Relaciones EDIP

Daniel Morán Sánchez
Elías Monge Sánchez
Jesús Muñoz Velasco

2023

Índice

1. Relación 1

1.1. Ejercicio 1:

El número de hijos de las familias de una determinada barriada de una ciudad es una variable estadística de la que se conocen los siguientes datos:

x_i	n_i	N_i	f_i
0	80		0.16
1	110		
2		320	
3			0.18
4	40		
5			
6	20		

n_i: frecuencias absolutas
N_i: frecuencias absolutas acumuladas
f_i: frecuencias relativas

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

a) Completar la tabla de frecuencias

x_i	n_i	N_i	f_i
0	80	**80**	0.16
1	110	**190**	**0.22**
2	**130**	320	**0.26**
3	**90**	**410**	0.18
4	40	**450**	**0.08**
5	**30**	**480**	**0.06**
6	20	**500**	**0.04**

b) Representar la distribución mediante un diagrama de barras y la curva de distribución

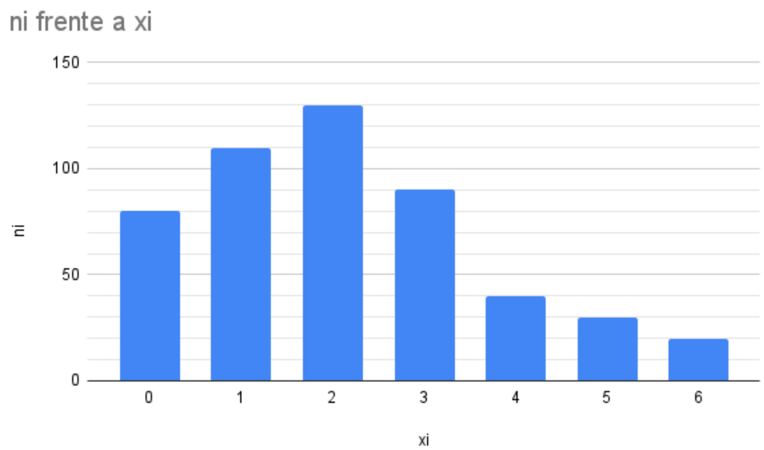
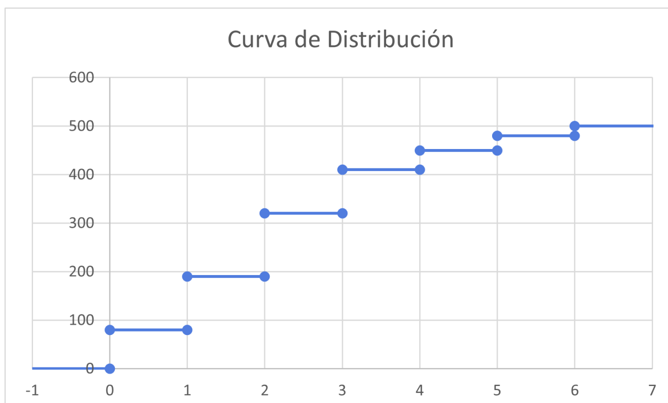


Diagrama de barras



c) Promediar los valores de la variable mediante diferentes medidas. Interpretarlas.

Las medias se obtienen mediante las siguientes fórmulas:

Media aritmética

$\bar{x} = \sum_{i=1}^{k} f_i x_i = 2,14$

La media aritmética nos da una idea bastante acertada de en torno a qué valor medio se agrupan los datos. Mirando la tabla y el diagrama de barras, vemos que en el entorno cercano al valor 2,14 las frecuencias son más altas, mientras que cuanto más nos alejamos más disminuyen las frecuencias.

Media armónica

$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$

Esta media no es posible de calcular, pues existe una modalidad nula que no tiene inverso. Lo único que se podría argumentar aquí es que tomando límite el valor de H es 0.

Media geométrica

$G = \sqrt[n]{\prod_{i=1}^{k} f_i x_i} = 0$

La media geométrica tiene el inconveniente de que, como se ha dado el caso, si alguna modalidad vale 0, su valor es cero, por lo que no aporta ninguna información añadida.

Media cuadrática

$Q = \sqrt{\sum_{i=1}^{k} f_i x_i^2} = 7,1$

La media cuadrática al menos ha dado un valor no nulo, pero no parece guardar mucha relación con la distribución ya que, de hecho, se sale del rango. Esto es porque no es una media adecuada para esta distribución al tener un cuadrado, y es que se usa sobre todo para promediar superficies.

1.2. Ejercicio 2:

La puntuación obtenida por 50 personas que se presentaros a una prueba de selección, sumadas las puntuaciones de los distintos test, fueron:

174, 185, 166, 176, 145, 166, 191, 175, 158, 156, 156, 187, 162, 172, 197, 181, 151, 161, 183, 172, 162, 147, 178, 176, 141, 170, 171, 158, 184, 173, 169, 162, 172, 181, 187, 177, 164, 171, 193, 183, 173, 179, 188, 179, 167, 178, 180, 168, 148, 173.

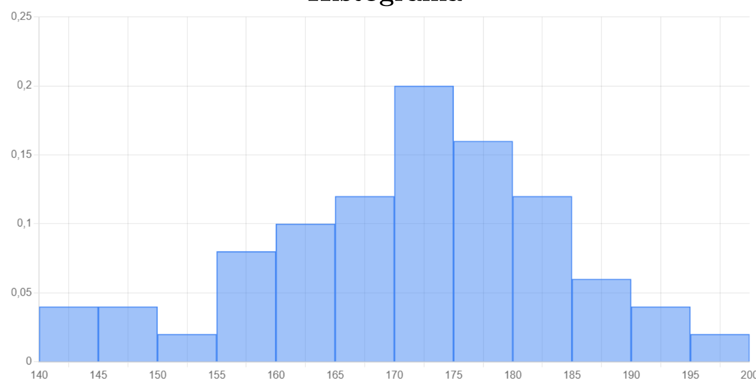
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2,$ ... ,x_k

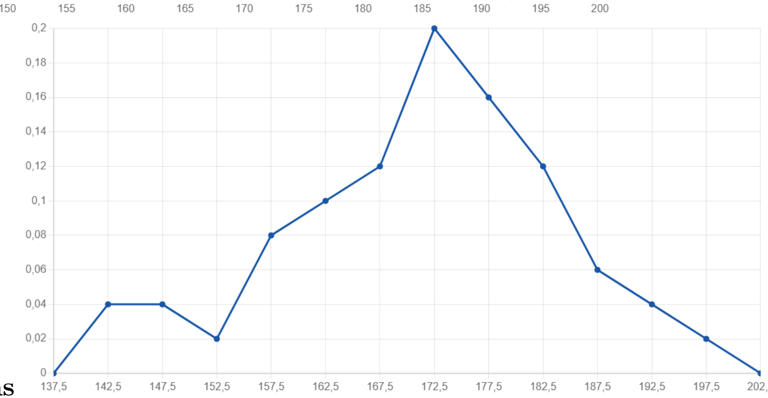
a) Agrupar los datos en intervalos de amplitud 5 desde 140 a 200 y dar la tabla de frecuencias

x_i	n_i	N_i	f_i	F_i	h_i
(140,145]	2	2	2/50	2/50	4/500
(145,150]	2	4	2/50	4/50	4/500
(150,155]	1	5	1/50	5/50	2/500
(155,160]	4	9	4/50	9/50	8/500
(160,165]	5	14	5/50	14/50	10/500
(165,170]	6	20	6/50	20/50	12/500
(170,175]	10	30	10/50	30/50	20/500
(175,180]	8	38	8/50	38/50	16/500
(180,185]	6	44	6/50	44/50	12/500
(185,190]	3	47	3/50	47/50	6/500
(190,195]	2	49	2/50	49/50	4/500
(195,200]	1	50	1/50	50/50	2/500

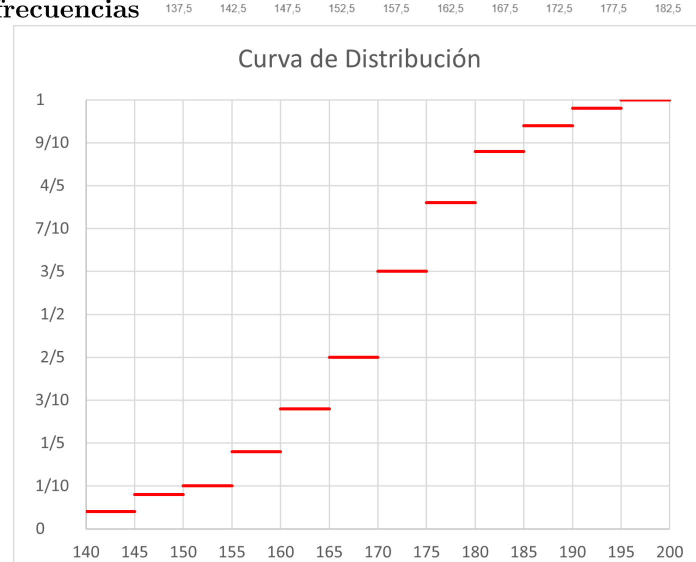
b) Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.

Histograma





Poligonal de frecuencias



1.3. Ejercicio 3:

La distribución de la renta familiar en el año 2003 por comunidades autónomas se recoge en la siguiente tabla:

I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
(8300,9300]	2						
(,10200]		5					
				10/18		1100	
			2/18		12000		
	4						0.005/18
		18					0.002/18

n_i: frecuencias absolutas
N_i: frec. absolutas acumuladas
f_i: frecuencias relativas
F_i: frec. relativas acumuladas
c_i: marcas de clase
a_i: amplitudes
h_i: densidades de frecuencia

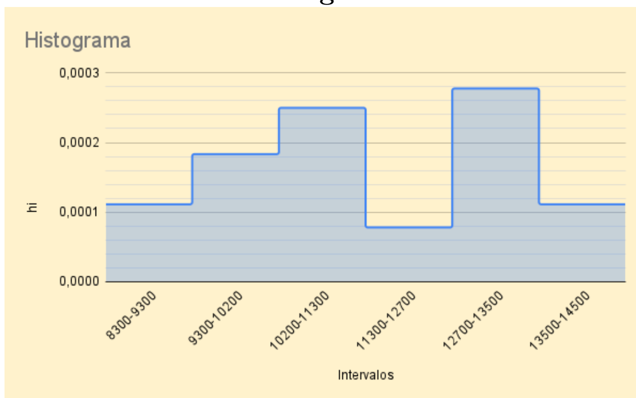
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

a) Completar la tabla

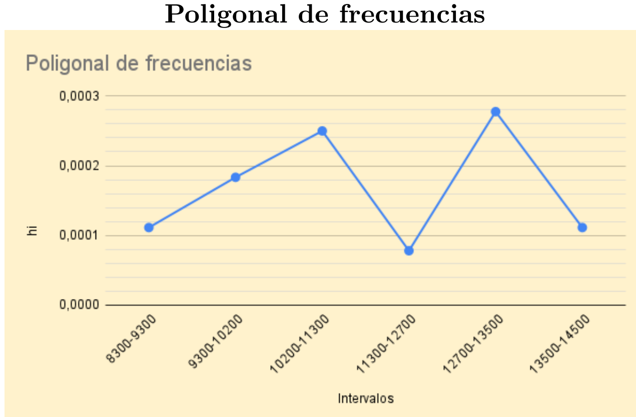
I_i	n_i	N_i	f_i	F_i	c_i	a_i	h_i
(8300,9300]	2	**2**	**2/18**	**2/18**	**8800**	**1000**	**0.002/18**
(**9300**,10200]	**3**	5	**3/18**	**5/18**	**9750**	**900**	**0.0033/18**
(10200,11300]	**5**	**10**	**5/18**	10/18	**10750**	1100	**0.0045/18**
(11300,12700]	**2**	**12**	2/18	**12/18**	12000	**1400**	**0.0014/18**
(12700,13500]	4	**16**	**4/18**	**16/18**	**13100**	**800**	0.005/18
(13500,14500]	**2**	18	**2/18**	**1**	**14000**	**1000**	0.002/18

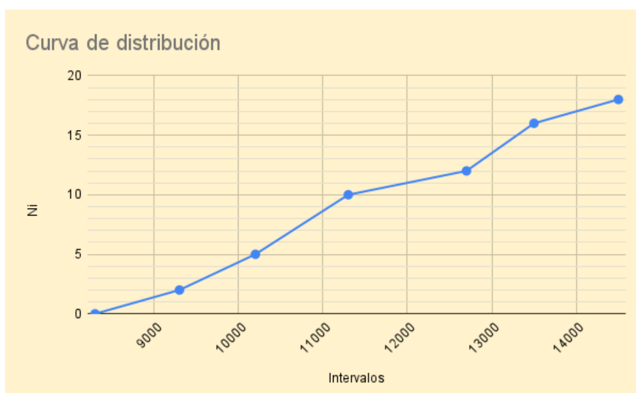
b) Representar la distribución mediante un histograma, poligonal de frecuencias y curva de distribución.

