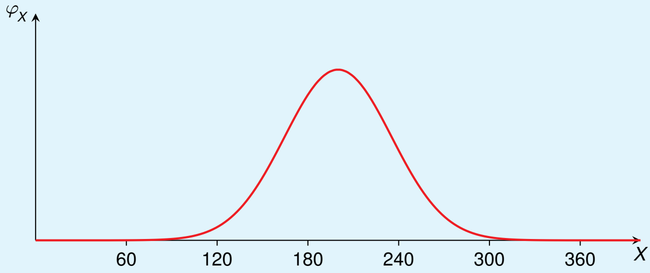


Distribuciones univariadas teóricas para variables continuas

Modelo estadístico y función de densidad

Supongamos que:

- Se aplica un examen en la computadora y, para una pregunta concreta, se registra el tiempo X (en segundos) entre el momento en que se muestra la pregunta en la pantalla y el momento de la selección de la respuesta;
- La variable X tiene la siguiente distribución:



Nota: φ_X se llama la función de densidad de la variable X.

Índice

Estadísticos y parámetros de tendencia central

La media aritmética y el valor esperado

- Para distribuciones empíricas (la variable X observada en una muestra), se define la media aritmética:

$$\overline{x} = \sum_{j=1}^{m} p(x_j) x_j,$$

donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la media poblacional o el valor esperado.
  - Para variables discretas:

$$\mathscr{E}(X) = \sum_{j=1}^{m} \pi_X(x_j)\, x_j,$$

donde la suma es entre todos los posibles valores que puede asumir la variable X.

  - Para variables continuas:

$$\mathscr{E}(X) = \int_{-\infty}^{+\infty} \varphi_X(x)\, x\, dx.$$

Estadísticos y parámetros de tendencia central

La media aritmética y el valor esperado

- Para distribuciones empíricas (la variable X observada en una muestra), se define la media aritmética:

  $$\overline{x} = \sum_{j=1}^{m} p(x_j) x_j ,$$

  donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la media poblacional o el valor esperado.
  - Para variables discretas:

    $$\mathscr{E}(X) = \sum_{j=1}^{m} \pi_X(x_j)\, x_j ,$$

    donde la suma es entre todos los posibles valores que puede asumir la variable X.
  - Para variables continuas:

    $$\mathscr{E}(X) = \int_{-\infty}^{+\infty} \varphi_X(x)\, x\, dx .$$

Estadísticos y parámetros de tendencia central

La media aritmética y el valor esperado

- Para distribuciones empíricas (la variable X observada en una muestra), se define la media aritmética:

$$\overline{x} = \sum_{j=1}^{m} p(x_j)x_j,$$

donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la media poblacional o el valor esperado.
  - Para variables discretas:

$$\mathscr{E}(X) = \sum_{j=1}^{m} \pi_X(x_j)\, x_j,$$

donde la suma es entre todos los posibles valores que puede asumir la variable X.

  - Para variables continuas:

$$\mathscr{E}(X) = \int_{-\infty}^{+\infty} \varphi_X(x)\, x\, dx.$$

Estadísticos y parámetros de tendencia central

Efecto de una transformación lineal en la media

- Definimos la nueva variable Y a partir de la variable X:

$$Y = aX + b,$$

donde a y b son dos constantes cualesquiera.

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\overline{y} = a\overline{x} + b.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = a\mathscr{E}(X) + b.$$

Estadísticos y parámetros de tendencia central

Efecto de una transformación lineal en la media

- Definimos la nueva variable Y a partir de la variable X:

$$Y = aX + b,$$

donde a y b son dos constantes cualesquiera.

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\bar{y} = a\bar{x} + b.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = a\mathscr{E}(X) + b.$$

Estadísticos y parámetros de tendencia central

Efecto de una transformación lineal en la media

- Definimos la nueva variable Y a partir de la variable X:

$$Y = aX + b,$$

donde a y b son dos constantes cualesquiera.

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\overline{y} = a\overline{x} + b.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = a\mathscr{E}(X) + b.$$

Estadísticos y parámetros de tendencia central

La media de una suma de variables

- Definimos la nueva variable Y a partir de las variable X_1 y X_2 como:

$$Y = X_1 + X_2.$$

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\overline{y} = \overline{x}_1 + \overline{x}_2.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = \mathscr{E}(X_1) + \mathscr{E}(X_2).$$

Estadísticos y parámetros de tendencia central

La media de una suma de variables

- Definimos la nueva variable Y a partir de las variable X_1 y X_2 como:

$$Y = X_1 + X_2.$$

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\overline{y} = \overline{x}_1 + \overline{x}_2.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = \mathscr{E}(X_1) + \mathscr{E}(X_2).$$

Estadísticos y parámetros de tendencia central

La media de una suma de variables

- Definimos la nueva variable Y a partir de las variable X_1 y X_2 como:

$$Y = X_1 + X_2.$$

- Entonces, la media aritmética de la variable Y se obtiene por:

$$\overline{y} = \overline{x}_1 + \overline{x}_2.$$

- Entonces, la media poblacional de la variable Y se obtiene por:

$$\mathscr{E}(Y) = \mathscr{E}(X_1) + \mathscr{E}(X_2).$$

Índice

Estadísticos y parámetros de variabilidad o dispersión

La varianza

- Para distribuciones empíricas, se define la varianza muestral:

$$s_X^2 = \sum_{j=1}^{m} p(x_j)(x_j - \overline{x})^2,$$

donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la varianza poblacional como:

$$\sigma_X^2 = \mathscr{E}\left[(X - \mu_X)^2\right],$$

donde μ_X es $\mathscr{E}(X)$.

  - Para variables discretas:

$$\sigma_X^2 = \sum_{j=1}^{m} \pi_X(x_j)\,(x_j - \mu_X)^2$$

  - para variables continuas:

$$\sigma_X^2 = \int_{-\infty}^{+\infty} \varphi_X(x)\,(x - \mu_X)^2\,dx$$

Estadísticos y parámetros de variabilidad o dispersión

La varianza

- Para distribuciones empíricas, se define la varianza muestral:

$$s_X^2 = \sum_{j=1}^{m} p(x_j)(x_j - \overline{x})^2,$$

donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la varianza poblacional como:

$$\sigma_X^2 = \mathscr{E}\left[(X - \mu_X)^2\right],$$

donde μ_X es $\mathscr{E}(X)$.

  - Para variables discretas:

$$\sigma_X^2 = \sum_{j=1}^{m} \pi_X(x_j)\,(x_j - \mu_X)^2$$

  - para variables continuas:

$$\sigma_X^2 = \int_{-\infty}^{+\infty} \varphi_X(x)\,(x - \mu_X)^2\,dx$$

Estadísticos y parámetros de variabilidad o dispersión

La varianza

- Para distribuciones empíricas, se define la varianza muestral:

$$s_X^2 = \sum_{j=1}^{m} p(x_j)(x_j - \overline{x})^2,$$

donde la suma es entre todos los *distintos* valores que se han observado en la muestra.

- Para distribuciones teóricas, se define la varianza poblacional como:

$$\sigma_X^2 = \mathscr{E}\left[(X - \mu_X)^2\right],$$

donde μ_X es $\mathscr{E}(X)$.

  - Para variables discretas:

$$\sigma_X^2 = \sum_{j=1}^{m} \pi_X(x_j)\,(x_j - \mu_X)^2$$

  - para variables continuas:

$$\sigma_X^2 = \int_{-\infty}^{+\infty} \varphi_X(x)\,(x - \mu_X)^2\,dx$$

Estadísticos y parámetros de variabilidad o dispersión

La desviación estándar

La desviación estándar es la raiz cuadrada de la varianza:

- Para distribuciones empíricas:

$$s_X = \sqrt{s_X^2}$$

- Para distribuciones teóricas:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Estadísticos y parámetros de variabilidad

Efecto de una transformación lineal en la varianza y desviación estándar

Si la variable Y se obtiene por una transformación lineal de la variable X, es decir $Y = aX + b$, entonces:

- la varianza y desviación estándar muestral de la variable Y se obtiene por:

$$s_Y^2 = a^2 \, s_X^2$$
$$s_Y = |a| \, s_X$$

- la varianza y desviación estándar poblacional de la variable Y se obtiene por:

$$\sigma_Y^2 = a^2 \, \sigma_X^2$$
$$\sigma_Y = |a| \, \sigma_X$$

Estadísticos y parámetros de variabilidad

Efecto de una transformación lineal en la varianza y desviación estándar

Si la variable Y se obtiene por una transformación lineal de la variable X, es decir $Y = aX + b$, entonces:

- la varianza y desviación estándar muestral de la variable Y se obtiene por:

$$s_Y^2 = a^2 \, s_X^2$$
$$s_Y = |a| \, s_X$$

- la varianza y desviación estándar poblacional de la variable Y se obtiene por:

$$\sigma_Y^2 = a^2 \, \sigma_X^2$$
$$\sigma_Y = |a| \, \sigma_X$$

Índice

Función de frecuencia y proporción acumulada

Definición

Definición

Para una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$ donde $x_1 < x_2 < x_3 < \cdots < x_m$, se define:

La función de frecuencia acumulada:

$$\boxed{afrec(x_j) = \sum_{k=1}^{j} frec(x_k)}$$

La función de proporción acumulada:

$$\boxed{F(x_j) = \sum_{k=1}^{j} p(x_k)}$$

Función de frecuencia y proporción acumulada

Definición

Definición

Para una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$ donde $x_1 < x_2 < x_3 < \cdots < x_m$, se define:

La función de frecuencia acumulada:

$$\boxed{afrec(x_j) = \sum_{k=1}^{j} frec(x_k)}$$

La función de proporción acumulada:

$$\boxed{F(x_j) = \sum_{k=1}^{j} p(x_k)}$$

Función de frecuencia y proporción acumulada

Definición

Definición

Para una variable X, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$ donde $x_1 < x_2 < x_3 < \cdots < x_m$, se define:

La función de frecuencia acumulada:

$$\boxed{afrec(x_j) = \sum_{k=1}^{j} frec(x_k)}$$

La función de proporción acumulada:

$$\boxed{F(x_j) = \sum_{k=1}^{j} p(x_k)}$$

Función de frecuencia y proporción acumulada

Función de frecuencia y proporción acumulada

j	x_j	$frec(x_j)$	$p(x_j)$
1	19	2	.04
2	20	3	.06
3	21	1	.02
4	22	7	.14
5	23	6	.12
6	24	7	.14
7	25	5	.10
8	26	5	.10
9	27	5	.10
10	28	3	.06
11	29	2	.04
12	30	1	.02
13	33	1	.02
14	34	1	.02
15	42	1	.02
Total		50	1.00

Función de frecuencia y proporción acumulada

Función de frecuencia y proporción acumulada

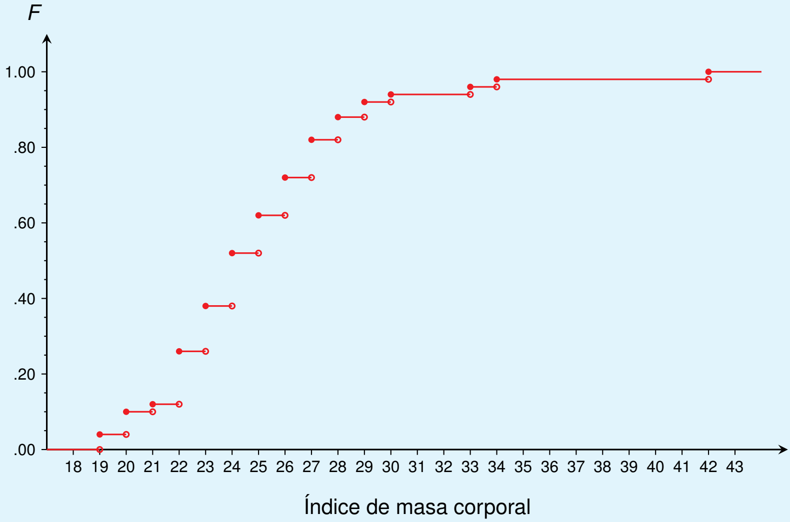
j	x_j	$frec(x_j)$	$p(x_j)$	$afrec(x_j)$
1	19	2	.04	2
2	20	3	.06	5
3	21	1	.02	6
4	22	7	.14	13
5	23	6	.12	19
6	24	7	.14	26
7	25	5	.10	31
8	26	5	.10	36
9	27	5	.10	41
10	28	3	.06	44
11	29	2	.04	46
12	30	1	.02	47
13	33	1	.02	48
14	34	1	.02	49
15	42	1	.02	50
Total		50	1.00	