Histograma



Poligonal de frecuencias





Curva de distribución

c) ¿Cuántas comunidades presentan una renta menor o igual que 12700 euros? ¿Y cuántas superior a 11300 euros?

12 comunidades presentan una renta menor o igual a 12700 euros (N_4) y 8 comunidades presentan una renta superior a 11300 euros ($n - N_3$).

1.4. Ejercicio 4:

En una determinada empresa se realiza un estudio sobre la calidad de su producción. La distribución siguiente informa sobre el número de piezas defectuosas encontradas en 100 cajas examinadas con 50 unidades cada una de ellas:

Nº de piezas defectuosas	0	1	2	3	4	5	6	7	8	9	10
Nº de cajas	6	9	10	11	14	16	16	9	4	3	2

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

x_i	n_i	N_i
0	6	6
1	9	15
2	10	25
3	11	36
4	14	50
5	16	66
6	16	82
7	9	91
8	4	95
9	3	98
10	2	100

a) Calcular el número medio de piezas defectuosas por caja.

$$\bar{x} = \frac{6+9+10+11+14+16+16+9+4+3+2}{11} = 9.0909\, piezas$$

b) ¿Cuántas piezas defectuosas se encuentra más frecuentemente en las cajas examinadas?

Este apartado hace referencia a la moda:

$$Mo_1 = 5$$
$$Mo_2 = 6$$

En este caso la variable es de tipo cuantitativa discreta con
$n_6, n_7 > n_i\, \forall i \in \{1, ..., 11\} \,/\, i \neq 6, 7$
Por lo que se encuentra 5 o 6 piezas más frecuentemente.

c) ¿Cuál es el número mediano de piezas defectuosas por caja?

$$Me = x_i \,/\, \frac{n}{2} = N_i \Rightarrow Me = 4\, piezas$$

d) Calcular los cuartiles de la distribución. Interpretarlos.

$$Q_1 = x_i \,/\, \frac{n}{4} = N_i \Rightarrow Q_1 = 2\, piezas$$

$$Q_2 = x_i \,/\, \frac{n}{2} = N_i \Rightarrow Q_2 = Me = 4\,piezas$$

$$Q_3 = x_i \,/\, N_{i-1} < \frac{3n}{4} \leq N_i \Rightarrow Q_2 = 6\,piezas$$

El cuartil 1 nos indica que el 25 % de las cajas tiene menos o hasta 2 piezas defectuosas.
El cuartil 2 nos indica que el 50 % de las cajas tiene menos o hasta 4 piezas defectuosas.
El cuartil 3 nos indica que el 75 % de las cajas tiene menos o hasta 6 piezas defectuosas.

e) Calcular los deciles de orden 3 y 7. Interpretarlos.

$$D_3 = x_i \,/\, N_{i-1} < \frac{3n}{10} \leq N_i \Rightarrow D_3 = 3\,piezas$$

$$D_7 = x_i \,/\, N_{i-1} < \frac{7n}{10} \leq N_i \Rightarrow D_7 = 6\,piezas$$

El decil 3 nos indica que el 30 % de las cajas tiene menos o hasta 3 piezas defectuosas.
El decil 7 nos indica que el 70 % de las cajas tiene menos o hasta 6 piezas defectuosas.

f) Cuantificar la dispersión de la distribución utilizando diferentes medidas, interpretando los resultados y señalando las ventajas e inconvenientes de cada una.

Medidas de dispersión absolutas:

$$Recorrido\ o\ rango : R = x_k - x_1 = 10 - 0 = 10\,piezas$$

Nos indica como de esparcidos están los datos.
Su principal desventaja es que es difícil de interpretar y es poco preciso ya que puede haber datos aislados que modifiquen ampliamente este resultado.

$$Recorrido\ intercuartilico : R_I = Q_3 - Q_1 = 6 - 2 = 4\,piezas$$

Indica como de esparcido está el 50 % central de los datos.
Al igual que el dato anterior es poco interpretable.

$$Desviación\ absoluta\ media\ respecto\ a\ \bar{x} : D_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{n} = 4.7673$$

Indica cómo de cercanos están los datos a la media (de media).
Cuanto más pequeño sea este resultado mayor representatividad tendrá la media.
Su desventaja es que al ser una medida absoluta es difícil de interpretar si es “grande” o “pequeña”.
Además es costosa de calcular.

$$Desviación\ absoluta\ media\ respecto\ a\ Me : D_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{n} = 2$$

Indica cómo de cercanos están los datos a la mediana (de media).

Cuanto más pequeño sea este resultado mayor representatividad tendrá la mediana.
Al igual que el resultado anterior, al ser una medida absoluta es difícil de interpretar si es "grande" o "pequeña".
También es difícil de calcular.

$$Varianza: Var(X) = \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i}{n} = 28.25$$

Representa la variabilidad de los datos con respecto a la media (al igual que la desviación absoluta media con respecto a la media).
De igual forma es difícil de interpretar y no nos da una medida representativa de la bondad de la media.

$$Dexviación\ tipica: \sigma = +\sqrt{\sigma^2} = +\sqrt{28.25} = 5.315$$

Es la raíz cuadrada de la Varianza y por ello hereda las ventajas e inconvenientes de la misma.
Al igual que la varianza, este resultado trata de indicar la medida de representatividad de la media.

Medidas de dispersión relativas:

$$Coeficiente\ de\ apertura: C_A = \frac{x_k}{x_1} = \frac{10}{0} \Rightarrow imposible\ de\ calcular\ para\ esta\ distribución$$

La principal desventaja es la que se puede apreciar en este caso, que al aparecer un valor de x mínimo igual a 0, esta medida es imposible de calcular.

$$Recorrido\ relativo: R_R = \frac{R}{\bar{x}} = \frac{10}{9.09} = 1.1$$

Indica el número de veces que la media está contenida en el recorrido.
Al igual que el recorrido su principal desventaja es que un valor extremo puede afectar en gran medida a este resultado.

$$Recorrido\ semi-intercuartilico: R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{6-2}{6+2} = \frac{4}{8} = 0.5$$

La ventaja de este dato es que no se ve afectado por los datos extremos.
Trabaja con el 50 % central de los datos.

$$Coeficiente\ de\ variación\ de\ Pearson: C.V.(X) = \frac{\sigma_x}{|\bar{x}|} = \frac{5.315}{9.09} = 0.585$$

Nos indica la dispersión relativa de los datos.
Al ser una medida relativa, es más fácil de interpretar y nos va a permitir comparar diferentes distribuciones.

$$Indice\ de\ dispersión\ respecto\ a\ la\ mediana : V_{Me} = \frac{D_{Me}}{Me} = \frac{2}{4} = 0.5$$

Al igual que el resultado anterior da una medida de dispersión de los datos pero en este caso con respecto a la mediana.
Igual que el anterior resultado, nos va a permitir comparar distintas distribuciones.

1.5. Ejercicio 5:

Dadas las siguientes distribuciones:

$I_i^{(1)}$	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]
$n_i^{(1)}$	12	13	11	8	6

$I_i^{(2)}$	(0,1]	(1,3]	(3,6]	(6,10]	(10,12]
$n_i^{(2)}$	1	6	7	12	2

Calcular para cada una de ellas:

a) Medias aritmética, armónica y geométrica.

Las medias se obtienen mediante las siguientes fórmulas, tomando las marcas de clase:

$$\bar{x} = \sum_{i=1}^{k} f_i c_i$$

$$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{c_i}}$$

$$G = \sqrt[n]{\prod_{i=1}^{k} f_i c_i}$$

Medias	Distribución 1	Distribución 2
Aritmética	2,16	5,7857
Armónica	1,2289	3,3991
Geométrica	1,3419	1,5850

b) El valor más frecuente

Las modas en variable continua se obtienen mediante la siguiente fórmula, aplicando la semejanza de triángulos en el intervalo modal (aquel con mayor frecuencia):

$$Mo = e_{i-1} + \frac{h_i - h_{i-1}}{2h_i - h_{i-1} - h_{i+1}}(e_i - e_{i-1})$$

$Mo(1) = 1,314285714$
$Mo(2) = 6,5$

c) El valor superado por el 50 % de las observaciones.

Las medianas se calculan haciendo una interpolación en el intervalo mediano, pero si ocurre, como es el caso, que para algún i, $N_i = \frac{n}{2}$, entonces $Me = e_i$. Así, las medianas son:

$Me(1) = e_2 = 2$
$Me(2) = e_3 = 6$

d) Recorrido, recorrido intercuartílico y desviación típica. Interpretarlos. ¿Qué distribución es más homogénea?

El recorrido no es más que el rango total en el que se encuentran los datos, es decir, el máximo menos el mínimo.

$R^{(1)} = 5$

$R^{(2)} = 12$

El recorrido intercuartílico requiere un cierto cálculo, ya que se obtiene al restar los cuartiles Q_3 y Q_1 de ambas distribuciones:

$Q_1 = P_{25} = e_{i-1} + \frac{0{,}25n - N_{i-1}}{n_i}(e_i - e_{i-1})$

$Q_3 = P_{75} = e_{j-1} + \frac{0{,}75n - N_{j-1}}{n_j}(e_j - e_{j-1})$

Siendo e_i y e_{i-1} los extremos del intervalo I_i y e_j y e_{j-1} los extremos del intervalo I_j de tal manera que $N_{i-1} < \frac{n}{4} \leq N_i$ y $N_{j-1} < \frac{3n}{4} \leq N_j$. Es decir que I_i es el intervalo bajo el que se encuentra el 25 por ciento de los datos y I_j el intervalo bajo el que se encuentra el 75 por ciento de los datos.

Para la primera distribución, $i = 2$ (segundo intervalo) y $j = 4$ (cuarto intervalo).
Para la segunda distribución, $i = 2$ y $j = 4$ también, aunque en este caso da la casualidad de que $N_2 = \frac{n}{4}$, luego $Q_1^{(1)} = e_2 = 3$.

$R_I^{(1)} = Q_3^{(1)} - Q_1^{(1)} = 3,1875 - 1,0385 = 2,1490$

$R_I^{(2)} = Q_3^{(2)} - Q_1^{(2)} = 8,3333 - 3 = 5,3333$

Comparando sendos valores con el recorrido de cada distribución, se interpreta que ambas distribuciones tienen una centralización del 50 por ciento de los datos similar:

$\frac{R_I^{(1)}}{R^{(1)}} = \frac{2{,}1490}{5} = 0,4280$

$\frac{R_I^{(2)}}{R^{(2)}} = \frac{5{,}3333}{12} = 0,4444$

De aquí se deduce que el recorrido intercuartílico de la segunda distribución representa una mayor fracción del recorrido total, y que por tanto dicha distribución es ligeramente más homogénea en cuanto a la centralización del 50 por ciento de los datos.
No obstante, cabe destacar que la primera distribución está desplazada a la izquierda y la segunda desplazada hacia la derecha, información que nos proporcionaría el coeficiente de asimetría de Fisher, por ejemplo.

La desviación típica se define como la raíz cuadrada de la varianza, y se calcula de la siguiente manera:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}$$

Por lo tanto:

$\sigma^{(1)} = \sqrt{\sigma^{2(1)}} = \sqrt{1,7444} = 1,3208$

$\sigma^{(2)} = \sqrt{\sigma^{2(2)}} = \sqrt{8,5255} = 2,9198$

Este coeficiente nos proporciona información sobre cuánto se desvían los datos de la media aritmética. Si lo comparamos con el recorrido total obtenemos:

$\frac{\sigma^{(1)}}{\frac{1}{2}R^{(1)}} = \frac{1,3208}{2,5} = 0,5280$

$\frac{\sigma^{(2)}}{\frac{1}{2}R^{(2)}} = \frac{2,9198}{6} = 0,4866$

Aquí observamos que en la primera distribución los datos se desvían de la media un 52,8 por ciento de la mitad del recorrido, mientras que en la segunda se desvían un 48,66 por ciento de la mitad del recorrido, lo que quiere decir que en la segunda se desvían menos en relación con el recorrido, concluyendo con esto y con la interpretación del recorrido intercuartílico que la segunda distribución es ligeramente más homogénea.

1.6. Ejercicio 6:

Un móvil efectúa un recorrido de 100 km en dos sentidos. En uno va a una velocidad constante de $V_1 = 60$ km/h y en el otro va a una velocidad constante de $V_2 = 70$ km/h. Calcular la velocidad media del recorrido.
Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, ... ,x_k$

La media más adecuada para esta situación es la media armónica ya que se trata de una magnitud compuesta (km/h). El cálculo sería:

$$\bar{x} = \frac{200}{\frac{100}{70} + \frac{100}{60}} = 64.615\, km/h$$

1.7. Ejercicio 7:

Las acciones de una empresa han producido los siguientes rendimientos netos anuales

Año	Rentabilidad
1994	12 %
1995	10 %
1996	7 %
1997	6 %
1998	5 %

Obtener el rendimiento neto medio en esos cinco años.

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) $x_1, x_2, ... ,x_k$
La media más adecuada para esta situación es la media geométrica ya que la variable presenta un comportamiento multiplicativo acumulativo. El cálculo sería:

$$\bar{x} = G = \sqrt[5]{1.12 \times 1.1 \times 1.07 \times 1.06 \times 1.05} = 1.08$$

Traducido en términos de rendimiento, el rendimiento neto medio ha sido del
8 %. $((1.08 - 1) \times 100 = 8)$

1.8. Ejercicio 8:

Un profesor califica a sus alumnos según el criterio siguiente: 40 % de suspensos, 30 % de aprobados, 15 % notables, 10 % sobresalientes y 5 % matrículas. Las notas obtenidas son las siguientes:

(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]
34	74	56	81	94	70	41	28	16	4

Calcular las notas máximas para obtener cada una de las calificaciones.

Supongamos una población estadística de n elementos o individuos, en la que se desea estudiar una variable (o atributo) que presenta los valores (o modalidades) x_1, x_2, ... ,x_k

l_i	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]
n_i	34	74	56	81	94
N_i	34	108	164	245	339
f_i	0,0682731	0,1485944	0,1124498	0,1626506	0,1887550
F_i	0,0682731	0,2168675	0,3293173	0,4919679	0,6807229
h_i	0,0682731	0,1485944	0,1124498	0,1626506	0,1887550

l_i	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]
n_i	70	41	28	16	4
N_i	409	450	478	494	498
f_i	0,1405622	0,0823293	0,0562249	0,0321285	0,0080321
F_i	0,8212851	0,9036145	0,9598394	0,9919679	1,0000000
h_i	0,1405622	0,0823293	0,0562249	0,0321285	0,0080321

Nota máxima para estar suspenso: percentil 40:

$$3 + \frac{498 \times \frac{40}{100} - 164}{81} = 3.435$$

Nota máxima para estar aprobado: percentil 70:

$$5 + \frac{498 \times \frac{70}{100} - 339}{70} = 5.137$$

Nota máxima para estar notable: percentil 85:

$$6 + \frac{498 \times \frac{85}{100} - 409}{41} = 6.348$$

Nota máxima para estar sobresaliente: percentil 95:

$$7 + \frac{498 \times \frac{95}{100} - 450}{28} = 7.825$$

Nota máxima para ser matricula: percentil 100: 10

$$9 + \frac{498 \times \frac{100}{100} - 494}{4} = 10$$

1.9. Ejercicio 9:

Se ha medido la altura de 110 jóvenes, obteniendo:

Altura	(1.55,1.60]	(1.60,1.70]	(1.70,1.80]	(1.80,1.90]	(1.90,2.00]
Nº Jóvenes	18	31	24	20	17

a) Si se consideran bajos el 3 % de los individuos de menor altura, ¿cuál es la altura máxima que pueden alcanzar?

La altura máxima para ser bajo es el percentil 3, porque los que queden por debajo del 3 % se consideran bajos: (hacer interpolación en (1.60, 1.70]):

$$1,60 + \frac{110 \times \frac{3}{100} - 0}{18} \times 0,05 = 1,5592$$

b) Si se consideran altos el 18 % de los individuos de mayor altura, ¿cual es su altura mínima?

La altura máxima para no ser alto es la misma que la altura mínima para ser alto, que es el percentil 82: (hacer interpolación en (1.80, 1.90]):

$$1,80 + \frac{110 \times \frac{82}{100} - 73}{20} \times 0,1 = 1,886$$

c) ¿Qué altura es superada sólo por 1/4 de los jóvenes?

Cuartil 3: (hacer interpolación en (1.80, 1.90]):

$$1,80 + \frac{110 \times \frac{3}{4} - 73}{20} \times 0,1 = 1,8475$$

d) Calcular el número de jóvenes cuya altura es superior a 1.75.

Nº de sujetos mayores de 1.75: n -(hacer interpolación inversa en (1.70, 1.80]):

$$110 - \frac{100}{110} \times (24 \times \frac{1,75 - 1,7}{0,1} + 49) = 54,5454$$

e) Calcular la altura máxima de los 11 jóvenes más bajos.

Altura superada por 11 de la muestra: percentil 10: (hacer interpolación en (1.55, 1.60])

$$1,55 + \frac{110 \times \frac{10}{100} - 0}{18} \times 0,05 = 1,5805$$

f) Calcular la altura mínima de los 11 jóvenes más altos.

Altura superada por 99 de la muestra: percentil 90:(hacer interpolación en (1.90, 2.00]):

$$1,90 + \frac{110 \times \frac{90}{100} - 93}{17} \times 0,1 = 1,9352$$

1.10. Ejercicio 10:

Realizando una prueba para el estudio del cáncer a 150 personas se obtuvo la siguiente tabla según la edad de los enfermos:

Edad	(10,30]	(30,40]	(40,50]	(50,60]	(60,90]
Nº Enfermos	15	22	48	40	25

Edad	(10, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 90]
Nº enfermos (n_i)	15	22	48	40	25
N_i	15	37	85	125	150
f_i	0,1	0,1467	0,32	0,2667	0,1667
F_i	0,1	0,2467	0,5667	0,8333	1
a_i	20	10	10	10	30
h_i	0,005	0,01467	0,032	0,02667	0,00556

a) Calcular la edad más común de los individuos estudiados. Para ello calcularemos la moda:

$$Mo = 40 + \frac{0.032 - 0.01467}{(2 \times 0.032 - 0.01467 - 0.02667) \times 10} = 47.65\, años$$

b) Calcular la edad mínima y máxima del 30 % central de los individuos. Percentil 35:

$$P_{35} = 40 + \frac{\frac{150 \times 35}{100} - 37}{48} \times 10 = 43.23\, años$$

Percentil 65:

$$P_{65} = 50 + \frac{\frac{150 \times 65}{100} - 85}{40} \times 10 = 53.125\, años$$

c) Calcular el recorrido intercuartílico y la desviación típica.

Recorrido intercuartílico:

$R_I = Q_3 - Q_1 = P_{75} - P_{25} = (50 + \frac{\frac{3}{4}150 - 85}{40} \times 10) - (40 + \frac{\frac{1}{4}150 - 37}{48} \times 10) =$
$56,875 - 40,104 = 16,771$

Desviación típica:

$\bar{x} = \sum_{i=1}^{k} f_i c_i = 0,1 \times 15 + 0,1467 \times 35 + 0,32 \times 45 + 0,2667 \times 55 + 0,1667 \times 75 = 48,2055$

$\sigma^2 = 0,1 \times (15 - 48,2055)^2 + 0,1467 \times (35 - 48,2055)^2 + 0,32 \times (45 - 48,2055)^2 + 0,2667 \times (55 - 48,2055)^2 + 0,1667 \times (75 - 48,2055)^2 = 271,125$

$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} f_i (c_i - \bar{x})^2} = \sqrt{271,125} = 16,466$

d) Calcular e interpretar los valores de los coeficientes de asimetría y curtosis.

Coeficiente de asimetría de Fisher:

$\gamma_1 = \sum_{i=1}^{k} f_i(\frac{c_i - \bar{x}}{\sigma})^3 = -0,1611$

Coeficiente de asimetría de Pearson:

$A_p = \frac{\bar{x} - Mo}{\sigma} = \frac{48,2055 - 47,65}{16,466} = 0,0337$

De aquí se puede deducir que la distribución está bastante centrada, pues ambos coeficientes dan valores de diferente signo aunque muy próximos a cero. El coeficiente de Fisher con signo negativo está más lejano a cero, así que se puede decir que la distribución es ligeramente asimétrica hacia la izquierda.

Coeficiente de curtosis de Fisher:

$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{212561,932}{73511,019} - 3 = -0,1084$

Coeficiente de curtosis de Kelley:

$K = \frac{R_I}{2(D_9 - D_1)} = \frac{16,771}{2(72-30)} = 0,1997$

Nuevamente ambos coeficientes tienen valores próximos a cero aunque con signos distintos. Esto da a entender que la distribución es más bien mesocúrtica, aunque al ser el coeficiente positivo el que más se aleja del cero, se puede decir que es ligeramente leptocúrtica.

2. Relación 2

2.1. Ejercicio 1:

Se han lanzado dos dados varias veces, obteniendo los resultados que se presentan en la siguiente tabla, donde X designa el resultado del primer dado e Y el resultado del segundo:

X	1	2	2	3	5	4	1	3	3	4	1	2	5	4	3	4	4	5	3	1	6	5	4	6
Y	2	3	1	4	3	2	6	4	1	6	6	5	1	2	5	1	1	2	6	6	2	1	2	5

a) Construir la tabla de frecuencias

X/Y	1	2	3	4	5	6	$f_{i.}$
1	0	0.04167	0	0	0	0.125	0.16667
2	0.4167	0	0.4167	0	0.04167	0	0.125
3	0.4167	0	0	0.0833	0.04167	0.04167	0.20833
4	0.08333	0.125	0	0	0	0.04167	0.25
5	0.08333	0.4167	0.4167	0	0	0	0.1667
6	0	0.04167	0	0	0.04167	0	0.08333
$f_{.j}$	0.25	0.25	0.08333	0.08333	0.125	0.20833	1

b) Calcular las puntuaciones medias obtenidas con cada dado y ver cuales son más homogéneas.