Función de frecuencia y proporción acumulada

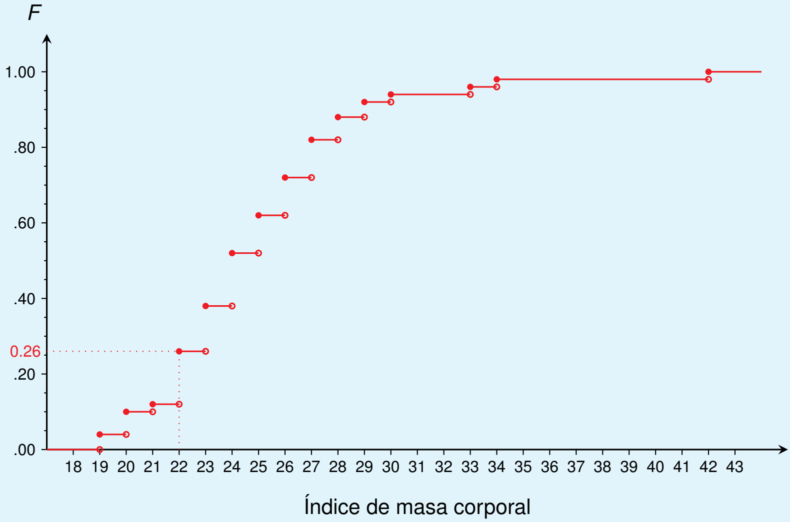
Función de frecuencia y proporción acumulada

j	x_j	$frec(x_j)$	$p(x_j)$	$afrec(x_j)$	$F(x_j)$
1	19	2	.04	2	.04
2	20	3	.06	5	.10
3	21	1	.02	6	.12
4	22	7	.14	13	.26
5	23	6	.12	19	.38
6	24	7	.14	26	.52
7	25	5	.10	31	.62
8	26	5	.10	36	.72
9	27	5	.10	41	.82
10	28	3	.06	44	.88
11	29	2	.04	46	.92
12	30	1	.02	47	.94
13	33	1	.02	48	.96
14	34	1	.02	49	.98
15	42	1	.02	50	1.00
Total		50	1.00		

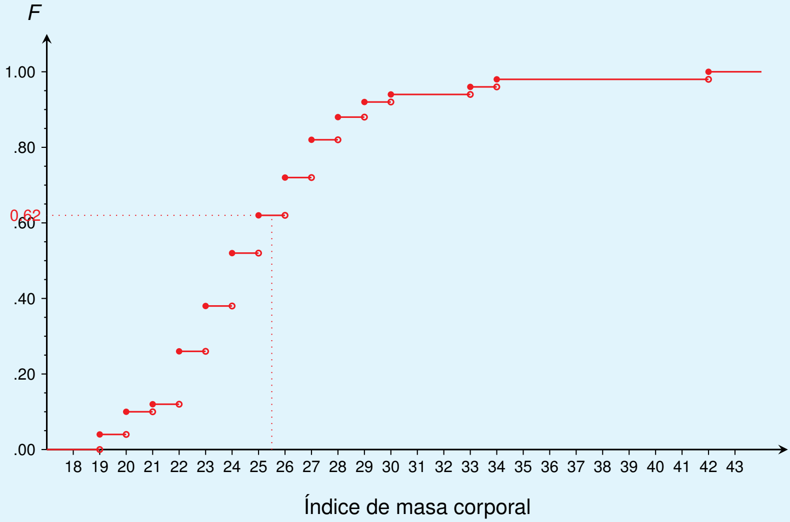
Representación gráfica de la función de proporción acumulada



Representación gráfica de la función de proporción acumulada



Representación gráfica de la función de proporción acumulada



Función de distribución

Definición

Definición

Para una variable X discreta, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$ donde $x_1 < x_2 < x_3 < \cdots < x_m$, se define la función de distribución como:

$$\boxed{\Phi_X(x_j) = \sum_{k=1}^{j} \pi_X(x_k)}$$

Definición

Para una variable X continua, se define la función de distribución como:

$$\Phi_X(x) = \int_{-\infty}^{x} \varphi_X(u)du,$$

lo cual corresponde con el área debajo de φ_X a la izquierda de x.

Función de distribución

Definición

Definición

Para una variable X discreta, que asume m distintos valores $x_1, x_2, x_3, \ldots, x_m$ donde $x_1 < x_2 < x_3 < \cdots < x_m$, se define la función de distribución como:

$$\boxed{\Phi_X(x_j) = \sum_{k=1}^{j} \pi_X(x_k)}$$

Definición

Para una variable X continua, se define la función de distribución como:

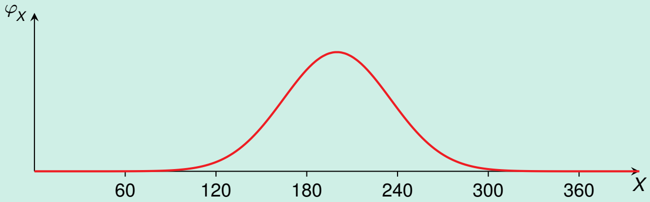
$$\Phi_X(x) = \int_{-\infty}^{x} \varphi_X(u)du,$$

lo cual corresponde con el área debajo de φ_X a la izquierda de x.

Función de distribución

Definición

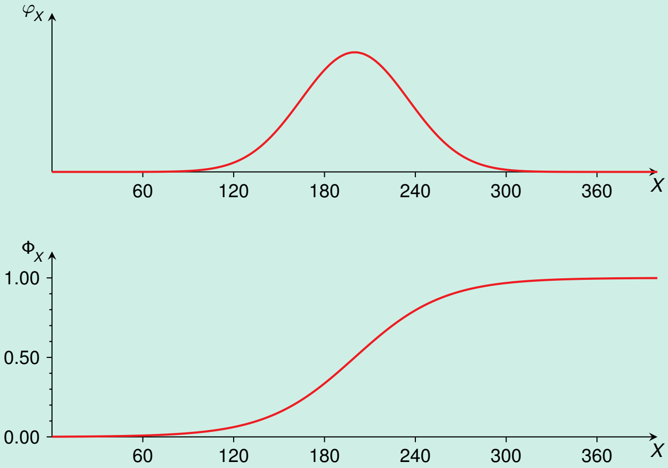
La distribución normal acumulada



Función de distribución

Definición

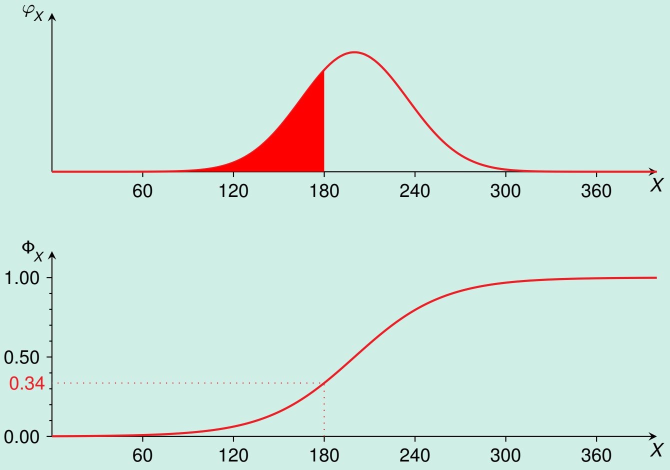
La distribución normal acumulada



Función de distribución

Definición

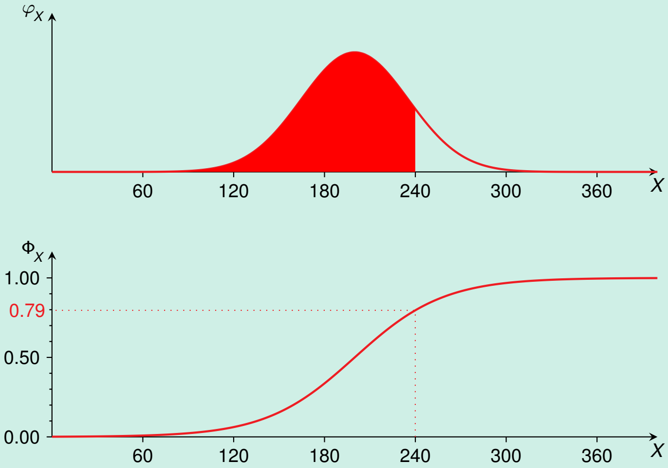
La distribución normal acumulada



Función de distribución

Definición

La distribución normal acumulada



Índice

Variables cuantitativas: Cuantiles

Los cuantiles se definen a partir de la distribución acumulada.

Para distribuciones empíricas:

Definición

Para cualquier r, $0 < r < 1$, el cuantil de orden r se define como sigue:

- Caso 1: Existe (al menos) un valor x, para el cual $F(x) = r$
  Todos los valores x que cumplen $F(x) = r$ son un cuantil de orden r.

  Si se quiere un valor único, se calcula:

  $$\frac{x_{\text{min}} + x_{\text{max}}}{2},$$

  donde x_{min} y x_{max} son los valores mínimos y máximos, respectivamente, que cumplen $F(x) = r$.

- Caso 2: No existe un valor x, para el cual $F(x) = r$
  El cuantil de orden r es el valor x más pequeño que cumple $F(x) > r$.

Notación: x_r

Variables cuantitativas: Cuantiles

Los cuantiles se definen a partir de la distribución acumulada.

Para distribuciones empíricas:

Definición

Para cualquier r, $0 < r < 1$, el cuantil de orden r se define como sigue:

- Caso 1: Existe (al menos) un valor x, para el cual $F(x) = r$
  Todos los valores x que cumplen $F(x) = r$ son un cuantil de orden r.

  Si se quiere un valor único, se calcula:

  $$\frac{x_{\text{min}} + x_{\text{max}}}{2},$$

  donde x_{min} y x_{max} son los valores mínimos y máximos, respectivamente, que cumplen $F(x) = r$.

- Caso 2: No existe un valor x, para el cual $F(x) = r$
  El cuantil de orden r es el valor x más pequeño que cumple $F(x) > r$.

Notación: x_r

Variables cuantitativas: Cuantiles

Los cuantiles se definen a partir de la distribución acumulada.

Para distribuciones empíricas:

Definición

Para cualquier r, $0 < r < 1$, el cuantil de orden r se define como sigue:

- Caso 1: Existe (al menos) un valor x, para el cual $F(x) = r$
  Todos los valores x que cumplen $F(x) = r$ son un cuantil de orden r.

  Si se quiere un valor único, se calcula:

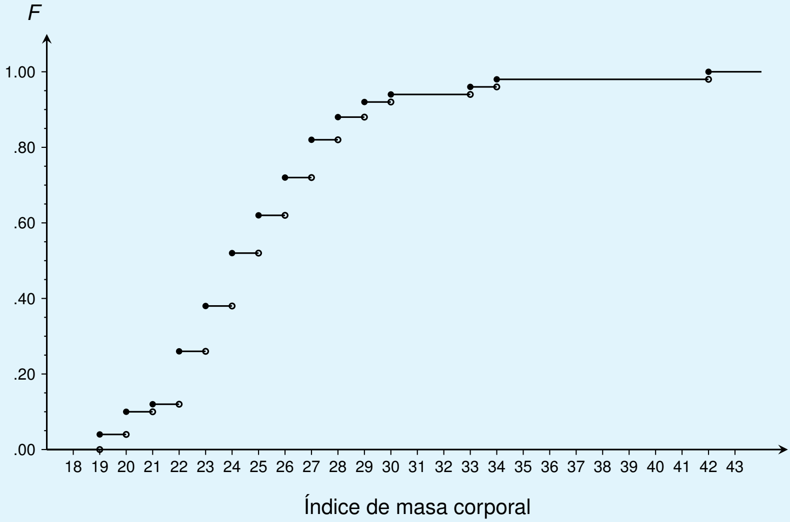
  $$\frac{x_{\min} + x_{\max}}{2},$$

  donde $x_{\min}$ y $x_{\max}$ son los valores mínimos y máximos, respectivamente, que cumplen $F(x) = r$.

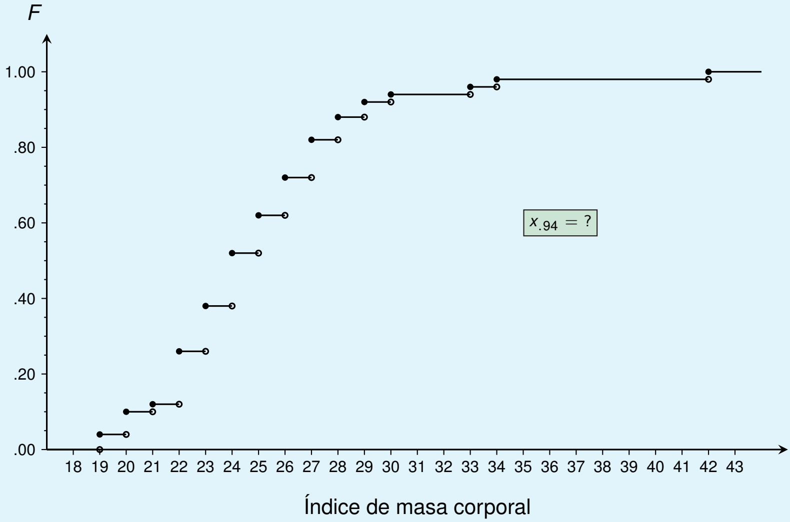
- Caso 2: No existe un valor x, para el cual $F(x) = r$
  El cuantil de orden r es el valor x más pequeño que cumple $F(x) > r$.