Empezaremos calculando las medias:
Media de los resultados del dado x:

$$\bar{x} = \frac{1}{6}\sum_{i=1}^{6}(x_i * n_{i.}) = 3,375$$

Media de los resultados del dado y:

$$\bar{y} = \frac{1}{6}\sum_{j=1}^{6}(y_j * n_{.j}) = 3,208333333$$

desv(x):

$$\sigma^2 = \frac{1}{6}\sum_{i=1}^{6}(x_{i.} - \bar{x})^2 = 2,317708333$$

desv(y):

$$\sigma^2 = \frac{1}{6}\sum_{i=1}^{6}(y_{i.} - \bar{y})^2 = 3,692708333$$

coeficiente de variación:

$$C.V.(x) = \frac{\sigma_x}{\bar{x}} = 0,4510821211$$

$$C.V.(y) = \frac{\sigma_y}{\bar{y}} = 0,5989533796$$

El dado x tiene menor coeficiente de variación, por tanto sus resultados son mas homogéneos

c) ¿Qué resultado del segundo dado es más frecuente cuando en el primero se obtiene un 3?

El resultado mas frecuente de la distribución marginal cuando X=3 es Y=4 con una frecuencia de 0,0833333.

d) Calcular la puntuación máxima del 50 % de las puntuaciones más bajas obtenidas con el primer dado si con el segundo se ha obtenido un 2 o un 5.

X/Y	2	5	2+5	F_i
1	1	0	1	1
2	0	1	1	2
3	0	1	1	3
4	3	0	3	**6**
5	1	0	1	7
6	1	1	2	9
$n_{.j}$	6	3	9	

La mitad de 9 (4.5) se obtiene para $x_i = 4$ por lo que este será el valor de la mediana.

2.2. Ejercicio 2:

Medidos los pesos, X (en Kg), y las alturas, Y (en cm), a un grupo de individuos, se han obtenido los siguientes resultados:

X/Y	160	162	164	166	168	170
48	3	2	2	1	0	0
51	2	3	4	2	2	1
54	1	3	6	8	5	1
57	0	0	1	2	8	3
60	0	0	0	2	4	4

a) Calcular el peso medio y la altura media y decir cuál es más representativo.

Antes de nada vamos a añadir las frecuencias marginales a la tabla.

X/Y	160	162	164	166	168	170	$n_{i.}$
48	3	2	2	1	0	0	8
51	2	3	4	2	2	1	14
54	1	3	6	8	5	1	24
57	0	0	1	2	8	3	14
60	0	0	0	2	4	4	10
$n_{.j}$	6	8	13	15	19	9	70

El peso viene dado por la variable X con lo cual podemos calcular el peso medio tomando la media marginal de x:

$\bar{x} = \sum_{i=1}^{5} f_{i.} x_i = 54.1714\ kg$

La altura media la da la media marginal de la variable Y:

$\bar{y} = \sum_{j=1}^{6} f_{.j} y_j = 165.7142\ cm$

Para decir cuál es más representativa se pueden tener en cuenta diversos aspectos. Por ejemplo, será más representativa aquella media de la que se alejen menos los datos de forma relativa. Esta información la da el coeficiente de variación de Pearson:

$$\sigma_x = \sqrt{\sum_{i=1}^{5} f_{i.}(x_i - \bar{x})^2} = 3,6074\ kg \quad C.V.(X) = \frac{\sigma_x}{\bar{x}} = 0.0666$$

$$\sigma_y = \sqrt{\sum_{j=1}^{6} f_{.j}(y_j - \bar{y})^2} = 2,9740 cm \quad C.V.(Y) = \frac{\sigma_y}{\bar{y}} = 0.0179$$

Con esto podemos afirmar que la altura media es más representativa que el peso medio.

b) Calcular el porcentaje de individuos que pesan menos de 55 kg y miden más de 165 cm.

Vamos a hacer una tabla con la submatriz que recoge a individuos que pesan menos de 55 kg y miden más de 165 cm:

X/Y	166	168	170	$n_{i.}$
48	1	0	0	1
51	2	2	1	5
54	8	5	1	14
$n_{.j}$	11	7	2	20

Como vemos hay 20 individuos que cumplen los requisitos con lo que el porcentaje será:

$\frac{20}{70} \times 100 = 28.57\,\%$ de la población

c) Entre los que miden más de 165 cm, ¿cuál es el porcentaje de los que pesan más de 52kg?

Hay un total de $15 + 19 + 9 = 43$ individuos que miden más de 165 cm.

X/Y	166	168	170	$n_{i.}$
54	8	5	1	14
57	2	8	3	13
60	2	4	4	10
$n_{.j}$	12	17	8	37

Y como vemos en la tabla superior hay un total de 37 individuos que miden más de 165 cm y además pesan más de 52 kg. Por lo tanto el porcentaje es:

$\frac{37}{43} \times 100 = 86.05\,\%$

Entonces un $86.05\,\%$ de los individuos que miden más de 165cm pesan además más de 52 kg.

d) ¿Cuál es la altura más frecuente entre los individuos cuyo peso oscila entre 51 y 57 kg?

Primero reduzcamos la población respecto del carácter X:

X/Y	160	162	164	166	168	170	$n_{i.}$
51	2	3	4	2	2	1	14
54	1	3	6	8	5	1	24
57	0	0	1	2	8	3	14
$n_{.j}$	3	6	11	12	15	5	52

De la tabla de arriba, tan solo nos tenemos que fijar en los $n_{.j}$ y ver cuál es el máximo. Cómo vemos el máximo se alcanza para $j = 5$ con $y_5 = 168\ cm$, siendo esta la altura más frecuente entre los individuos de entre 51 y 57 kg.

e) ¿Qué peso medio es más representativo, el de los individuos que miden 164 cm o el de los que miden 168 cm?

Calculamos $\bar{x_j}$ para $j = 3$ e $j = 5$:

$$\bar{x_3} = \frac{1}{n_{.3}} \sum_{i=1}^{5} n_{i3} x_i = \frac{2 \times 48 + 4 \times 51 + 6 \times 54 + 1 \times 57 + 0 \times 60}{13} = 52.38\ kg$$

$$\bar{x_5} = \frac{1}{n_{.5}} \sum_{i=1}^{5} n_{i5} x_i = \frac{0 \times 48 + 2 \times 51 + 5 \times 54 + 8 \times 57 + 4 \times 60}{19} = 56.21\ kg$$

El peso medio condicionado más representativo será aquel que se aproxime más al peso medio, es decir, $\bar{x_3}$.

2.3. Ejercicio 3:

En una encuesta de familias sobre el número de individuos que la componen (X) y el número de personas activas en ellas (Y) se han obtenido los siguientes resultados:

X/Y	1	2	3	4	$n_{i.}$
1	7	0	0	0	7
2	10	2	0	0	12
3	11	5	1	0	17
4	10	6	6	0	22
5	8	6	4	2	20
6	1	2	3	1	7
7	1	0	0	1	2
8	0	0	1	1	2
$n_{.j}$	48	21	15	5	89

a) Calcular la recta de regresión de Y sobre X.

Para ello calcularé los puntos $(x_i, \bar{y}_i)$:

$$\bar{y}_1 = \frac{1 \times 7 + 2 \times 0 + 3 \times 0 + 4 \times 0}{7} = 1 \Rightarrow (1; 1)$$

$$\bar{y}_2 = \frac{1 \times 10 + 2 \times 2 + 3 \times 0 + 4 \times 0}{12} = 1.1667 \Rightarrow (2; 1.1667)$$

$$\bar{y}_3 = \frac{1 \times 11 + 2 \times 5 + 3 \times 1 + 4 \times 0}{17} = 1.412 \Rightarrow (3; 1.412)$$

$$\bar{y}_4 = \frac{1 \times 10 + 2 \times 6 + 3 \times 6 + 4 \times 0}{22} = 1.818 \Rightarrow (4; 1.818)$$

$$\bar{y}_5 = \frac{1 \times 8 + 2 \times 6 + 3 \times 4 + 4 \times 2}{20} = 2 \Rightarrow (5; 2)$$

$$\bar{y}_6 = \frac{1 \times 1 + 2 \times 2 + 3 \times 3 + 4 \times 1}{7} = 2.571 \Rightarrow (6; 2.571)$$

$$\bar{y}_7 = \frac{1 \times 1 + 2 \times 0 + 3 \times 0 + 4 \times 1}{2} = 2.5 \Rightarrow (7; 2.5)$$

$$\bar{y}_8 = \frac{1 \times 0 + 2 \times 0 + 3 \times 1 + 1 \times 0}{2} = 3.5 \Rightarrow (8; 3.5)$$

La recta de regresión será aquella que pase por los puntos calculados.

b) ¿Es adecuado suponer una relación lineal para explicar el comportamiento de Y a partir de X?

Si existiese dependencia funcional lineal, la recta de regresión X/Y y la recta Y/X deberían coincidir con la recta de dependencia.

Esto además implicaría que $r = 1$ por lo que pasaré a calcular el coeficiente de correlación lineal. En este caso:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.791}{1.586 \times 0.93} = 0.536 \neq 1$$

Por lo que no sería correcta tal suposición.

2.4. Ejercicio 4:

Se realiza un estudio sobre la tensión de vapor de agua (Y , en ml. de Hg.) a distintas temperaturas (X, en ºC). Efectuadas 21 medidas, los resultados son:

X/Y	(0.5,1.5]	(1.5,2.5]	(2.5,5.5]
(1,15]	4	2	0
(15,25]	1	4	2
(25,30]	0	3	5

Explicar el comportamiento de la tensión de vapor en términos de la temperatura mediante una función lineal. ¿Es adecuado asumir este tipo de relación?
Con las marcas de clase la distribución en la tabla de frecuencias quedaría:

X/Y	1	2	4	$n_{i.}$	$c_i \times n_{i.}$	$c_i \times n_{i.}^2$
8	4	2	0	6	48	384
20	1	4	2	7.	140	2800
27.5	0	3	5	8	220	6050
$n_{.j}$	5	9	7	21	408	9234
$d_j \times n_{.j}$	5	18	28	51	-	-
$d_j \times n_{.j}^2$	5	36	112	153	-	-

Media de x:

$$\bar{x} = \frac{1}{4}\sum_{i=1}^{4}(x_i * n_{i.}) =$$

Media de y:

$$\bar{y} = \frac{1}{4}\sum_{j=1}^{4}(y_i * n_{.j}) =$$

desv(x):

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i - \bar{x})^2 =$$

desv(y):

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(y_i - \bar{y})^2 =$$

2.5. Ejercicio 5:

Estudiar la dependencia o independencia de las variables en cada una de las siguientes distribuciones. Dar, en cada caso, las curvas de regresión y la covarianza de las dos variables.

X/Y	1	2	3	4	5
10	2	4	6	10	8
20	1	2	3	5	4
30	3	6	9	15	12
40	4	8	12	20	16

X/Y	1	2	3
-1	0	1	0
0	1	0	1
1	0	1	0

Completamos ambas tablas:

X/Y	1	2	3	4	5	$n_{i.}$
10	2	4	6	10	8	30
20	1	2	3	5	4	15
30	3	6	9	15	12	45
40	4	8	12	20	16	60
$n_{.j}$	10	20	30	50	40	150

Podemos ver que en la primera tabla se da independencia entre las variables por el **Teorema de caracterización de la independencia**, que dice lo siguiente:

$$X \ e \ Y \ independientes \iff n_{ij} = \frac{n_{i.}n_{.j}}{n} \ \forall i = 1, ..., k \ \forall j = 1, ..., p$$

Podemos comprobar que dicha caracterización se verifica $\forall i, j = 1, 2, 3, 4$ por lo que podemos afirmar que en esta primera distribución las variables son independientes.
Para esta segunda tabla, aplicando el mismo teorema, comprobamos que no se da dicha caracterización:

$$\frac{n_{1.}n_{.j}}{n} = \frac{1 \times 1}{4} = \frac{1}{4} \neq 0 = n_{11}$$

A continuación, procedemos a trazar las curvas de regresión de tipo I de ambas distribuciones. Se define la curva de regresión de tipo I de Y sobre X como aquella que pasa por todos los puntos $(x_i, \bar{y_i})$ y la curva de regresión de tipo I de X sobre Y como aquella que pasa por todos los puntos $(\bar{x_j}, y_j)$. Para el caso de variables discretas, las curvas de regresión no serán más que puntos aislados.