Notación: x_r

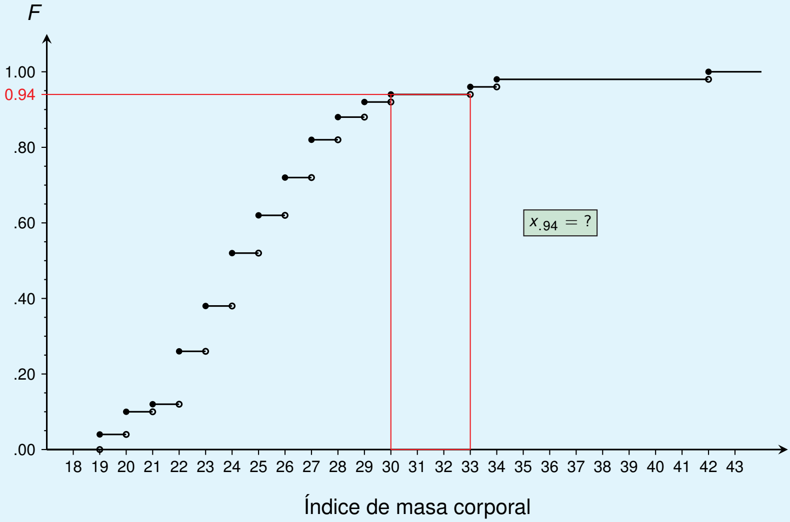
Hallar cuantiles gráficamente



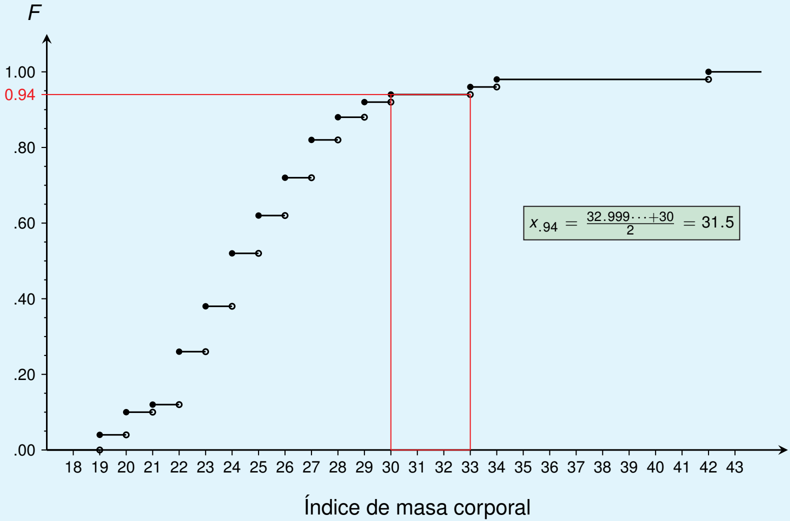
Hallar cuantiles gráficamente



Hallar cuantiles gráficamente

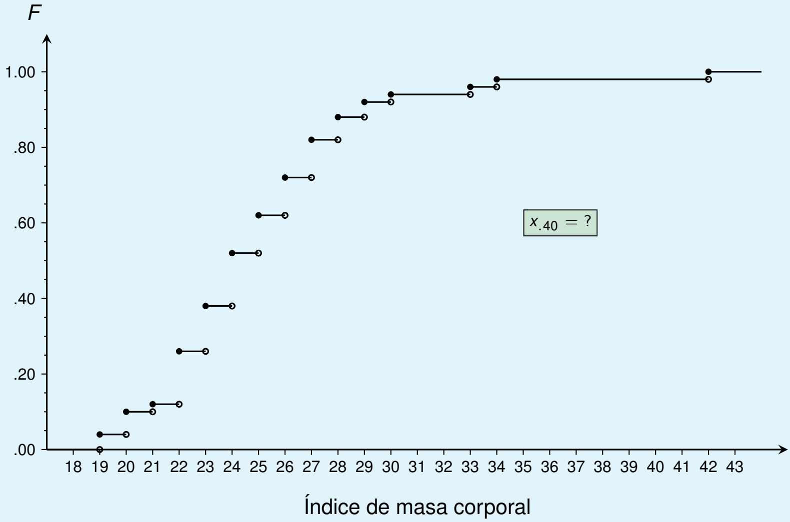


Hallar cuantiles gráficamente

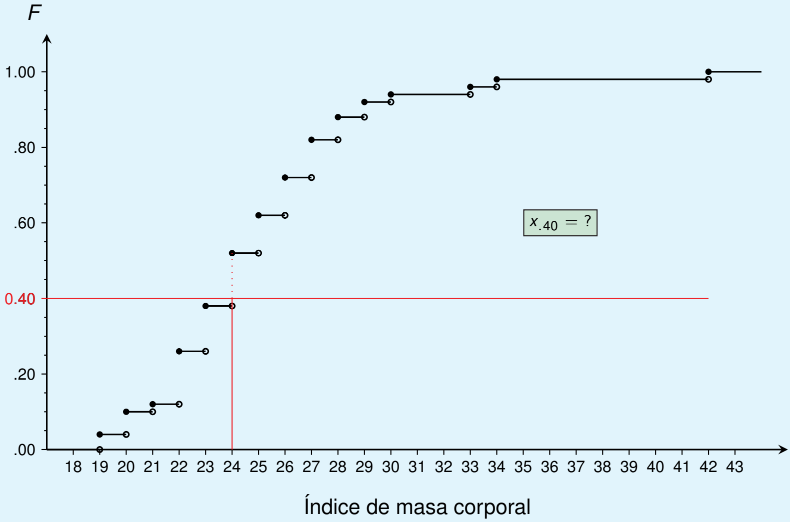


$$x_{.94} = \frac{32.999\cdots+30}{2} = 31.5$$

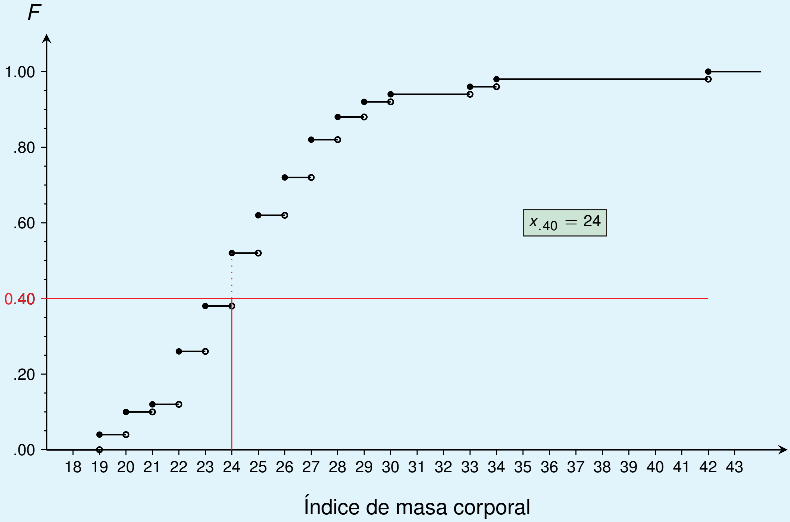
Hallar cuantiles gráficamente



Hallar cuantiles gráficamente



Hallar cuantiles gráficamente



Variables cuantitativas: Cuantiles

Los cuantiles se definen a partir de la distribución acumulada.

Para distribuciones teóricas:

Definición

Para cualquier r, $0 < r < 1$, el cuantil de orden r se define como sigue:

- Caso 1: Existe (al menos) un valor x, para el cual $\Phi_X(x) = r$
  Todos los valores x que cumplen $\Phi_X(x) = r$ son un cuantil de orden r.

  Si se quiere un valor único, se calcula:

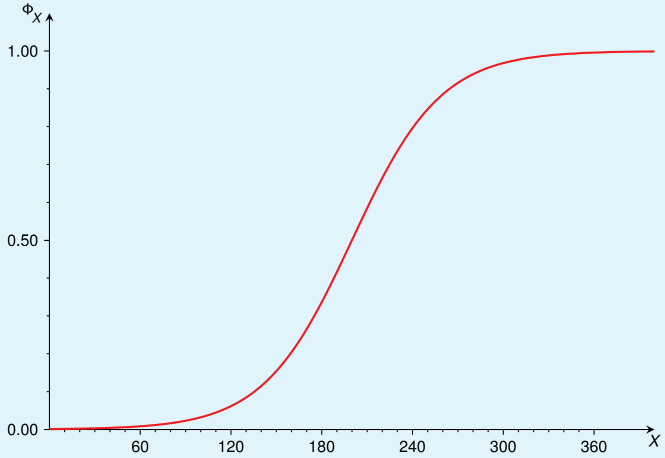
  $$\frac{x_{\text{min}} + x_{\text{max}}}{2},$$

  donde x_{min} y x_{max} son los valores mínimos y máximos, respectivamente, que cumplen $\Phi_X(x) = r$.

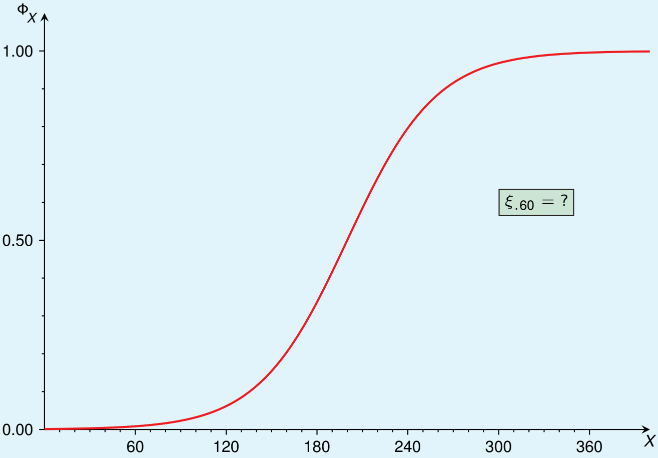
- Caso 2: No existe un valor x, para el cual $\Phi_X(x) = r$
  El cuantil de orden r es el valor x más pequeño que cumple $\Phi_X(x) > r$.

Notación: ξ_r

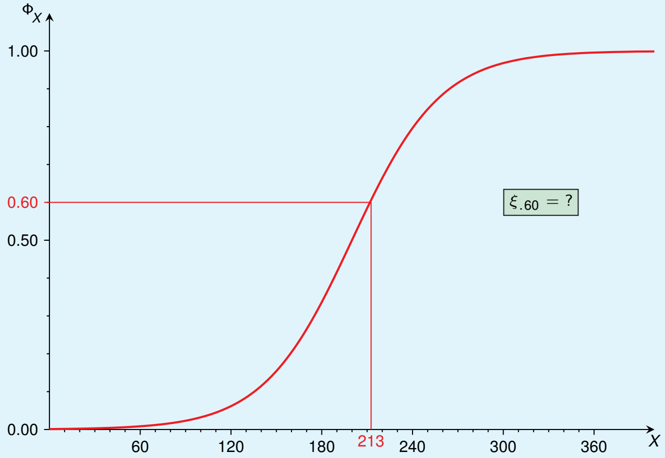
Hallar cuantiles gráficamente



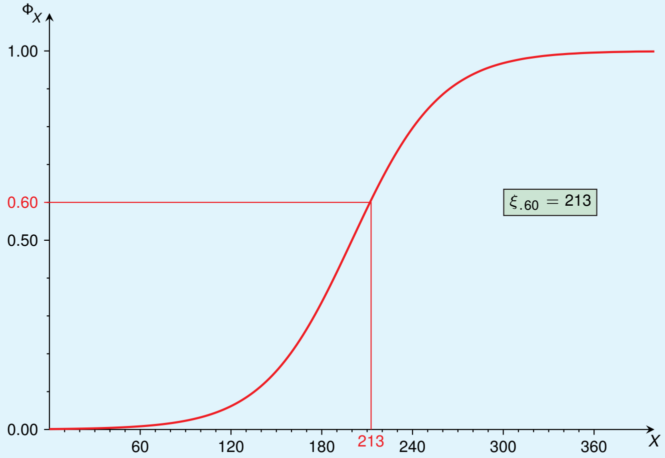
Hallar cuantiles gráficamente



Hallar cuantiles gráficamente



Hallar cuantiles gráficamente



Índice

Datos bivariados

- Datos bivariados $\rightarrow$ Disponemos de los valores en dos variables para cada una de las *n* observaciones

De forma abstracta

	X	Y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	y_4
$\vdots$	$\vdots$	$\vdots$
i	x_i	y_i
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

Datos bivariados

- Datos bivariados $\rightarrow$ Disponemos de los valores en dos variables para cada una de las n observaciones

Ejemplo

	Nivel educativo	**Fumador**
1	Secundaria	No fuma
2	Bachillerato	No fuma
3	Universitario	Poco
4	Secundaria	No fuma
$\vdots$	$\vdots$	$\vdots$
i	Primaria	Intensivo
$\vdots$	$\vdots$	$\vdots$
50	Secundaria	No fuma

Índice

Función de frecuencia y proporción bivariada

Representación tabular

Se puede presentar la función de frecuencia o proporción bivariada en una tabla de frecuencias y proporciones:

Frecuencia y proporción bivariada para *Nivel Educativo* (X) y *Fumador* (Y)

j	k	x_j	y_k	$frec_{XY}(x_j, y_k)$	$p_{XY}(x_j, y_k)$
1	1	Primaria	No fuma	2	.04
1	2	Primaria	Poco	1	.02
1	3	Primaria	Intensivo	1	.02
2	1	Secundaria	No fuma	11	.22
2	2	Secundaria	Poco	2	.04
2	3	Secundaria	Intensivo	5	.10
3	1	Bachillerato	No fuma	8	.16
3	2	Bachillerato	Poco	5	.10
3	3	Bachillerato	Intensivo	3	.06
4	1	Universitario	No fuma	11	.22
4	2	Universitario	Poco	1	.02
4	3	Universitario	Intensivo	0	.00
				50	1.00

Función de frecuencia y proporción bivariada

Tabla de contingencia

Pero es mejor presentar las frecuencias o proporciones en una tabla de contingencia:

Tabla de contingencia para *Nivel Educativo* y *Fumador*

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Función de frecuencia y proporción bivariada

Tabla de contingencia

Pero es mejor presentar las frecuencias o proporciones en una tabla de contingencia:

Tabla de contingencia para *Nivel Educativo* y *Fumador*

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Frecuencias marginales

Función de frecuencia y proporción bivariada

Tabla de contingencia

Pero es mejor presentar las frecuencias o proporciones en una tabla de contingencia:

Tabla de contingencia para *Nivel Educativo* y *Fumador*

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Gran total

Función de frecuencia y proporción bivariada

Tabla de contingencia

Pero es mejor presentar las frecuencias o proporciones en una tabla de contingencia:

Tabla de contingencia para *Nivel Educativo* y *Fumador*

	Nivel Educativo				
Fumador	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	.04	.22	.16	.22	.64
Poco	.02	.04	.10	.02	.18
Intensivo	.02	.10	.06	.00	.18
Total	.08	.36	.32	.24	1.00

Distribución condicional

Definición

Definición

La función de proporción condicional, $p_{X|Y}(x_j|y_k)$, indica cuál es la proporción de observaciones que tienen el valor x_j en la variable X, **solo considerando la submuestra de observaciones con el valor y_k en la variable Y**.