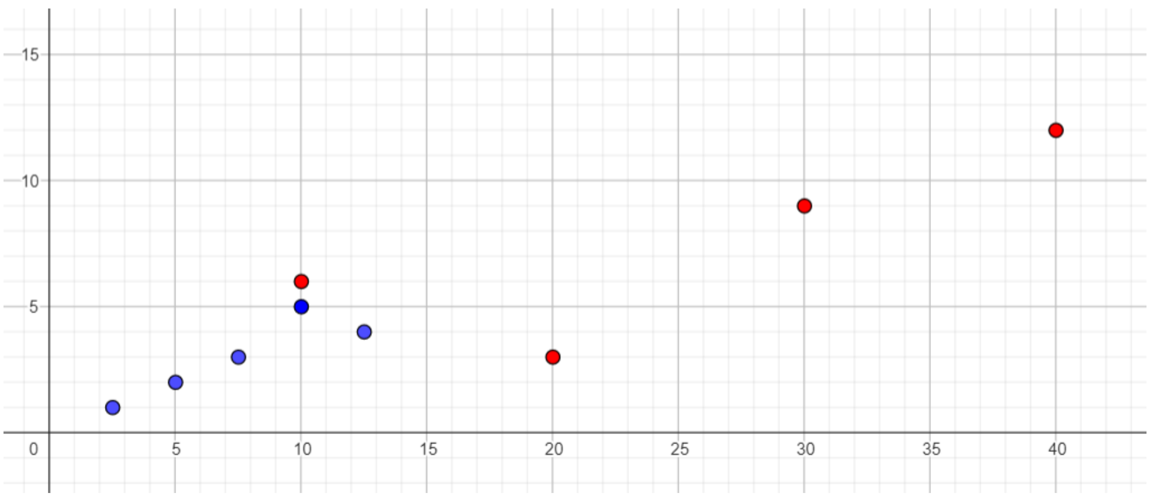
Curvas de regresión de tipo I:

Tabla 1:

X/Y	1	2	3	$n_{i.}$
-1	0	1	0	1
0	1	0	1	2
1	0	1	0	1
$n_{.j}$	1	2	1	4

x_i	$\bar{y}_i$
10	6
20	3
30	9
40	12

$\bar{x_j}$	y_j
2.5	1
5	2
7.5	3
12.5	4
10	5

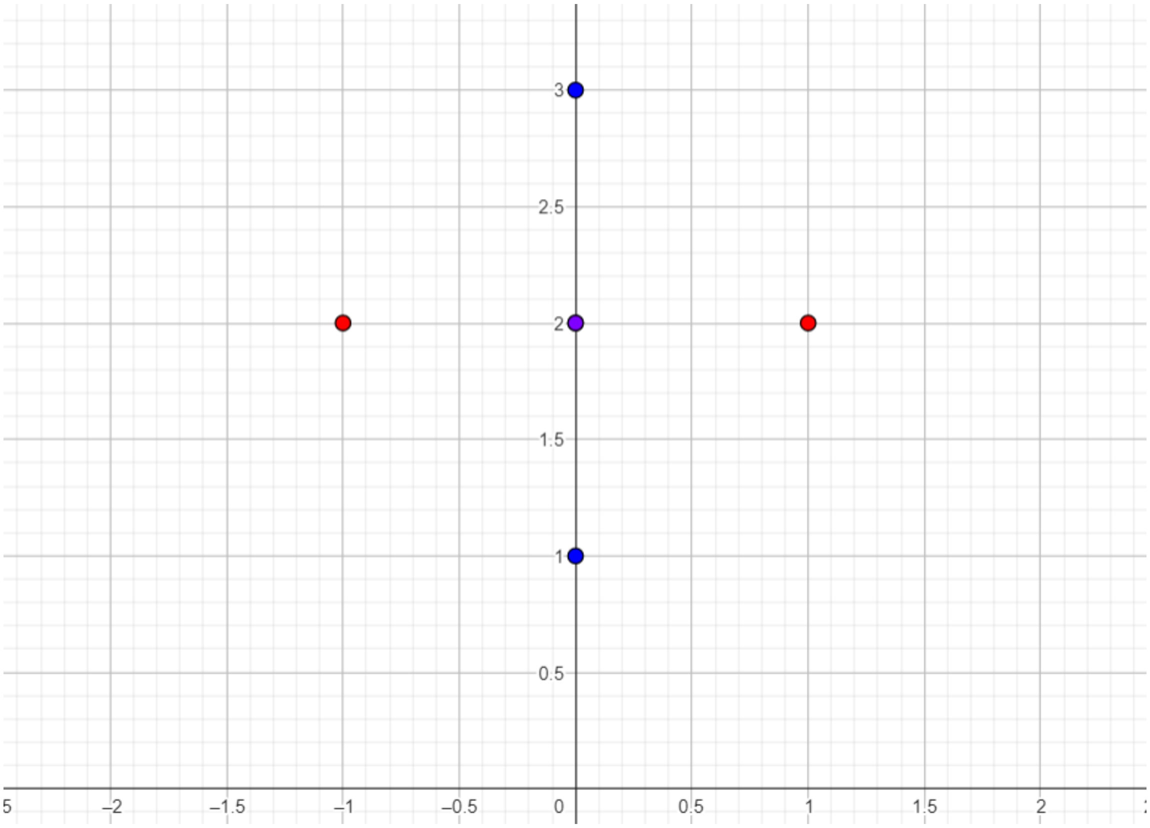


Curva de regresión tabla 1 (Y/X en rojo y X/Y en azul)

Tabla 2:

x_i	$\bar{y}_i$
-1	2
0	2
1	2

$\bar{x_j}$	y_j
0	1
0	2
0	3



Curva de regresión tabla 2 (Y/X en rojo y X/Y en azul)

En el anterior gráfico el punto (0,2) es doble, por eso está en morado.

Para terminar el ejercicio, calculamos las covarianzas de ambas distribuciones:

$\sigma_{xy}^{(1)} = \frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}(x_i-\bar{x})(y_j-\bar{y}) = 0$

$\sigma_{xy}^{(2)} = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij}(x_i-\bar{x})(y_j-\bar{y}) = 0$

Se ha dado el caso de que ambas covarianzas son nulas, lo cual indica, por ejemplo, que el coeficiente de correlación es nulo, que las rectas de regresión son perpendiculares a los ejes y de la forma $y = \bar{y}$ y $x = \bar{x}$

Cabe mencionar que el valor de la primera covarianza ya lo sabíamos, puesto que las variables X e Y eran independientes y esto implica que la covarianza tiene que ser 0. Como vemos, no se da la implicación contraria, puesto que la segunda distribución también tiene covarianza nula pero las variables no son independientes.

2.6. Ejercicio 6:

Dada la siguiente distribución bidimensional:

X/Y	1	2	3	4	
10	1	3	0	0	4
12	0	1	4	3	8
14	2	0	0	2	4
16	4	0	0	0	4
	7	4	4	5	20

a) ¿Son estadísticamente independientes X e Y?

La independencia de 2 caracteres estadísticos viene dada por la caracterización:

$$\text{X e Y son independientes} \iff n_{ij} = \tfrac{n_{i.}n_{.j}}{n} \qquad \forall i,j = 1,2,3,4$$

En el caso i=1 , j=1 tenemos:

$$n_{1,1} = 1 \neq \frac{n_{1.}n_{.1}}{n} = \frac{4\times 7}{20} = \frac{28}{20} = 1.4$$

Por lo que ambos caracteres estadísticos no son independientes.

b) Calcular y representar las curvas de regresión de X/Y e Y/X.

Primero calcularé la curva de regresión de X/Y indicando por qué puntos deberá pasar: Dichos puntos serán $(\bar{x}_j, y_j)$; $j = 1,2,3,4$;

$$\bar{x}_1 = \frac{10\times 1 + 12\times 0 + 14\times 2 + 16\times 4}{7} = 14.57 \Rightarrow (14.57; 1)$$

$$\bar{x}_2 = \frac{10\times 3 + 12\times 1 + 14\times 0 + 16\times 0}{4} = 10.5 \Rightarrow (10.5; 2)$$

$$\bar{x}_3 = \frac{10\times 0 + 12\times 3 + 14\times 2 + 16\times 0}{5} = 12 \Rightarrow (12; 3)$$

$$\bar{x}_4 = \frac{10\times 1 + 12\times 0 + 14\times 2 + 16\times 4}{7} = 12.8 \Rightarrow (12.8; 4)$$

Del mismo modo calculando la curva de Y/X calcularé $(x_i, \bar{y}_i)$:

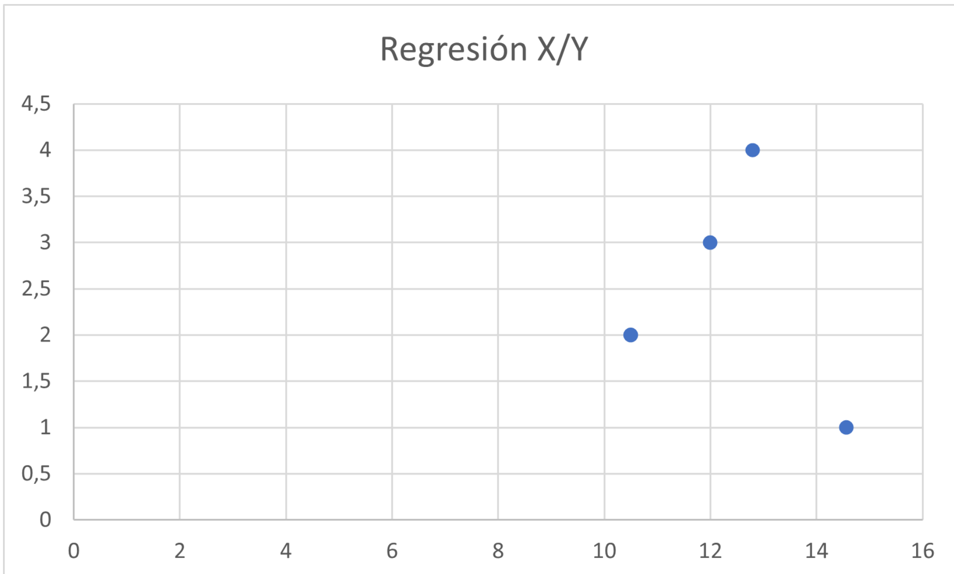
$$\bar{y}_1 = \frac{1\times 1 + 2\times 3 + 3\times 0 + 4\times 0}{4} = 1.75 \Rightarrow (1; 1.75)$$

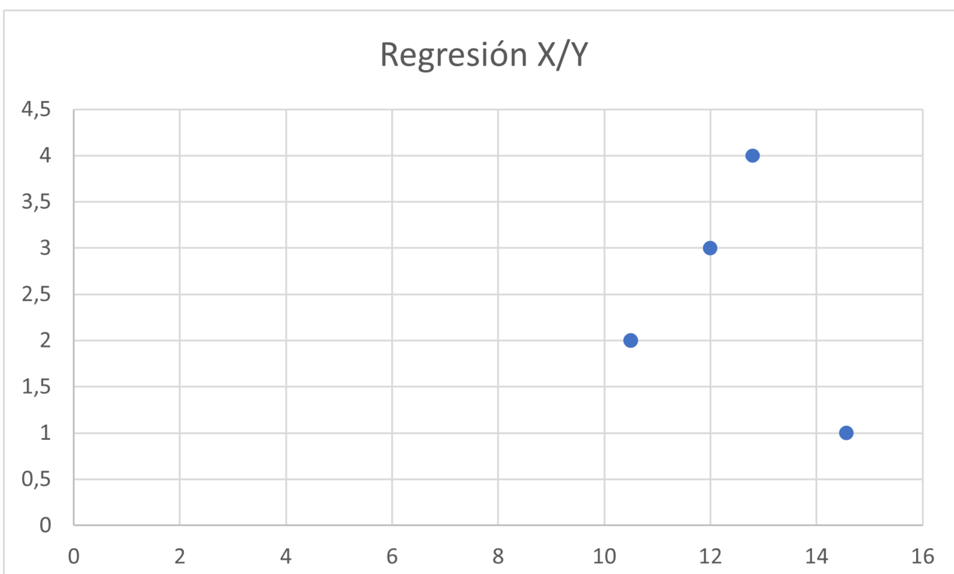
$$\bar{y}_2 = \frac{1\times 0 + 2\times 1 + 3\times 4 + 4\times 3}{8} = 3.25 \Rightarrow (2; 3.25)$$

$$\bar{y}_3 = \frac{1 \times 2 + 2 \times 0 + 3 \times 0 + 4 \times 2}{4} = 2.5 \Rightarrow (3; 2.5)$$

$$\bar{y}_4 = \frac{1 \times 4 + 2 \times 0 + 3 \times 0 + 4 \times 0}{4} = 1 \Rightarrow (4; 1)$$

Las representaciones gráficas quedarían:





c) Cuantificar el grado en que cada variable es explicada por la otra mediante la correspondiente curva de regresión.

Para ello calcularé la varianza explicada por cada una de las regresiones:

$$\sigma_{ey}^2 = \sum_{i=1}^{4}\sum_{j=1}^{4} f_{ij}(\hat{y}_j - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{4} n_{ij}(\hat{y}_j - \bar{y})^2$$

$$\bar{y} = \frac{1}{n}\sum_{j=0}^{4} y_j n_{.j} = \frac{1\times 7 + 2\times 4 + 3\times 4 + 4\times 5}{20} = 2.35$$

$$\sigma_{ey}^2 = \frac{(1+2+4)(1.75-2.35)^2 + (3+1)(3.25-2.35)^2 + 4(2.5-2.35)^2 + (3+2)(1-2.35)^2}{20} =$$

d) ¿Están X e Y correlacionadas linealmente? Dar las expresiones de las rectas de regresión.

Calcularé

2.7. Ejercicio 7:

Para cada una de las distribuciones:

a) ¿Dependen funcionalmente X de Y o Y de X?

b) Calcular curvas de regresión y comentar los resultados

Distribución A:

X/Y	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3
4	0	1	0

a) Es fácil ver que Y depende funcionalmente de X ya que a cada valor de X le corresponde un único valor no nulo de Y, pero no al revés.
b) Calcularé las curvas de regresión:
Y/X: puntos de la forma $(x_k, \bar{y_k})$

$$(x_1, \bar{y_1}) = (1, \frac{15 \times 2}{2}) = (1, 15)$$

$$(x_2, \bar{y_2}) = (2, \frac{10 \times 1}{1}) = (2, 10)$$

$$(x_3, \bar{y_3}) = (3, \frac{20 \times 3}{3}) = (3, 20)$$

$$(x_4, \bar{y_4}) = (4, \frac{5 \times 1}{1}) = (4, 15)$$

Y/X: puntos de la forma $(\bar{x_k}, y_k)$

$$(\bar{x_1}, y_1) = (\frac{2 \times 1}{1}, 10) = (2, 10)$$

$$(\bar{x_2}, y_2) = (\frac{1 \times 2 + 4 \times 1}{3}, 15) = (2, 15)$$

$$(\bar{x_3}, y_3) = (\frac{3 \times 3}{3}, 20) = (3, 20)$$

Distribución B:

X/Y	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3

a) En este caso Y depende funcionalmente de X por lo comentado anteriormente y además ahora también depende X de Y por el mismo motivo.
b) Calcularé las curvas de regresión:
Y/X: puntos de la forma $(x_k, \bar{y_k})$

$$(x_1, \bar{y_1}) = (1, \frac{15 \times 2}{2}) = (1, 15)$$

$$(x_2, \bar{y_2}) = (2, \frac{10 \times 1}{1}) = (2, 10)$$

$$(x_3, \bar{y_3}) = (3, \frac{20 \times 3}{3}) = (3, 20)$$

Y/X: puntos de la forma $(\bar{x_k}, y_k)$

$$(\bar{x_1}, y_1) = (\frac{2 \times 1}{1}, 10) = (2, 10)$$

$$(\bar{x_2}, y_2) = (\frac{1 \times 2}{3}, 15) = (1, 15)$$

$$(\bar{x_3}, y_3) = (\frac{3 \times 3}{3}, 20) = (3, 20)$$

Como se puede apreciar ambas curvas son iguales (o por lo menos en los puntos calculados).

Distribución C:

X/Y	10	15	20	25
1	0	3	0	1
2	0	0	1	0
3	2	0	0	0

a) Es fácil ver que X depende funcionalmente de Y ya que a cada valor de Y le corresponde un único valor no nulo de X, pero no al revés.
b) Calcularé las curvas de regresión:
Y/X: puntos de la forma $(x_k, \bar{y_k})$

$$(x_1, \bar{y_1}) = (1, \frac{15 \times 3 + 25 \times 1}{4}) = (1; 17.5)$$

$$(x_2, \bar{y_2}) = (2, \frac{20 \times 1}{1}) = (2, 20)$$
$$(x_3, \bar{y_3}) = (3, \frac{10 \times 2}{2}) = (3, 10)$$

Y/X: puntos de la forma $(\bar{x_k}, y_k)$

$$(\bar{x_1}, y_1) = (\frac{3 \times 2}{2}, 10) = (3, 10)$$
$$(\bar{x_2}, y_2) = (\frac{1 \times 3}{3}, 15) = (1, 15)$$
$$(\bar{x_3}, y_3) = (\frac{2 \times 1}{1}, 20) = (2, 20)$$
$$(\bar{x_4}, y_4) = (\frac{1 \times 1}{1}, 25) = (1, 25)$$

2.8. Ejercicio 8:

De una muestra de 24 puestos de venta en un mercado de abastos se ha recogido información sobre el número de balanzas (X) y el número de dependientes (Y). Los resultados aparecen en la siguiente tabla:

X/Y	1	2	3	4
1	1	2	0	0
2	1	2	3	1
3	0	1	2	6
4	0	0	2	3

a) Determinar las rectas de regresión.

b) ¿Es apropiado suponer que existe una relación lineal entre las variables?

c) Predecir, a partir de los resultados, el número de balanzas que puede esperarse en un puesto con seis dependientes. ¿Es fiable esta predicción?

a) Para las rectas de regresión, utilizaremos las fórmulas que se obtienen al realizar el método de mínimos cuadrados:

Recta de regresión Y/X

$$y = a + bx$$

$$a = \bar{y} - b\bar{x}$$
$$b = \frac{\sigma_{xy}}{\sigma_x^2}$$

Recta de regresión X/Y

$$x = a' + b'y$$

$$a' = \bar{x} - b'\bar{y}$$
$$b' = \frac{\sigma_{xy}}{\sigma_y^2}$$

X/Y	1	2	3	4	$n_{i.}$
1	1	2	0	0	3
2	1	2	3	1	7
3	0	1	2	6	9
4	0	0	2	3	5
$n_{.j}$	2	5	7	10	24

Cálculo de medias:

$\bar{x} = \frac{1}{n}\sum_{i=1}^{4} n_{i.}x_i = 2.667$

$\bar{y} = \frac{1}{n}\sum_{j=1}^{4} n_{.j}y_j = 3.042$

Cálculo de varianzas y covarianza:

$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{4} n_{i.}(x_i - \bar{x})^2 = 0.889$

$\sigma_y^2 = \frac{1}{n}\sum_{j=1}^{4} n_{.j}(y_j - \bar{y})^2 = 0.957$

$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{4} n_{ij}(x_i - \bar{x})(y_j - \bar{y}) = 0.597$

Cálculo de parámetros de las rectas:

$b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0.597}{0.889} = 0.672$

$b' = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{0.597}{0.957} = 0.624$

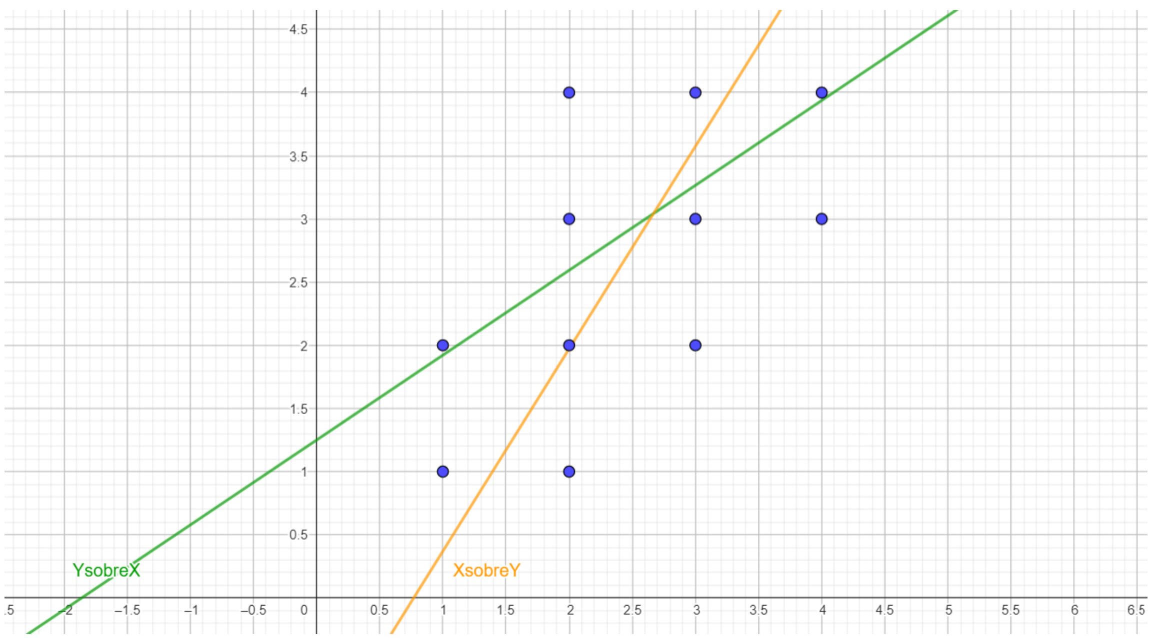
$a = \bar{y} - b\bar{x} = 3.042 - 0.672 \times 2.667 = 1.250$

$a' = \bar{x} - b'\bar{y} = 2.667 - 0.624 \times 3.042 = 0.769$

Por lo tanto, las rectas que nos quedan son:

Recta Y/X: $y = 1.25 + 0.672x$

Recta X/Y: $x = 0.769 + 0.624y$



b) Para ver si la relación lineal es fuerte, nos apoyamos en el coeficiente de correlación lineal:

$$R^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{0.597^2}{0.889 \times 0.957} = 0.419$$

Este valor se interpreta como que las rectas de regresión explican un 41.9 % de la distribución, lo cual indica que no es un buen modelo de aproximación y no sería apropiado suponer que la relación entre las variables X e Y es lineal.

c) Un puesto con 6 dependientes significa $y = 6$. Utilizamos la recta X/Y para obtener el valor esperado de x.

$$x = 0.769 + 0.624 \times 6 = 4.513 \ balanzas$$

Esta predicción no es fiable a pesar de no estar muy alejados del rango de y, puesto que se trata de un ajuste con una baja correlación lineal entre las variables.