Para todos los x_j y y_k:

$$\boxed{p_{X|Y}(x_j|y_k) = \frac{frec_{XY}(x_j, y_k)}{frec_Y(y_k)}}$$

Distribución condicional

Definición

Definición

La función de proporción condicional, $p_{X|Y}(x_j|y_k)$, indica cuál es la proporción de observaciones que tienen el valor x_j en la variable X, **solo considerando la submuestra de observaciones con el valor y_k en la variable Y**.

Para todos los x_j y y_k:

$$\boxed{p_{X|Y}(x_j|y_k) = \frac{frec_{XY}(x_j, y_k)}{frec_Y(y_k)}}$$

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma					
Poco					
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{32} = .063$				
Poco					
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{32} = .063$	$\frac{11}{32} = .344$	$\frac{8}{32} = .250$	$\frac{11}{32} = .344$	$\frac{32}{32} = 1.00$
Poco					
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

Fumador (X)	*Nivel Educativo* (Y) Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y\|X}$

Fumador (X)	*Nivel Educativo* (Y) Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{32} = .063$	$\frac{11}{32} = .344$	$\frac{8}{32} = .250$	$\frac{11}{32} = .344$	$\frac{32}{32} = 1.00$
Poco	$\frac{1}{9} = .111$	$\frac{2}{9} = .222$	$\frac{5}{9} = .555$	$\frac{1}{9} = .111$	$\frac{9}{9} = 1.00$
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{32} = .063$	$\frac{11}{32} = .344$	$\frac{8}{32} = .250$	$\frac{11}{32} = .344$	$\frac{32}{32} = 1.00$
Poco	$\frac{1}{9} = .111$	$\frac{2}{9} = .222$	$\frac{5}{9} = .555$	$\frac{1}{9} = .111$	$\frac{9}{9} = 1.00$
Intensivo	$\frac{1}{9} = .111$	$\frac{5}{9} = .555$	$\frac{3}{9} = .333$	$\frac{0}{9} = .000$	$\frac{9}{9} = 1.00$
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{32} = .063$	$\frac{11}{32} = .344$	$\frac{8}{32} = .250$	$\frac{11}{32} = .344$	$\frac{32}{32} = 1.00$
Poco	$\frac{1}{9} = .111$	$\frac{2}{9} = .222$	$\frac{5}{9} = .555$	$\frac{1}{9} = .111$	$\frac{9}{9} = 1.00$
Intensivo	$\frac{1}{9} = .111$	$\frac{5}{9} = .555$	$\frac{3}{9} = .333$	$\frac{0}{9} = .000$	$\frac{9}{9} = 1.00$
p_Y	$\frac{4}{50} = .080$	$\frac{18}{50} = .360$	$\frac{16}{50} = .320$	$\frac{12}{50} = .240$	$\frac{50}{50} = 1.00$

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma					
Poco					
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

Fumador (X)	*Nivel Educativo* (Y) Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

Fumador (X)	*Nivel Educativo* (Y) Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$				
Poco					
Intensivo					
p_Y					

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$				
Poco	$\frac{1}{4} = .250$				
Intensivo	$\frac{1}{4} = .250$				
p_Y	$\frac{4}{4} = 1.000$				

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X|Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$	$\frac{11}{18} = .611$			
Poco	$\frac{1}{4} = .250$	$\frac{2}{18} = .111$			
Intensivo	$\frac{1}{4} = .250$	$\frac{5}{18} = .278$			
p_Y	$\frac{4}{4} = 1.000$	$\frac{18}{18} = 1.000$			

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$	$\frac{11}{18} = .611$	$\frac{8}{16} = .500$		
Poco	$\frac{1}{4} = .250$	$\frac{2}{18} = .111$	$\frac{5}{16} = .313$		
Intensivo	$\frac{1}{4} = .250$	$\frac{5}{18} = .278$	$\frac{3}{16} = .188$		
p_Y	$\frac{4}{4} = 1.000$	$\frac{18}{18} = 1.000$	$\frac{16}{16} = 1.000$		

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\,|\,Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$	$\frac{11}{18} = .611$	$\frac{8}{16} = .500$	$\frac{11}{12} = .917$	
Poco	$\frac{1}{4} = .250$	$\frac{2}{18} = .111$	$\frac{5}{16} = .313$	$\frac{1}{12} = .083$	
Intensivo	$\frac{1}{4} = .250$	$\frac{5}{18} = .278$	$\frac{3}{16} = .188$	$\frac{0}{12} = .000$	
p_Y	$\frac{4}{4} = 1.000$	$\frac{18}{18} = 1.000$	$\frac{16}{16} = 1.000$	$\frac{12}{12} = 1.000$	

Distribución condicional

Ejemplo

Tabla de contingencia: Distribución de frecuencia bivariada

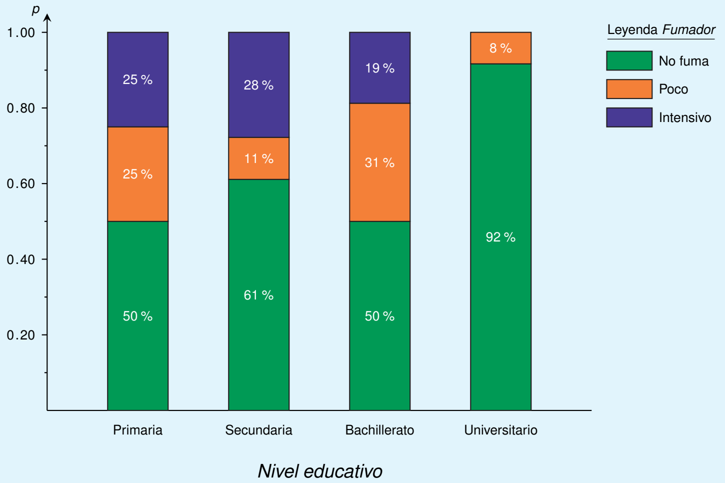
	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	2	11	8	11	32
Poco	1	2	5	1	9
Intensivo	1	5	3	0	9
Total	4	18	16	12	50

Tabla de contingencia: Distribución condicional $p_{X\,|\,Y}$

	Nivel Educativo (Y)				
Fumador (X)	Primaria	Secundaria	Bachillerato	Universitario	**Total**
No fuma	$\frac{2}{4} = .500$	$\frac{11}{18} = .611$	$\frac{8}{16} = .500$	$\frac{11}{12} = .917$	$\frac{32}{50} = .64$
Poco	$\frac{1}{4} = .250$	$\frac{2}{18} = .111$	$\frac{5}{16} = .313$	$\frac{1}{12} = .083$	$\frac{9}{50} = .18$
Intensivo	$\frac{1}{4} = .250$	$\frac{5}{18} = .278$	$\frac{3}{16} = .188$	$\frac{0}{12} = .000$	$\frac{9}{50} = .18$
p_Y	$\frac{4}{4} = 1.000$	$\frac{18}{18} = 1.000$	$\frac{16}{16} = 1.000$	$\frac{12}{12} = 1.000$	$\frac{50}{50} = 1.00$

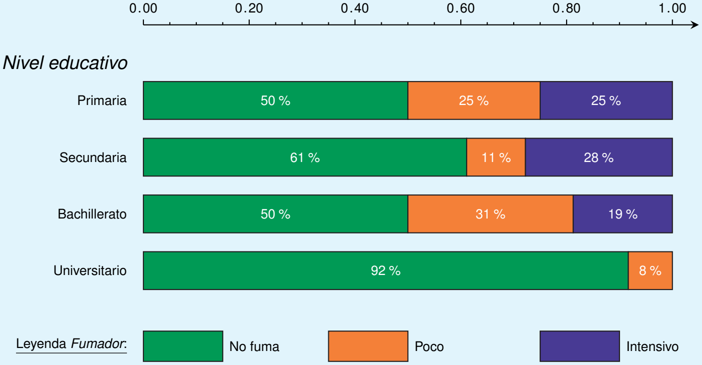
Función de proporción condicional

Representaciones gráficas: Diagrama de rectángulos partidos



Función de proporción condicional

Representaciones gráficas: Diagrama de rectángulos partidos



Índice

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma					
Poco					
Intensivo					
p_Y					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	$\frac{35}{84} = .417$	$\frac{28}{84} = .333$	$\frac{14}{84} = .167$	$\frac{7}{84} = .083$	$\frac{84}{84} = 1.00$
Poco					
Intensivo					
p_Y					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	$\frac{35}{84} = .417$	$\frac{28}{84} = .333$	$\frac{14}{84} = .167$	$\frac{7}{84} = .083$	$\frac{84}{84} = 1.00$
Poco	$\frac{10}{24} = .417$	$\frac{8}{24} = .333$	$\frac{4}{24} = .167$	$\frac{2}{24} = .083$	$\frac{24}{24} = 1.00$
Intensivo					
p_Y					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	$\frac{35}{84} = .417$	$\frac{28}{84} = .333$	$\frac{14}{84} = .167$	$\frac{7}{84} = .083$	$\frac{84}{84} = 1.00$
Poco	$\frac{10}{24} = .417$	$\frac{8}{24} = .333$	$\frac{4}{24} = .167$	$\frac{2}{24} = .083$	$\frac{24}{24} = 1.00$
Intensivo	$\frac{5}{12} = .417$	$\frac{4}{12} = .333$	$\frac{2}{12} = .167$	$\frac{1}{12} = .083$	$\frac{12}{12} = 1.00$
p_Y					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{Y|X}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	$\frac{35}{84} = .417$	$\frac{28}{84} = .333$	$\frac{14}{84} = .167$	$\frac{7}{84} = .083$	$\frac{84}{84} = 1.00$
Poco	$\frac{10}{24} = .417$	$\frac{8}{24} = .333$	$\frac{4}{24} = .167$	$\frac{2}{24} = .083$	$\frac{24}{24} = 1.00$
Intensivo	$\frac{5}{12} = .417$	$\frac{4}{12} = .333$	$\frac{2}{12} = .167$	$\frac{1}{12} = .083$	$\frac{12}{12} = 1.00$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.00$

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X|Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma					
Poco					
Intensivo					
Total					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X|Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .70$				
Poco	$\frac{10}{50} = .20$				
Intensivo	$\frac{5}{50} = .10$				
Total	$\frac{50}{50} = 1.00$				

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .70$	$\frac{28}{40} = .70$			
Poco	$\frac{10}{50} = .20$	$\frac{8}{40} = .20$			
Intensivo	$\frac{5}{50} = .10$	$\frac{4}{40} = .10$			
Total	$\frac{50}{50} = 1.00$	$\frac{40}{40} = 1.00$			

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X|Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .70$	$\frac{28}{40} = .70$	$\frac{14}{20} = .70$		
Poco	$\frac{10}{50} = .20$	$\frac{8}{40} = .20$	$\frac{4}{20} = .20$		
Intensivo	$\frac{5}{50} = .10$	$\frac{4}{40} = .10$	$\frac{2}{20} = .10$		
Total	$\frac{50}{50} = 1.00$	$\frac{40}{40} = 1.00$	$\frac{20}{20} = 1.00$		

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X|Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .70$	$\frac{28}{40} = .70$	$\frac{14}{20} = .70$	$\frac{7}{10} = .70$	
Poco	$\frac{10}{50} = .20$	$\frac{8}{40} = .20$	$\frac{4}{20} = .20$	$\frac{2}{10} = .20$	
Intensivo	$\frac{5}{50} = .10$	$\frac{4}{40} = .10$	$\frac{2}{20} = .10$	$\frac{1}{10} = .10$	
Total	$\frac{50}{50} = 1.00$	$\frac{40}{40} = 1.00$	$\frac{20}{20} = 1.00$	$\frac{10}{10} = 1.00$	

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución condicional $p_{X\mid Y}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{50} = .70$	$\frac{28}{40} = .70$	$\frac{14}{20} = .70$	$\frac{7}{10} = .70$	$\frac{84}{120} = .70$
Poco	$\frac{10}{50} = .20$	$\frac{8}{40} = .20$	$\frac{4}{20} = .20$	$\frac{2}{10} = .20$	$\frac{24}{120} = .20$
Intensivo	$\frac{5}{50} = .10$	$\frac{4}{40} = .10$	$\frac{2}{20} = .10$	$\frac{1}{10} = .10$	$\frac{12}{120} = .10$
Total	$\frac{50}{50} = 1.00$	$\frac{40}{40} = 1.00$	$\frac{20}{20} = 1.00$	$\frac{10}{10} = 1.00$	$\frac{120}{120} = 1.00$

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución de proporción bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma					
Poco					
Intensivo					
p_Y					

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución de proporción bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{120} = .292$	$\frac{28}{120} = .233$	$\frac{14}{120} = .117$	$\frac{7}{120} = .058$	$\frac{84}{120} = .70$
Poco	$\frac{10}{120} = .083$	$\frac{8}{120} = .067$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{24}{120} = .20$
Intensivo	$\frac{5}{120} = .042$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{1}{120} = .008$	$\frac{12}{120} = .10$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.00$