2.9. Ejercicio 9:

Se eligen 50 matrimonios al azar y se les pregunta la edad de ambos al contraer matrimonio. Los resultados se recogen en la siguiente tabla, en la que X denota la edad del hombre e Y la de la mujer:

X/Y	(10,20]	(20,25]	(25,30]	(30,35]	(35,40]	$n_{i.}$
(15,18]	3	2	3	0	0	8
(18,21]	0	4	2	2	0	8
(21,24]	0	7	10	6	1	24
(24,27]	0	0	2	5	3	10
$n_{.j}$	3	13	17	13	4	50

Estudiar la interdependencia lineal entre ambas variables.
Calcularé para ello el correlación lineal que me indicará el grado de interdependencia de las 2 variables. Su expresión viene dada por:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$\sigma_{xy} = \mu_{11} = \sum_{i=1}^{4}\sum_{j=1}^{5} f_{ij}(c_i - \bar{x})(c_j - \bar{y}) = \frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}(c_i - \bar{x})(c_j - \bar{y})$$

Construiré una tabla más completa y calcularé las medias:

X/Y	(10,20]	(20,25]	(25,30]	(30,35]	(35,40]	$n_{i.}$	c_i
(15,18]	3	2	3	0	0	8	16.5
(18,21]	0	4	2	2	0	8	19.5
(21,24]	0	7	10	6	1	24	22.5
(24,27]	0	0	2	5	3	10	25.5
$n_{.j}$	3	13	17	13	4	50	
c_j	15	22.5	27.5	32.5	37.5		

$$\bar{x} = \frac{1}{n} c_i n_{i}. = \frac{16.5 \times 8 + 19.5 \times 8 + 22.5 \times 24 + 25.5 \times 10}{50} = 21.66$$

$$\bar{y} = \frac{1}{n} c_j n_{.j} = \frac{15 \times 3 + 22.5 \times 13 + 27.5 \times 17 + 32.5 \times 13 + 37.5 \times 4}{50} = 27.55$$

$$\sigma_{xy} = \frac{1}{50}(194.274 + 52.116 + 0.774 + 43.632 + 0.216 - 21.384 - 29.694 - 0.42 + 24.948 +$$
$$8.358 - 0.384 + 95.04 + 114.624) = 9.642$$

$$\sigma_x = \sqrt{\mu_{20}} = \sqrt{\frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}(c_i - \bar{x})^2} = 2.88$$

$$\sigma_y = \sqrt{\mu_{02}} = \sqrt{\frac{1}{n}\sum_{i=1}^{4}\sum_{j=1}^{5} n_{ij}(c_j - \bar{y})^2} = 5.51$$

$$r = \frac{9.642}{2.88 \times 5.51} = 0.6076$$

Tienen por tanto una interdependencia lineal del 60.76 %

2.10. Ejercicio 10:

Calcular el coeficiente de correlación lineal de dos variables cuyas rectas de regresión son:

$$x + 4y = 1$$

$$x + 5y = 2$$

Como no sabemos cual es la recta de regresión lineal Y /X y cual es la X/Y , supondremos que la primera es la que explica la variable x en función de la variable y (X/Y) y la segunda la que explica la variable y en función de la variable x y (Y/X).
Sabemos que

$$x = \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2}\bar{y} + \frac{\sigma_{xy}}{\sigma_y^2}y$$

$$y = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} + \frac{\sigma_{xy}}{\sigma_x^2}x$$

Por lo tanto, despejando de las ecuaciones tenemos:

$$x + 4y = 1 \;\Rightarrow\; x = 1 - 4y \;\Rightarrow\; \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2}\bar{y} = 1 \;\; \frac{\sigma_{xy}}{\sigma_y^2} = -4$$

$$x + 5y = 2 \;\Rightarrow\; y = 0.4 - 0.2x \;\Rightarrow\; \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} = 0.4 \;\; \frac{\sigma_{xy}}{\sigma_x^2} = -0.2$$

De estos resultados saco que $\sigma_{xy} < 0 \Rightarrow r < 0$
Trabajando con la expresión de r^2 tenemos:

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x} \times \frac{\sigma_{xy}}{\sigma_y} = (-4) \times (-0.2) = 0.8$$

Por lo tanto

$$r = -\sqrt{r^2} = -0.8944 \quad \Rightarrow \quad \text{Tienen una relación inversa del } 89.44\,\%$$

2.11. Ejercicio 11:

Consideremos una distribución bidimensional en la que la recta de regresión de Y sobre X es: $y = 5x - 20$, y $\sum y_j^2 n_{.j} = 3240$. Supongamos, además, que la distribución marginal de X es:

x_i	3	5	8	9
$n_{i.}$	5	1	2	1

Determinar la recta de regresión de X sobre Y, y la bondad de los ajustes lineales.

La recta de regresión X sobre Y tiene la forma $x = a + by$, con $a = \bar{x} + b\bar{y}$ y con $b = \frac{\sigma_{xy}}{\sigma_y^2}$, por lo que los datos que necesitamos son $\bar{x},\ \bar{y},\ \sigma_{xy}$ y σ_y^2

Primero, calculamos las medias. La media de x se obtiene de la tabla:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{4} n_{i.}x_i = \frac{3\times 5 + 5\times 1 + 8\times 2 + 9\times 1}{9} = 5$$

Sabiendo que ambas rectas de regresión pasan por el punto $(\bar{x}, \bar{y})$ deducimos que $\bar{y} = 5\bar{x} - 20 = 5\times 5 - 20 = 5$.

Teniendo ya la media de x, podemos calcular la varianza de x:

$$\sigma_x^2 = \frac{1}{9}\sum_{i=1}^{4} n_{i.}(x_i - \bar{x})^2 = 6$$

De la recta de regresión Y/X, sabemos que $5 = \frac{\sigma_{xy}}{\sigma_x^2}$ y por tanto:

$$\sigma_{xy} = 5\sigma_x^2 = 5\times 6 = 30$$

Haciendo uso del dato restante, calculamos la varianza de y:

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{4} n_{.j}(y_j - \bar{y})^2 = \frac{1}{9}\sum_{i=1}^{4} n_{.j}(y_j^2 - 2y_j\bar{y} + \bar{y}^2) =$$

$$= \frac{1}{9}(\sum_{i=1}^{4} n_{.j}y_j^2 - 2\sum_{i=1}^{4} n_{.j}y_j\bar{y} + \sum_{i=1}^{4}\bar{y}^2) = \frac{1}{9}\sum_{i=1}^{4} y_j^2 n_{.j} - \bar{y}^2 = \frac{3240}{9} - 25 = 335$$

Ya finalmente calculamos la recta de regresión:

$$b = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{30}{335} = 0.0896 \qquad a = \bar{x} - b\bar{y} = 5 - 0.0896\times 5 = 4.55$$

y = 4.55 + 0.09x

2.12. Ejercicio 12:

De las estadísticas de ”Tiempos de vuelo y consumos de combustible”de una compañía aérea, se han obtenido datos relativos a 24 trayectos distintos realizados por el avión DC-9. A partir de esos datos se han obtenido las siguientes medidas:

$$\sum y_i = 219.719 \qquad \sum {y_i}^2 = 2396.504 \qquad \sum x_i y_i = 349.486$$
$$\sum x_i = 31.470 \qquad \sum {x_i}^2 = 51.075 \qquad \sum {x_i}^2 y_i = 633.993$$
$$\sum {x_i}^4 = 182.977 \qquad \sum {x_i}^3 = 93.6$$

La variable Y expresa el consumo total de combustible, en miles de libras, correspondiente a un vuelo de duración X (el tiempo se expresa en horas, y se utilizan como unidades de orden inferior fracciones decimales de la hora):

a) Ajustar un modelo del tipo $Y = aX + b$. ¿Qué consumo total se estimaría para un programa de vuelos compuesto de 100 vuelos de media hora, 200 de una hora y 100 de dos horas? ¿Es fiable esta estimación?

b) Ajustar un modelo del tipo $Y = a + bX + cX^2$. ¿Qué consumo total se estimaría para el mismo programa de vuelos del apartado a)?

c) ¿Cuál de los dos modelos se ajusta mejor? Razonar la respuesta.

Empezamos calculando los momentos que se puede ver, que son estas expresiones divididas entre el numero de vuelos estudiados: (cabe destacar que $n_{ij} = 1$ para cualesquiera i, j.

$$m_{10} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i = \frac{1}{24}31.470 = 1.31125$$

$$m_{20} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i^2 = \frac{1}{24}51.075 = 2.128125$$

$$m_{21} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i^2 y_j = \frac{1}{24}633.993 = 26.416375$$

$$m_{01} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}y_j = \frac{1}{24}219.719 = 9.154958333$$

$$m_{02} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}y_j^2 = \frac{1}{24}2396.504 = 99.85433333$$

$$m_{11} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i y_j = \frac{1}{24}349.486 = 14.56191667$$

$$m_{30} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i^3 = \frac{1}{24}93.6 = 4.0125$$

$$m_{40} = \sum_{i=1}^{k}\sum_{j=1}^{p}\frac{n_{ij}}{n}x_i^4 = \frac{1}{24}182.977 = 7.624041667$$

$$\sigma_{xy} = \mu_{11} = m11 - m10m01 = 14.56191667 - 1.31125 * 9.154958333 = 2.557477556$$

$$\sigma_x^2 = \sum_{i=1}^{k}\frac{n_{i.}}{n}x_i^2 - \bar{x}^2 = m20 - m10^2 = 2.128125 - 1.31125^2 = 0.4087484375$$

$$\sigma_y^2 = \sum_{j=1}^{p}\frac{n_{.j}}{n}y_j^2 - \bar{y}^2 = m02 - m01^2 = 99.85433333 - 9.154958333^2 = 16.041071251$$

a)Una recta de regresión lineal es d la forma $y = ax + b$ con

$$a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{2.557477556}{0.4087484375} = 6.2568497427$$

$$b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} = 9.154958333 - 6.2568497427 * 1.31125 = 0.95066410788$$

entonces $y = 6.2568497427x + 0.95066410788$

$$400f(30) = 400 * 188.656156389 = 75462.4625556$$

$$200f(60) = 200 * 376.36164867 = 75272.329734$$

$$100f(120) = 100 * 751.772633232 = 75177.2633232$$

es bastante fiable pues estos resultados se parecen mucho, pero vamos a hacer el coeficiente de estimación:

$$r = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{2.557477556}{\sqrt{0.4087484375}\sqrt{16.041071251}} = 0,9987537291$$

los resultados son prácticamente idénticos como podemos observar.

b) Una parábola de la forma $y = a + bx + cx^2$, lo haremos obviamente por el método de mínimos cuadrados, debemos minimizar la siguiente función:

$$\sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j - f(x_i))^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j - a - bx_i - cx_i^2)^2$$

hagamos la derivada parcial respecto de cada variable a, b y c, y igualamos a cero para averiguar la función mínima.

$\phi_{(a,b,c)} = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j - a - bx_i - cx_i^2)^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j - a - bx_i - cx_i^2)^2 =$
$\sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j^2 - ay_j - bx_iy_j - cx_i^2y_j - ay_j + a^2 + abx_i + acx_i^2 - bx_iy_j + abx_i + b^2x_i^2 + bcx_i^3 - y_jcx_i^2 + acx_i^2 + bcx_i^3 + c^2x_i^4) =$
$\sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j^2 - 2ay_j + a^2 + 2abx_i - 2bx_iy_j - 2cx_i^2y_j + 2acx_i^2 + b^2x_i^2 + 2bcx_i^3 + c^2x_i^4)$

$$\frac{\partial\phi_{(a,b,c)}}{\partial a} = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j^2 - 2y_j + 2a + 2bx_i - 2bx_iy_j - 2cx_i^2y_j + 2cx_i^2 + b^2x_i^2 + 2bcx_i^3 + c^2x_i^4)$$

$$\frac{\partial\phi_{(a,b,c)}}{\partial b} = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j^2 - 2ay_j + a^2 + 2ax_i - 2x_iy_j - 2cx_i^2y_j + 2acx_i^2 + 2bx_i^2 + 2cx_i^3 + c^2x_i^4)$$

$$\frac{\partial\phi_{(a,b,c)}}{\partial c} = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}(y_j^2 - 2ay_j + a^2 + 2abx_i - 2bx_iy_j - 2x_i^2y_j + 2ax_i^2 + b^2x_i^2 + 2bx_i^3 + 2cx_i^4)$$

Nos queda entonces el sistema de ecuaciones:

$$2m_{01} + 2a + 2bm_{10} + 2cm_{20} = 0 \rightarrow a + 1,3113b + 2,1281c = 9,155$$

$$2m_{11} + 2am_{10} + 2bm_{20} + 2cm_{30} = 0 \rightarrow 1,3113a + 2,1281b + 3,9c = 14,5619$$

$$2m_{21} + 2am_{20} + 2bm_{30} + 2cm_{20} \rightarrow 2,1281a + 3,9b + 7,624c = 26,4139$$

$$\left.\begin{array}{rcl} a + 1,3113b + 2,1281c &=& 9,155 \\ 1,3113a + 2,1281b + 3,9c &=& 14,5619 \\ 2,1281a + 3,9b + 7,624c &=& 26,4139 \end{array}\right\}$$

$$\left.\begin{array}{rcl} a + 1,3113b + 2,1281c &=& 9,155 \\ 1,3113a + 2,1281b + 3,9c &=& 14,5619 \\ 2,1281a + 3,9b + 7,624c &=& 26,4139 \end{array}\right\}\left\{\begin{array}{ll} a & = 0,7491 \\ b & = 6,6368 \\ c & = -0,1395 \end{array}\right.$$

Entonces $y = 0,7491 + 6,6368x - 0,1395x^2$.

$$400f(30) = 400 * 188.656156389 = 75462.4625556$$

$$200f(60) = 200 * 376.36164867 = 75272.329734$$

$$100f(120) = 100 * 751.772633232 = 75177.2633232$$

c) Es claro ver que en el segundo modelo lo único que se añade es un insignificante $c = -0,1395$, y el resto de la ecuación es casi idéntica, pero no ganamos casi nada de precisión a cambio de una gran complicación en los cálculos.