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución de proporción bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{120} = .292$	$\frac{28}{120} = .233$	$\frac{14}{120} = .117$	$\frac{7}{120} = .058$	$\frac{84}{120} = .70$
Poco	$\frac{10}{120} = .083$	$\frac{8}{120} = .067$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{24}{120} = .20$
Intensivo	$\frac{5}{120} = .042$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{1}{120} = .008$	$\frac{12}{120} = .10$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.00$

$$\frac{84}{120} \times \frac{50}{120} = \frac{35}{120}$$

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución de proporción bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{120} = .292$	$\frac{28}{120} = .233$	$\frac{14}{120} = .117$	$\frac{7}{120} = .058$	$\frac{84}{120} = .70$
Poco	$\frac{10}{120} = .083$	$\frac{8}{120} = .067$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{24}{120} = .20$
Intensivo	$\frac{5}{120} = .042$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{1}{120} = .008$	$\frac{12}{120} = .10$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.00$

$$\frac{12}{120} \times \frac{40}{120} = \frac{4}{120}$$

Independencia de variables

Ejemplo

Datos de 120 personas sobre su grupo sanguíneo y conducta fumadora:

Tabla de contingencia: Distribución de frecuencia bivariada $frec_{XY}$

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	**Total**
No fuma	35	28	14	7	84
Poco	10	8	4	2	24
Intensivo	5	4	2	1	12
Total	50	40	20	10	120

Tabla de contingencia: Distribución de proporción bivariada p_{XY}

	Grupo Sanguíneo (Y)				
Fumador (X)	O	A	B	AB	p_X
No fuma	$\frac{35}{120} = .292$	$\frac{28}{120} = .233$	$\frac{14}{120} = .117$	$\frac{7}{120} = .058$	$\frac{84}{120} = .70$
Poco	$\frac{10}{120} = .083$	$\frac{8}{120} = .067$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{24}{120} = .20$
Intensivo	$\frac{5}{120} = .042$	$\frac{4}{120} = .033$	$\frac{2}{120} = .017$	$\frac{1}{120} = .008$	$\frac{12}{120} = .10$
p_Y	$\frac{50}{120} = .417$	$\frac{40}{120} = .333$	$\frac{20}{120} = .167$	$\frac{10}{120} = .083$	$\frac{120}{120} = 1.00$

$$\frac{24}{120} \times \frac{10}{120} = \frac{2}{120}$$

Independencia de variables

Definición

Definición

Dos variables X y Y son *independientes* si, y sólo si, se cumple cualquiera de las siguientes tres condiciones:

(a) $p_{X|Y}(x|y) = p_X(x)$ para todos los pares (x, y) con $frec_Y(y) > 0$.

(b) $p_{Y|X}(y|x) = p_Y(y)$ para todos los pares (x, y) con $frec_X(x) > 0$.

(c) $p_{XY}(x, y) = p_X(x)\, p_Y(y)$ para todos los pares (x, y).

Independencia de variables

Definición

Definición

Dos variables X y Y son *independientes* si, y sólo si, se cumple cualquiera de las siguientes tres condiciones:

(a) $p_{X|Y}(x|y) = p_X(x)$ para todos los pares (x, y) con $frec_Y(y) > 0$.

(b) $p_{Y|X}(y|x) = p_Y(y)$ para todos los pares (x, y) con $frec_X(x) > 0$.

(c) $p_{XY}(x, y) = p_X(x)\, p_Y(y)$ para todos los pares (x, y).

No son condiciones diferentes; ¡son **equivalentes**!
Si una se cumple, entonces las otras tambien.

Independencia de variables

Definición

La definición anterior se generaliza para independencia de m ($m \geqslant 2$) variables:

Definición

Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$p_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = p_{X_1}(x_1)\, p_{X_2}(x_2)\ \ldots\ p_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$

Independencia de variables
Definición

Las definiciones anteriores aplican también a variables en distribuciones teóricas:

Definición

Para variables discretas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\pi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \pi_{X_1}(x_1)\, \pi_{X_2}(x_2)\, \ldots\, \pi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Definición

Para variables continuas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\varphi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \varphi_{X_1}(x_1)\, \varphi_{X_2}(x_2)\, \ldots\, \varphi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Independencia de variables

Definición

Las definiciones anteriores aplican también a variables en distribuciones teóricas:

Definición

Para variables discretas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\pi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \pi_{X_1}(x_1)\, \pi_{X_2}(x_2) \ \ldots \ \pi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Definición

Para variables continuas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\varphi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \varphi_{X_1}(x_1)\, \varphi_{X_2}(x_2) \ \ldots \ \varphi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Independencia de variables

Definición

Las definiciones anteriores aplican también a variables en distribuciones teóricas:

Definición

Para variables discretas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\pi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \pi_{X_1}(x_1)\, \pi_{X_2}(x_2) \; \ldots \; \pi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Definición

Para variables continuas:
Las variables $X_1, X_2, \ldots X_m$ son *independientes* si, y sólo si:

$$\varphi_{X_1 X_2 \ldots X_m}(x_1, x_2, \ldots, x_m) = \varphi_{X_1}(x_1)\, \varphi_{X_2}(x_2) \; \ldots \; \varphi_{X_m}(x_m)$$

para todos los vectores $(x_1, x_2, \ldots, x_m)$.

Índice

Representación gráfica de variables cuantitativas

Diagrama de dispersión

Si las dos variables son cuantitativas, es común representar la asociación gráficamente en un diagrama de dispersión:

Datos bivariados

i	*IMC*	*PAS*
1	21	109
2	25	129
3	27	128
4	20	112
5	33	129
6	23	121
7	24	116
8	28	177
9	23	128
10	27	122
11	22	164
12	26	114
13	29	133
14	24	111
⋮	⋮	⋮
49	26	114
50	26	121

Representación gráfica de variables cuantitativas

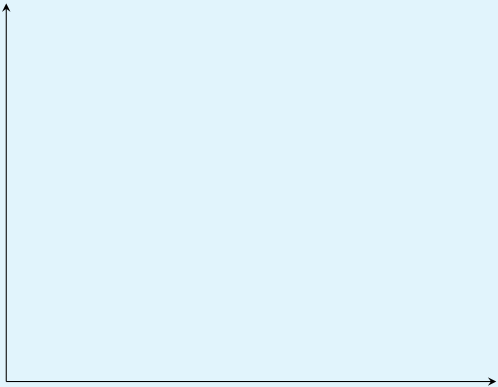
Diagrama de dispersión

Si las dos variables son cuantitativas, es común representar la asociación gráficamente en un diagrama de dispersión:

Datos bivariados

i	IMC	PAS
1	21	109
2	25	129
3	27	128
4	20	112
5	33	129
6	23	121
7	24	116
8	28	177
9	23	128
10	27	122
11	22	164
12	26	114
13	29	133
14	24	111
⋮	⋮	⋮
49	26	114
50	26	121

Diagrama de dispersión



Representación gráfica de variables cuantitativas

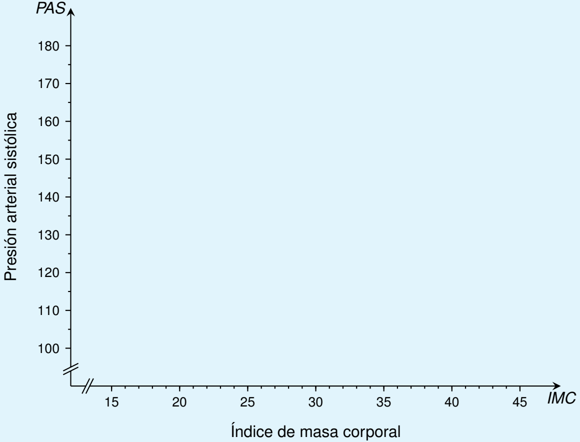
Diagrama de dispersión

Si las dos variables son cuantitativas, es común representar la asociación gráficamente en un diagrama de dispersión:

Datos bivariados

i	*IMC*	*PAS*
1	21	109
2	25	129
3	27	128
4	20	112
5	33	129
6	23	121
7	24	116
8	28	177
9	23	128
10	27	122
11	22	164
12	26	114
13	29	133
14	24	111
⋮	⋮	⋮
49	26	114
50	26	121

Diagrama de dispersión



Representación gráfica de variables cuantitativas

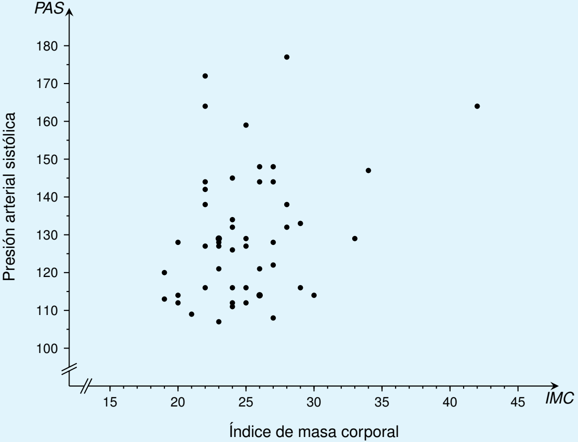
Diagrama de dispersión

Si las dos variables son cuantitativas, es común representar la asociación gráficamente en un diagrama de dispersión:

Datos bivariados

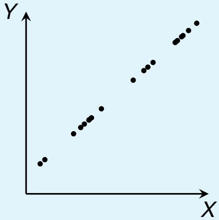
i	IMC	PAS
1	21	109
2	25	129
3	27	128
4	20	112
5	33	129
6	23	121
7	24	116
8	28	177
9	23	128
10	27	122
11	22	164
12	26	114
13	29	133
14	24	111
⋮	⋮	⋮
49	26	114
50	26	121

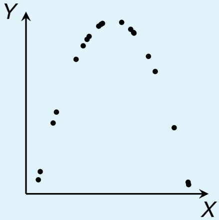
Diagrama de dispersión

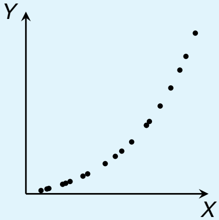


Covariación entre variables cuantitativas

Algunos ejemplos

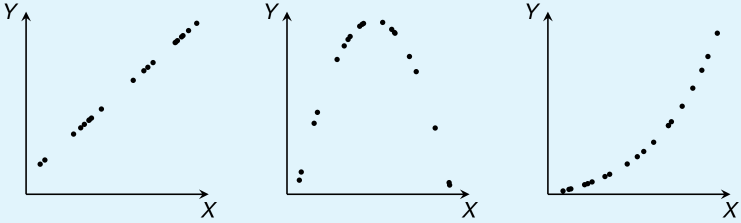






Covariación entre variables cuantitativas

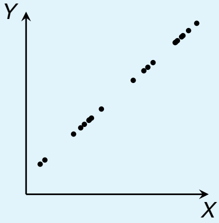
Algunos ejemplos



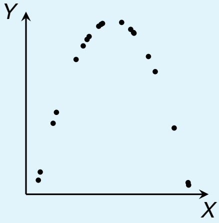
Covariación lineal

Covariación entre variables cuantitativas

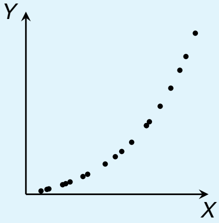
Algunos ejemplos



Covariación lineal



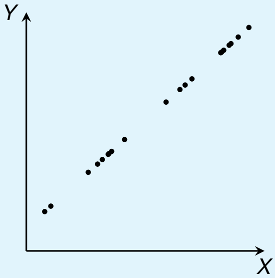
Covariación cuadrática

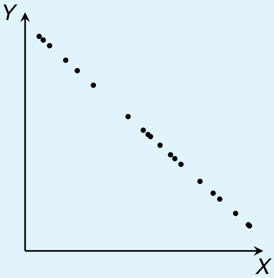


Covariación exponencial

Covariación entre variables cuantitativas

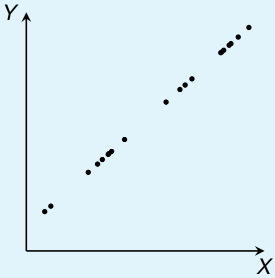
Algunos ejemplos



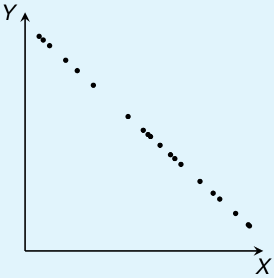


Covariación entre variables cuantitativas

Algunos ejemplos



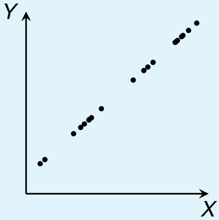
Covariación
lineal directa

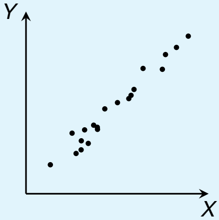


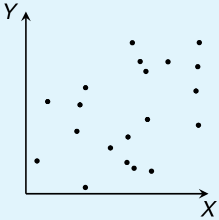
Covariación
lineal inversa

Covariación entre variables cuantitativas

Algunos ejemplos

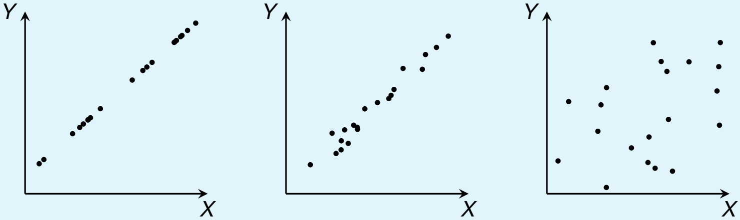






Covariación entre variables cuantitativas

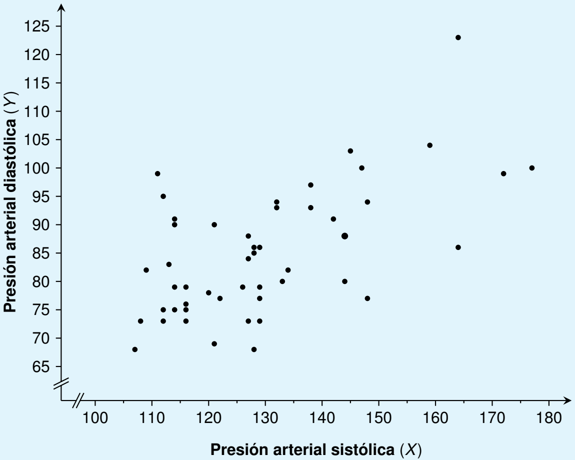
Algunos ejemplos



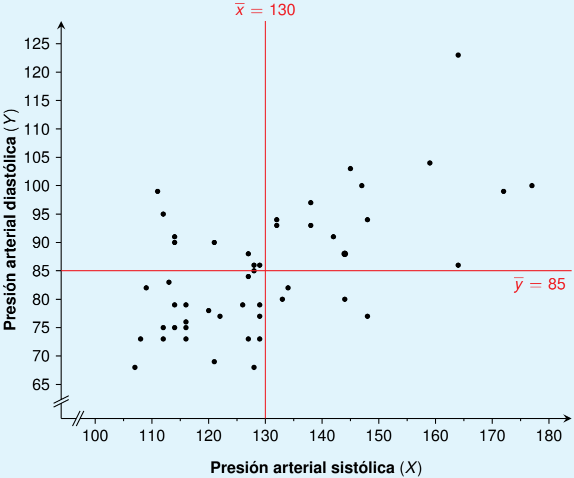
Covariación lineal perfect | Covariación lineal fuerte | Covariación lineal débil

Índice

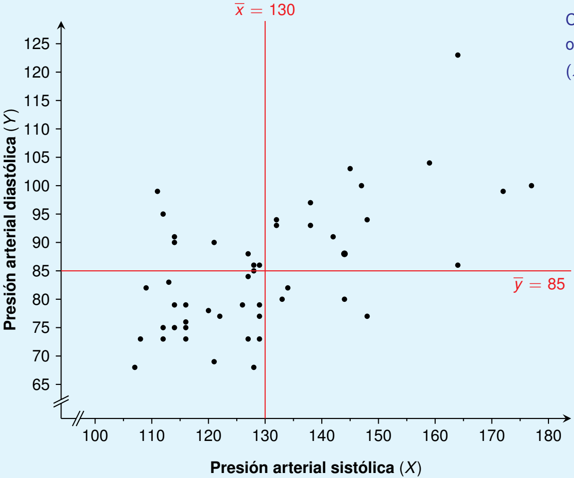
Covarianza: Ejemplo introductorio



Covarianza: Ejemplo introductorio



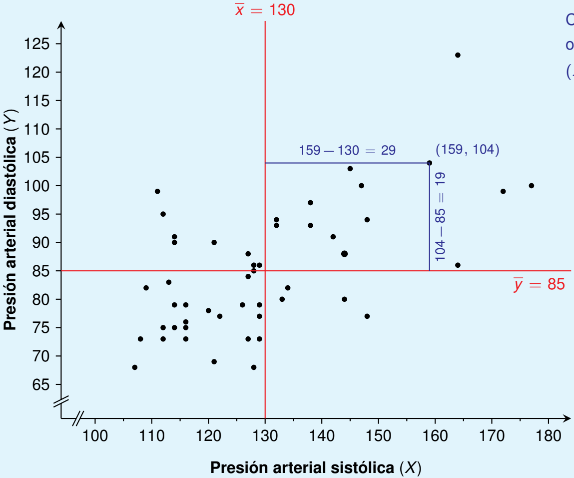
Covarianza: Ejemplo introductorio



Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

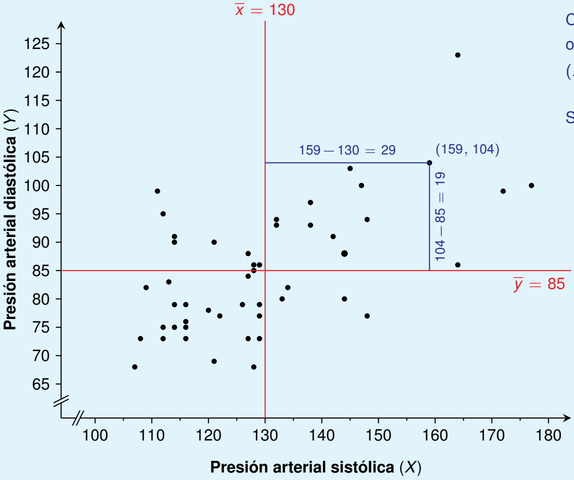
Covarianza: Ejemplo introductorio



Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Covarianza: Ejemplo introductorio



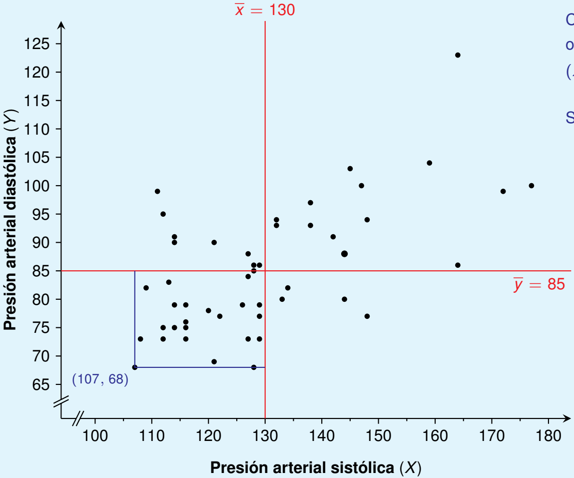
Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I: $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$

Covarianza: Ejemplo introductorio



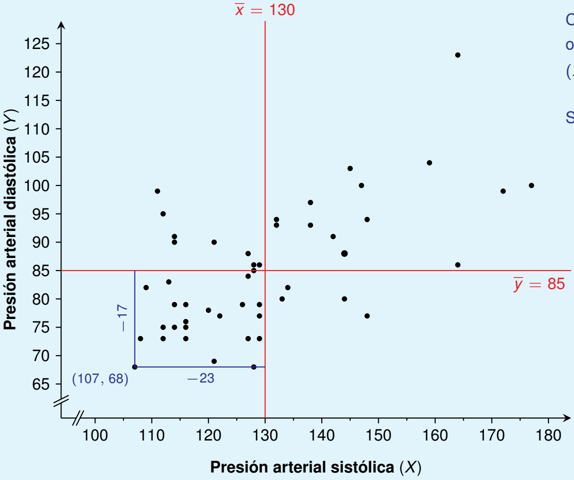
Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I:
  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$

Covarianza: Ejemplo introductorio



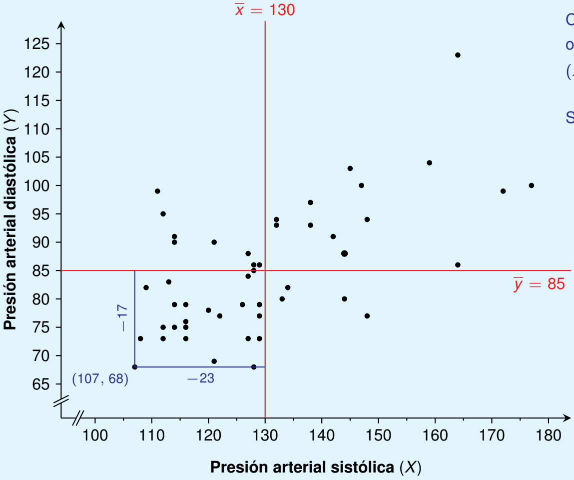
Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I: $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$

Covarianza: Ejemplo introductorio



Calculemos para cada observación i:

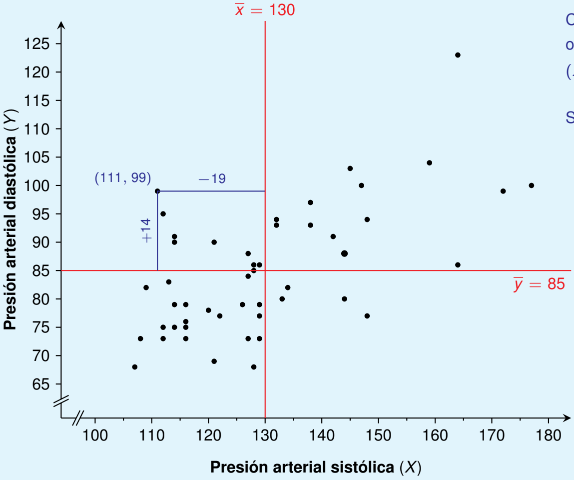
$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:

  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$

Covarianza: Ejemplo introductorio

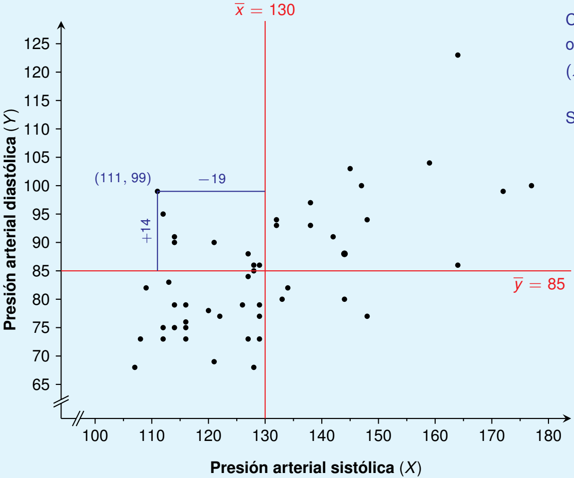


Calculemos para cada observación i:
$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:
  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$
- en el cuadrante II

Covarianza: Ejemplo introductorio



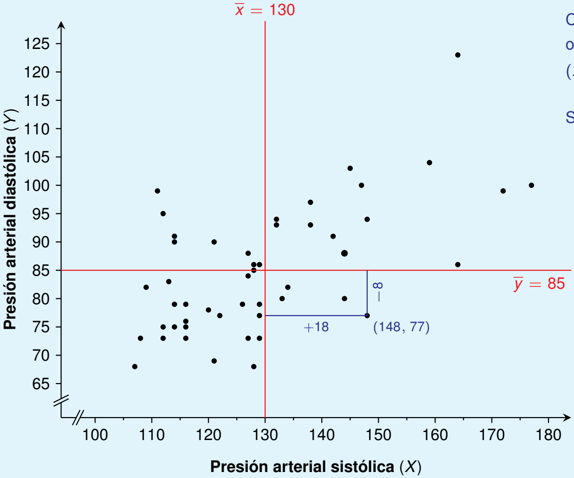
Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III: $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$
- en el cuadrante II $(x_i - \overline{x}) \times (y_i - \overline{y}) < 0$

Covarianza: Ejemplo introductorio

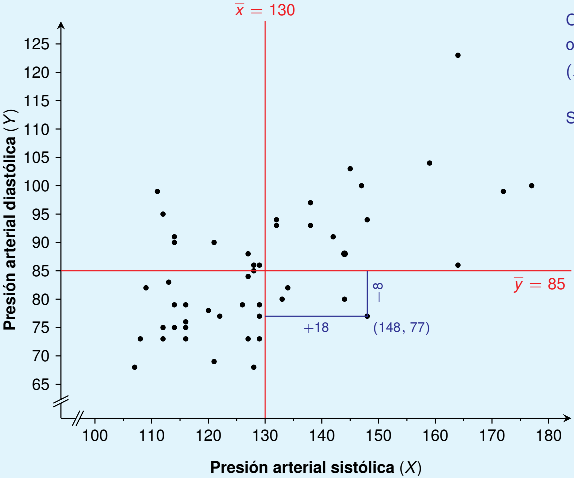


Calculemos para cada observación i:
$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:
  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$
- en el cuadrante II
  $(x_i - \overline{x}) \times (y_i - \overline{y}) < 0$

Covarianza: Ejemplo introductorio



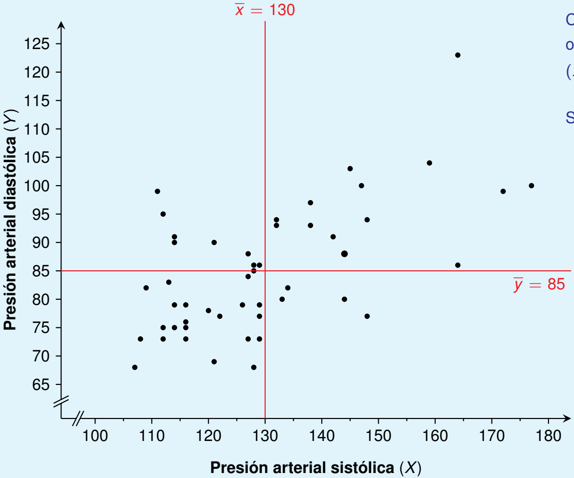
Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:
  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$
- en el cuadrante II y IV
  $(x_i - \overline{x}) \times (y_i - \overline{y}) < 0$

Covarianza: Ejemplo introductorio



Calculemos para cada observación i:

$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:

  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$

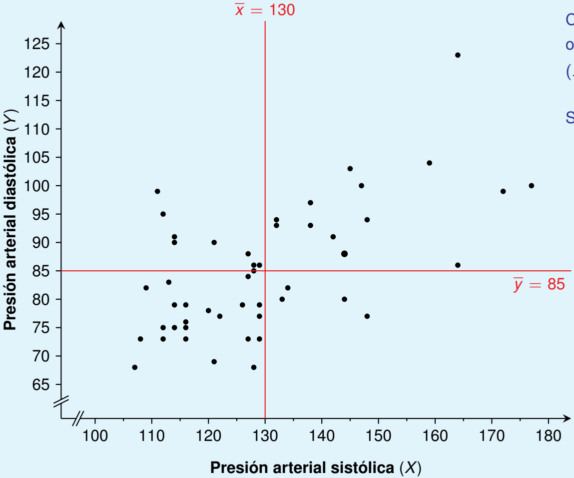
- en el cuadrante II y IV

  $(x_i - \overline{x}) \times (y_i - \overline{y}) < 0$

Covarianza entre X y Y:

$$\boxed{s_{XY} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})}{n}}$$

Covarianza: Ejemplo introductorio



Calculemos para cada observación i:
$(x_i - \overline{x}) \times (y_i - \overline{y})$

Si i está

- en el cuadrante I y III:
  $(x_i - \overline{x}) \times (y_i - \overline{y}) > 0$
- en el cuadrante II y IV
  $(x_i - \overline{x}) \times (y_i - \overline{y}) < 0$

Covarianza entre X y Y:

$$\boxed{s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n}}$$

Para estos datos:

$$s_{XY} = \frac{5610}{50} = 112.2$$

Covarianza: Definición

Definición

Para dos variables X y Y, que tienen valores en n observaciones, la covarianza, representada por s_{XY}, se define como:

$$s_{XY} = \frac{\sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y})]}{n}.$$

Se puede interpretar la covarianza como la media del producto de las puntuaciones diferenciales en ambas variables.

Covarianza: Definición

Definición

Para dos variables X y Y, que tienen valores en n observaciones, la covarianza, representada por s_{XY}, se define como:

$$s_{XY} = \frac{\sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y})]}{n}.$$

Se puede interpretar la covarianza como la media del producto de las puntuaciones diferenciales en ambas variables.

Covarianza: Definición

Para distribuciones teóricas, se define la covarianza de forma análoga:

Definición

Para dos variables X y Y con una distribución conjunta de probabilidad (π_{XY}) o densidad (φ_{XY}), la covarianza, representada por σ_{XY}, se define como:

$$\sigma_{XY} = \mathscr{E}\left[(X - \mu_X)(Y - \mu_Y)\right] .$$

Es decir, la covarianza es el valor esperado del producto de las puntuaciones diferenciales en ambas variables.

Covarianza: Definición

Para distribuciones teóricas, se define la covarianza de forma análoga:

Definición

Para dos variables X y Y con una distribución conjunta de probabilidad (π_{XY}) o densidad (φ_{XY}), la covarianza, representada por σ_{XY}, se define como:

$$\sigma_{XY} = \mathscr{E}\left[(X - \mu_X)(Y - \mu_Y)\right].$$

Es decir, la covarianza es el valor esperado del producto de las puntuaciones diferenciales en ambas variables.

Covarianza: Propiedades

Propiedad 1

Propiedad 1: Covarianza entre dos variables independientes

Si las variables X y Y son independientes, entonces:

$$s_{XY} =$$

Covarianza: Propiedades

Propiedad 1

Propiedad 1: Covarianza entre dos variables independientes

Si las variables X y Y son independientes, entonces:

$$s_{XY} = 0$$

Covarianza: Propiedades

Propiedad 1

Propiedad 1: Covarianza entre dos variables independientes

Si las variables X y Y son independientes, entonces:

$$s_{XY} = 0$$

¡Ojo! La relación inversa no necesariamente es verdad

La propiedad anterior dice:

$$\text{independencia} \implies \text{covarianza nula}$$

Sin embargo,

$$\text{covarianza nula} \not\Longrightarrow \text{independencia}$$

Covarianza: Propiedades

Propiedad 1

Propiedad 1: Covarianza entre dos variables independientes

Si las variables X y Y son independientes, entonces:

$$s_{XY} = 0$$

¡Ojo! La relación inversa no necesariamente es verdad

La propiedad anterior dice:

$$\text{independencia} \implies \text{independencia lineal}$$

Sin embargo,

$$\text{independencia lineal} \not\Longrightarrow \text{independencia}$$

Covarianza: Propiedades

Propiedad 2

Propiedad 2: Efecto de una transformación lineal en la covarianza

Si la variable T, se obtuvo a través de una transformación lineal de X, es decir, si

$$T = aX + b$$

para constantes a y b cualesquiera,

entonces la covarianza de T con otra variable Y cumple

$$s_{TY} =$$

Covarianza: Propiedades

Propiedad 2

Propiedad 2: Efecto de una transformación lineal en la covarianza

Si la variable T, se obtuvo a través de una transformación lineal de X, es decir, si

$$T = aX + b$$

para constantes a y b cualesquiera,

entonces la covarianza de T con otra variable Y cumple

$$s_{TY} = as_{XY}$$

Covarianza: Propiedades

Propiedad 3

Propiedad 3: Propiedad distributiva de la covarianza respecto de la suma

Si la variable T, se obtuvo por la suma de las variables X y Y, es decir, si

$$T = X + Y$$

entonces la covarianza de T con cualquier otra variable Z cumple

$$s_{TZ} =$$

Covarianza: Propiedades

Propiedad 3

Propiedad 3: Propiedad distributiva de la covarianza respecto de la suma

Si la variable T, se obtuvo por la suma de las variables X y Y, es decir, si

$$T = X + Y$$

entonces la covarianza de T con cualquier otra variable Z cumple

$$s_{TZ} = s_{XZ} + s_{YZ}$$

Varianza de una suma de variables

Si la variable T, se obtuvo por la suma de las variables X y Y, es decir, si

$$T = X + Y$$

entonces la varianza de la variable T se da por:

$$s_T^2 =$$

Varianza de una suma de variables

Si la variable T, se obtuvo por la suma de las variables X y Y, es decir, si

$$T = X + Y$$

entonces la varianza de la variable T se da por:

$$s_T^2 = s_X^2 + s_Y^2 + 2s_{XY}$$

Covarianza: Interpretación

Covarianza: Interpretación

Vimos que multiplar la(s) variable(s) con una constante afecta la covarianza.

Esto implica que la covarianza depende de la “unidad de la escala”

Por ejemplo, si

$$
\begin{array}{llll}
X_1 & : \text{Peso en kilogramos} \qquad & X_2 & : \text{Peso en gramos} \qquad (X_2 = 1000X_1) \\
Y_1 & : \text{Talla en metros} \qquad & Y_2 & : \text{Talla en centimetros} \qquad (Y_2 = 100Y_1),
\end{array}
$$

entonces

$$s_{X_2Y_2} \neq s_{X_1Y_1}$$

Covarianza: Interpretación

Vimos que multiplar la(s) variable(s) con una constante afecta la covarianza.

Esto implica que la covarianza depende de la "unidad de la escala"

Por ejemplo, si

$$\begin{array}{llll} X_1 & : \text{Peso en kilogramos} & X_2 : \text{Peso en gramos} & (X_2 = 1000 X_1) \\ Y_1 & : \text{Talla en metros} & Y_2 : \text{Talla en centimetros} & (Y_2 = 100 Y_1), \end{array}$$

entonces

$$s_{X_2 Y_2} \neq s_{X_1 Y_1} \qquad (s_{X_2 Y_2} = 100000 s_{X_1 Y_1})$$

Covarianza: Interpretación

Vimos que multiplar la(s) variable(s) con una constante afecta la covarianza.

Esto implica que la covarianza depende de la “unidad de la escala”

Por ejemplo, si

X_1 : Peso en kilogramos	X_2 : Peso en gramos	$(X_2 = 1000X_1)$
Y_1 : Talla en metros	Y_2 : Talla en centimetros	$(Y_2 = 100Y_1)$,

entonces

$$s_{X_2Y_2} \neq s_{X_1Y_1} \qquad (s_{X_2Y_2} = 100000 s_{X_1Y_1})$$

$\longrightarrow$ Es difícil interpretar la covarianza

Correlación: Definición

Definición

Para dos variables X y Y, que tienen valores en n observaciones, el coeficiente de correlación, representada por r_{XY}, se define como:

$$r_{XY} = \frac{s_{XY}}{s_X \, s_Y}.$$

Nota

La correlación es definida si y solo si la desviación estándar es diferente de 0 para ambas variables

Correlación: Definición

Definición

Para dos variables X y Y, que tienen valores en n observaciones, el coeficiente de correlación, representada por r_{XY}, se define como:

$$r_{XY} = \frac{s_{XY}}{s_X \, s_Y}.$$

Nota

La correlación es definida si y solo si la desviación estándar es diferente de 0 para ambas variables

Correlación: Definición

Para distribuciones teóricas, se define la correlación de forma análoga:

Definición

Para dos variables X y Y con una distribución conjunta de probabilidad (π_{XY}) o densidad (φ_{XY}), el coeficiente de correlación, representada por ρ_{XY}, se define como:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}.$$

Correlación: Propiedades

Propiedad 1

Propiedad 1: Correlación entre dos variables independientes

Si las variables X y Y (con $s_X > 0$ and $s_Y > 0$) son independientes, entonces:

$$r_{XY} =$$

Correlación: Propiedades

Propiedad 1

Propiedad 1: Correlación entre dos variables independientes

Si las variables X y Y (con $s_X > 0$ and $s_Y > 0$) son independientes, entonces:

$$r_{XY} = 0$$

Correlación: Propiedades

Propiedad 1

Propiedad 1: Correlación entre dos variables independientes

Si las variables X y Y (con $s_X > 0$ and $s_Y > 0$) son independientes, entonces:

$$r_{XY} = 0$$

¡Ojo! La relación inversa no necesariamente es verdad

$$r_{XY} = 0 \quad \not\Rightarrow \quad \text{independencia}$$

Correlación: Propiedades

Propiedad 2

Propiedad 2: Efecto de una transformación lineal en la correlación

Si la variable T, se obtuvo a través de una transformación lineal de X, es decir, si

$$T = aX + b$$

para constantes $a \neq 0$ y b,

entonces la correlación de T con cualquier otra variable Y cumple

$$r_{TY} =$$

Correlación: Propiedades

Propiedad 2

Propiedad 2: Efecto de una transformación lineal en la correlación

Si la variable T, se obtuvo a través de una transformación lineal de X, es decir, si

$$T = aX + b$$

para constantes $a \neq 0$ y b,

entonces la correlación de T con cualquier otra variable Y cumple

$$r_{TY} = \begin{cases} r_{XY} & \text{si } a > 0 \\ -r_{XY} & \text{si } a < 0 \end{cases}$$

Correlación: Propiedades

Propiedad 3

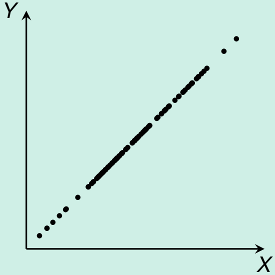
Propiedad 3: Límites de la correlación

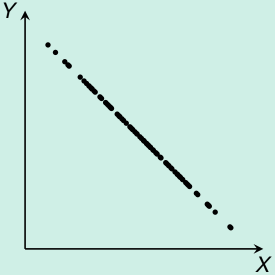
Para variables X y Y (con $s_X > 0$ y $s_Y > 0$), siempre se cumple:

$$-1 \leqslant r_{XY} \leqslant +1$$

Visualizar correlaciones

Ejemplo



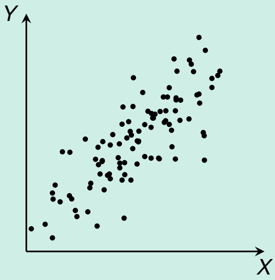


$r_{XY} = +1.00$

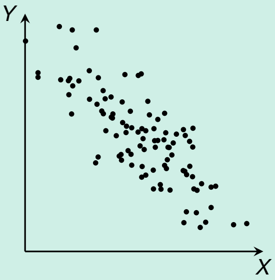
$r_{XY} = -1.00$

Visualizar correlaciones

Ejemplo



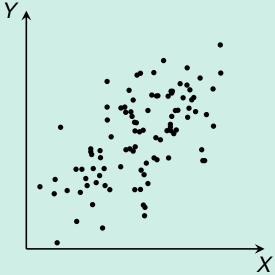
$r_{XY} = +.80$



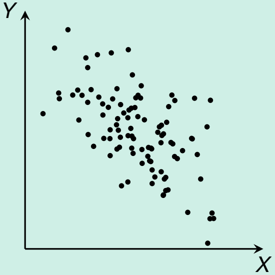
$r_{XY} = -.80$

Visualizar correlaciones

Ejemplo



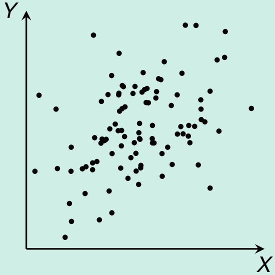
$r_{XY} = +.60$

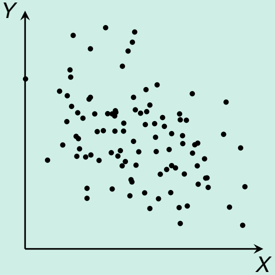


$r_{XY} = -.60$

Visualizar correlaciones

Ejemplo



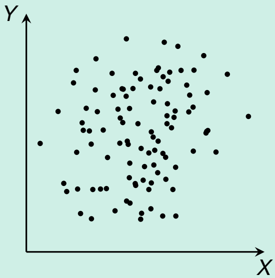


$r_{XY} = +0.40$

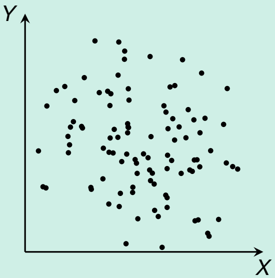
$r_{XY} = -0.40$

Visualizar correlaciones

Ejemplo



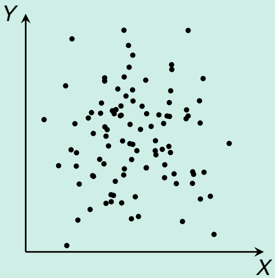
$r_{XY} = +0.20$



$r_{XY} = -0.20$

Visualizar correlaciones

Ejemplo



$r_{XY} = 0.00$

Índice

Índice

Estadística inferencial: Objetivos

- Objetivo general: Hacer inferencias sobre características de un modelo estadístico a partir de (datos observados en) una muestra.
- La especificación de un modelo estadístico $\rightarrow$ función de probabilidad/densidad de las variables aleatorias de interés.
- En muchas ocasiones, se especifica el modelo estadístico parcialmente.

  Por ejemplo:
  Se especifica que las calificaciones en un examen diagnóstico tienen una distribución normal, pero no se dan valores para la media y la varianza de esta distribución:

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \begin{cases} \mu &= ? \\ \sigma^2 &= ? \end{cases}$$

Estadística inferencial: Objetivos

- Objetivo general: Hacer inferencias sobre características de un modelo estadístico a partir de (datos observados en) una muestra.
- La especificación de un modelo estadístico $\rightarrow$ función de probabilidad/densidad de las variables aleatorias de interés.
- En muchas ocasiones, se especifica el modelo estadístico parcialmente.

  Por ejemplo:
  Se especifica que las calificaciones en un examen diagnóstico tienen una distribución normal, pero no se dan valores para la media y la varianza de esta distribución:

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \begin{cases} \mu &= ? \\ \sigma^2 &= ? \end{cases}$$

Estadística inferencial: Objetivos

- Objetivo general: Hacer inferencias sobre características de un modelo estadístico a partir de (datos observados en) una muestra.
- La especificación de un modelo estadístico $\rightarrow$ función de probabilidad/densidad de las variables aleatorias de interés.
- En muchas ocasiones, se especifica el modelo estadístico parcialmente.

  Por ejemplo:
  Se especifica que las calificaciones en un examen diagnóstico tienen una distribución normal, pero no se dan valores para la media y la varianza de esta distribución:

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \begin{cases} \mu & = ? \\ \sigma^2 & = ? \end{cases}$$

Estadística inferencial: Objetivos

Objetivos específicos:

1. Estimación de parámetros
   A partir de los datos de una muestra asignar valores a los parámetros desconocidos en el modelo estadístico.

2. Contraste de hipótesis
   Evaluar, a la luz de los datos observados en una muestra, si el modelo estadístico especificado (la “hipótesis nula”) es razonable/plausible.

Estadística inferencial: Objetivos

Objetivos específicos:

1. Estimación de parámetros
   A partir de los datos de una muestra asignar valores a los parámetros desconocidos en el modelo estadístico.

2. Contraste de hipótesis
   Evaluar, a la luz de los datos observados en una muestra, si el modelo estadístico especificado (la “hipótesis nula”) es razonable/plausible.

Estadística inferencial

Ejemplos

- Unos expertos en educación elaboraron un examen de admisión a la universidad.

  Una pregunta de investigación importante al respecto puede ser:
  ¿Cuál es la probabilidad de que alguien que pasa el examen de admisión apruebe los exámenes del primer semestre de la carrera de psicología?

  Se formaliza en términos de dos variables aleatorias:

$$X = \begin{cases} 1 & \text{si pasa el examen de admisión} \\ 0 & \text{si no pasa el examen de admisión} \end{cases}$$

$$Y = \begin{cases} 1 & \text{aprueba los exámenes del primer semestre} \\ 0 & \text{no aprueba los exámenes del primer semestre} \end{cases}$$

  Se desea conocer la probabilidad $\Pr(Y = 1 | X = 1)$.

  $\rightarrow$ Estimación de parámetros.

Estadística inferencial

Ejemplos

- Un profesor quiere investigar la hipótesis de que las notas de las mujeres en su examen son más altas que las de los hombres.

  Supone que tanto en la población de mujeres como en la población de hombres las notas tienen una distribución normal:

$$\begin{cases} \text{Mujeres:} & X \sim \mathcal{N}(\mu_X, \sigma^2) \\ \text{Hombres:} & Y \sim \mathcal{N}(\mu_Y, \sigma^2) \end{cases}$$

  y contrasta la hipótesis:

$$\mathrm{H}_0 : \ \mu_X = \mu_Y$$

  $\rightarrow$ Contraste de hipótesis.

Índice

Estimación puntual

Supongamos que deseamos estimar un parámetro θ de una distribución teórica/poblacional π_X o φ_X.

¿Cómo encontramos un buen estimador de θ?

¿Qué estadístico utilizar para el parámetro?

Métodos de estimación

Se han propuesto varios métodos. Los más conocidos son:

1. El método de los momentos (o el método por analogía)

   El estimador es el homólogo muestral del parámetro poblacional

2. El método de mínimos cuadrados

   El estimador minimiza las diferencias cuadráticas entre los valores observados en la muestra y los valores derivados del modelo estadístico.

3. El método de máxima verosimilitud

   El estimador maximiza la verosimilitud (la plausibilidad) del parámetro a la luz de los datos observados. El valor que se asigna al parámetro maximiza la probabilidad de los datos observados.

Métodos de estimación

Se han propuesto varios métodos. Los más conocidos son:

1. El método de los momentos (o el método por analogía)

   El estimador es el homólogo muestral del parámetro poblacional

2. El método de mínimos cuadrados

   El estimador minimiza las diferencias cuadráticas entre los valores observados en la muestra y los valores derivados del modelo estadístico.

3. El método de máxima verosimilitud

   El estimador maximiza la verosimilitud (la plausibilidad) del parámetro a la luz de los datos observados. El valor que se asigna al parámetro maximiza la probabilidad de los datos observados.

Métodos de estimación

Se han propuesto varios métodos. Los más conocidos son:

1. El método de los momentos (o el método por analogía)

   El estimador es el homólogo muestral del parámetro poblacional

2. El método de mínimos cuadrados

   El estimador minimiza las diferencias cuadráticas entre los valores observados en la muestra y los valores derivados del modelo estadístico.

3. El método de máxima verosimilitud

   El estimador maximiza la verosimilitud (la plausibilidad) del parámetro a la luz de los datos observados. El valor que se asigna al parámetro maximiza la probabilidad de los datos observados.

Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios.
Los más importantes son:

- Carencia de sesgo
  Un estadístico T es un estimador insesgado de un parámetro θ si y solo si:
  $$\mathscr{E}(T) = \theta$$
  - La media aritmética $\overline{X}$ es un estimador insesgado de la media poblacional μ_X.
  - La varianza muestral S_X^2 no es un estimador insesgado de la varianza poblacional σ_X^2.
  - La cuasivarianza muestral $\tilde{S}_X^2$ sí es un estimador insesgado de la varianza poblacional σ_X^2.

Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios. Los más importantes son:

- Carencia de sesgo
  Un estadístico T es un estimador insesgado de un parámetro θ si y solo si:
  $$\mathscr{E}(T) = \theta$$
  - La media aritmética $\overline{X}$ es un estimador insesgado de la media poblacional μ_X.
  - La varianza muestral S_X^2 no es un estimador insesgado de la varianza poblacional σ_X^2.
  - La cuasivarianza muestral $\tilde{S}_X^2$ sí es un estimador insesgado de la varianza poblacional σ_X^2.

Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios. Los más importantes son:

- Carencia de sesgo
  Un estadístico T es un estimador insesgado de un parámetro θ si y solo si:
  $$\mathscr{E}(T) = \theta$$

  - La media aritmética $\overline{X}$ es un estimador insesgado de la media poblacional μ_X.
  - La varianza muestral S_X^2 no es un estimador insesgado de la varianza poblacional σ_X^2.
  - La cuasivarianza muestral $\tilde{S}_X^2$ sí es un estimador insesgado de la varianza poblacional σ_X^2.

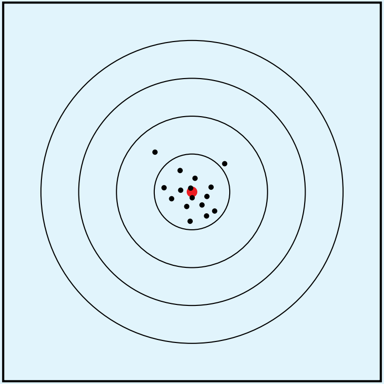
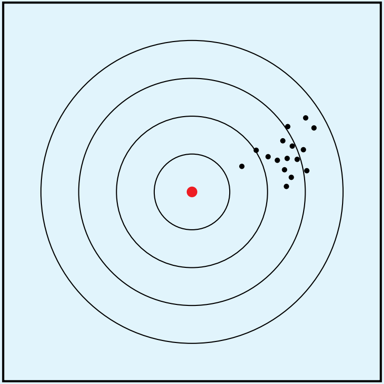
Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios. Los más importantes son:

- Carencia de sesgo
  Un estadístico T es un estimador insesgado de un parámetro θ si y solo si:
  $$\mathscr{E}(T) = \theta$$
  - La media aritmética $\overline{X}$ es un estimador insesgado de la media poblacional μ_X.
  - La varianza muestral S_X^2 no es un estimador insesgado de la varianza poblacional σ_X^2.
  - La cuasivarianza muestral $\tilde{S}_X^2$ sí es un estimador insesgado de la varianza poblacional σ_X^2.

Evaluar la bondad de los estimadores

Carencia de sesgo

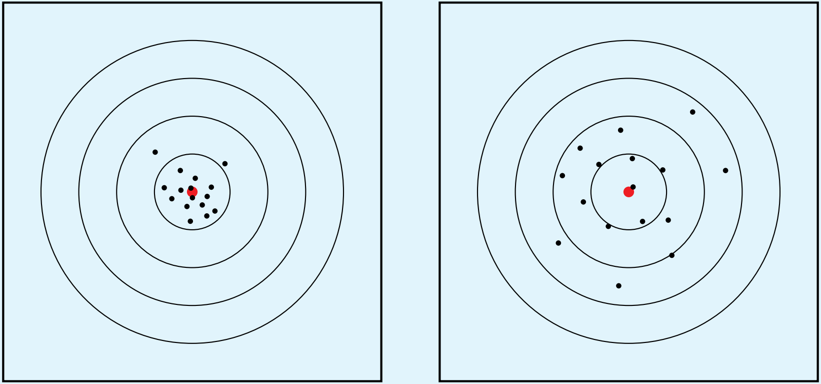
Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios. Los más importantes son:

- Eficiencia
  Un estimador T de un parámetro θ es más eficiente conforme
  $$\mathscr{E}\left[(T-\theta)^2\right]$$
  se acerca a 0.

Evaluar la bondad de los estimadores

Eficiencia



Evaluar la bondad de los estimadores

Para evaluar la bondad de un estimador, se consideran varios criterios.
Los más importantes son:

- Consistencia
  Un estimador es consistente si, conforme el tamaño de la muestra aumenta, tiene mejor eficiencia.

  En otras palabras:
  Un estimador es consistente si muestras más grandes dan mejores estimaciones.

  Un estimador $T_{(n)}$ de un parámetro θ es consistente si para cualquier $\varepsilon > 0$:

$$\lim_{n\to\infty} ¶(\theta - \varepsilon < T_{(n)} < \theta + \varepsilon) = 1.$$

Índice

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *“con cierto nivel de confianza”*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *“con cierto nivel de confianza”*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *"con cierto nivel de confianza"*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *"con cierto nivel de confianza"*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *"con cierto nivel de confianza"*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalos de confianza: Concepto

- Es deseable acompañar la estimación puntual con información sobre el margen de error.
- Se construye un intervalo $[a, b]$ que contiene el parámetro de interés *“con cierto nivel de confianza”*.
- El nivel de confianza se determina a priori.
- El complemento del nivel de confianza se llama el nivel de significancia o el nivel de riesgo.

  Se suele representar el nivel de significancia por α.
  $\rightarrow$ el nivel de confianza es $1 - \alpha$.

  Valores comunes para el nivel de confianza son 95 %, 99 %, 90 %.

Intervalo de confianza: Interpretación

- Interpretaciones erróneas del intervalo de confianza son muy comunes.

  Por ejemplo, no es correcto interpretar el intervalo de confianza de 95 % de $[a, b]$ para un parámetro de interés como:

  "La probabilidad de que el parámetro se encuentra entre a y b es 95 %".

- Interpretación correcta:

  "Si se repitiera el estudio un gran número de veces bajo las mismas condiciones y se calculara cada vez un intervalo de 95 % de confianza,
  entonces para el 95 % de los casos el intervalo de confianza contendría el parámetro de interés."

Intervalo de confianza: Interpretación

- Interpretaciones erróneas del intervalo de confianza son muy comunes.

  Por ejemplo, no es correcto interpretar el intervalo de confianza de 95 % de $[a, b]$ para un parámetro de interés como:

  "La probabilidad de que el parámetro se encuentra entre a y b es 95 %".

- Interpretación correcta:

  "Si se repitiera el estudio un gran número de veces bajo las mismas condiciones y se calculara cada vez un intervalo de 95 % de confianza,
  entonces para el 95 % de los casos el intervalo de confianza contendría el parámetro de interés."

Intervalo de confianza: Interpretación