2.13. Ejercicio 13:

La curva de Engel, que expresa el gasto en un determinado bien en función de la renta, adopta en ocasiones la forma de una hipérbola equilátera. Ajustar dicha curva a los siguientes datos, en los que X denota la renta en miles de euros e Y el gasto en euros. Cuantificar la bondad del ajuste:

X	10	12.5	20	25
Y	50	90	160	180

para hacer ajuste de hipérbola equilátera cambiamos el dato X por z=1/x, así:
$y = a\frac{1}{x} + b \rightarrow y = az + b$, y aplicamos regresión lineal.

	x	y	z
	10	50	0,1
	12,5	90	0,08
	20	160	0,05
	25	180	0,04
media:	16,875	120	0,0675
desv tipica	35,546875	2750	0,00056875
var	5,962120009	52,44044241	0,02384848004

Covarianza:

$$\sigma_{zy} = m11 - m10m01 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}z_iy_j - (\bar{z}\bar{y}) = 6,85 - 120 * 0,0675 = -1,25$$

Entonces la pendiente de la recta sería:

$$\frac{\sigma_{zy}}{\sigma_z^2} = \frac{-1,25}{0,00056875} = -2197,802198$$

$y = az + b$, con $a = -2197,802198$ y $b = \frac{\sigma_{zy}}{\sigma_z^2}\bar{y} - \bar{z} = \frac{-2197.802198}{0.00056875}120 - 0.0675 = 268,3516484$
sustituimos, entonces $y = -2197,802198z + 268,3516484 \rightarrow y = \frac{-2197,802198}{x} + 268,3516484$
Veamos ahora como de bueno es este ajuste, calculemos los residuos:

$$\sigma_{ry}^2 = \frac{1}{4}\sum_{i=1}^{4} n_i(f(x_i) - y))^2 = 2,237048666$$

Veamos entonces el coeficiente de determinación:

$$\frac{\sigma_{ey}^2}{\sigma_y^2} = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = 1 - \frac{2,237048666}{2750} = 0,9991865278$$

es prácticamente perfecto, casi igual a 1. Porque los residuos son casi despreciables

2.14. Ejercicio 14:

Se dispone de la siguiente información referente al gasto en espectáculos (Y, en euros) y la renta disponible mensual (X, en cientos de euros) de 6 familias:

Y	30	50	70	80	120	140
X	9	10	12	15	22	32

Explicar el comportamiento de Y por X mediante:

a) Relación lineal.

b) Hipérbola equilátera.

c) Curva potencial.

d) Curva exponencial.

¿Qué ajuste es más adecuado?

a) Para el ajuste lineal, el más sencillo, realizaremos un ajuste por mínimos cuadrados a la distribución dada realizando primero los cálculos de los coeficientes necesarios:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{6} n_{i.}x_i = \frac{9+10+12+15+22+32}{6} = 16.67$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{6} n_{.j}y_j = \frac{30+5+70+80+120+140}{6} = 81.67$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{6} n_{i.}(x_i-\bar{x})^2 = 65.22$$

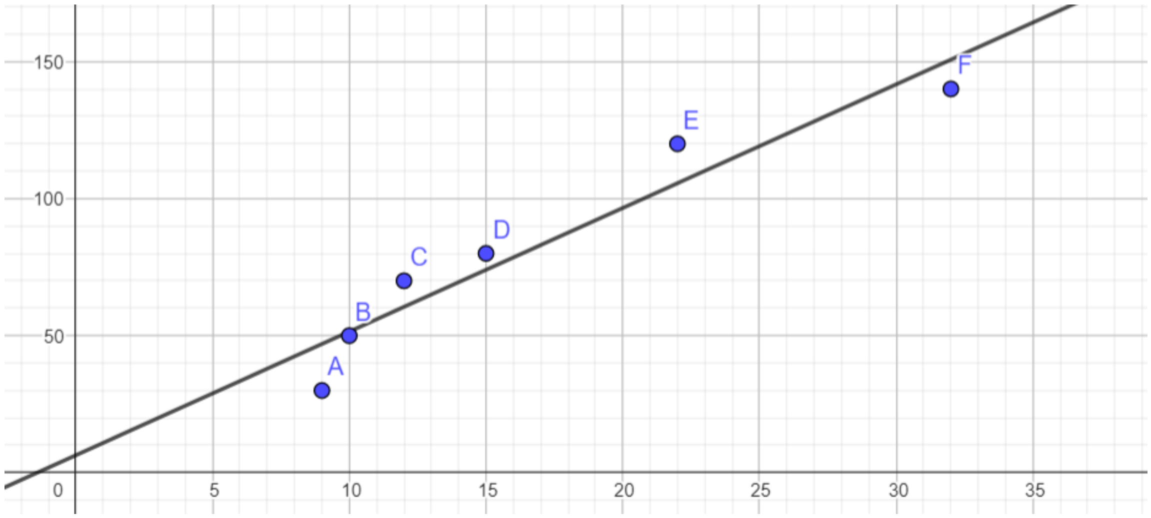
$$\sigma_y^2 = \frac{1}{n}\sum_{j=1}^{6} n_{.j}(y_j-\bar{y})^2 = 1447.22$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{6}\sum_{j=1}^{6} n_{ij}(x_i-\bar{x})(y_j-\bar{y}) = 293.89$$

Por tanto, los coeficientes de la recta de regresión $y = a + bx$ son:

$$b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{293.89}{65.22} = 4.51$$

$$a = \bar{y} - b\bar{x} = 81.67 - 4.51 \times 16.67 = 6.49$$



Ajuste lineal ($y = 6.49 + 4.51x$)

b) Para el ajuste hiperbólico, realizaremos un cambio de variable $z = \frac{1}{x}$ de manera que la ecuación quede como $y = a + \frac{b}{x} = a + bz$. Así, el problema se reduce a hacer un ajuste lineal entre z e y para obtener los parámetros a y b.

Y	30	50	70	80	120	140
$z = \frac{1}{x}$	0.111	0.1	0.083	0.067	0.045	0.031

Ahora calculamos la media, varianza y covarianza:

$$\bar{z} = \bar{x} = \frac{1}{n}\sum_{i=1}^{6} n_{i.} z_i = \frac{0.111 + 0.1 + 0.083 + 0.067 + 0.045 + 0.031}{6} = 0.073$$

$$\sigma_z^2 = \frac{1}{n}\sum_{i=1}^{6} n_{i.}(z_i - \bar{z})^2 = 0.000805$$

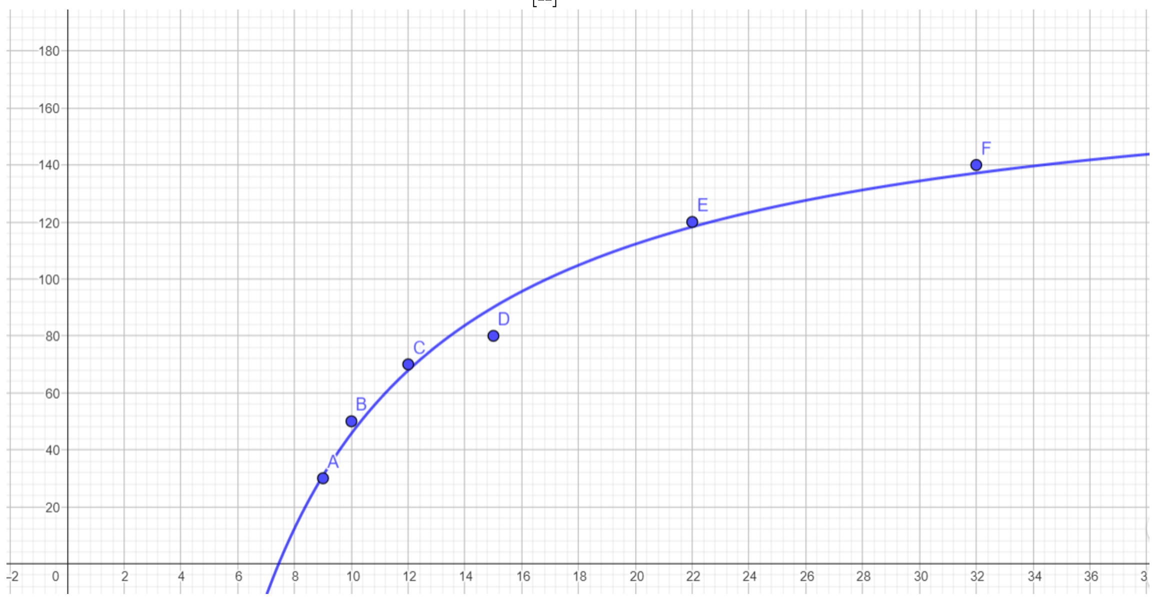
$$\sigma_{zy} = \frac{1}{n}\sum_{i=1}^{6}\sum_{j=1}^{6} n_{ij}(z_i - \bar{z})(y_j - \bar{y}) = -1.071$$

Calculamos los parámetros de la hipérbola:

$$b = \frac{\sigma_{zy}}{\sigma_z^2} = \frac{-1.071}{0.000805} = -1330.27$$

$$a = \bar{y} - b\bar{z} = 81.67 - (-1330.43) \times 0.073 = 178.73$$

[h]



Ajuste hiperbólico ($y = 178.73 - \frac{1330.27}{x}$)

c) Para el ajuste potencial es necesario tomar logaritmos a ambos lados de la ecuación. De manera que la ecuación $y = ax^b$ se nos queda como $ln(y) = ln(a) + b\ ln(x)$.
Ahora tomamos la ecuación $Y = A + b\ X$, donde $Y = ln(y),\ A = ln(a)$ y $X = ln(X)$, con lo que el problema se nos queda en un ajuste lineal.

$Y = ln(y)$	3.4	3.91	4.24	4.38	4.79	4.94
$X = ln(x)$	2.2	2.3	2.485	2.71	3.09	3.47

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{6} n_{i.}X_i = \frac{2.2 + 2.3 + 2.485 + 2.71 + 3.09 + 3.47}{6} = 2.71$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{6} n_{.j}y_j = \frac{3.4 + 3.91 + 4.24 + 4.38 + 4.79 + 4.94}{6} = 4.28$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{6} n_{i.}(X_i - \bar{X})^2 = 1.196$$

$$\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^{6} n_{.j}(Y_j - \bar{Y})^2 = 1.61$$

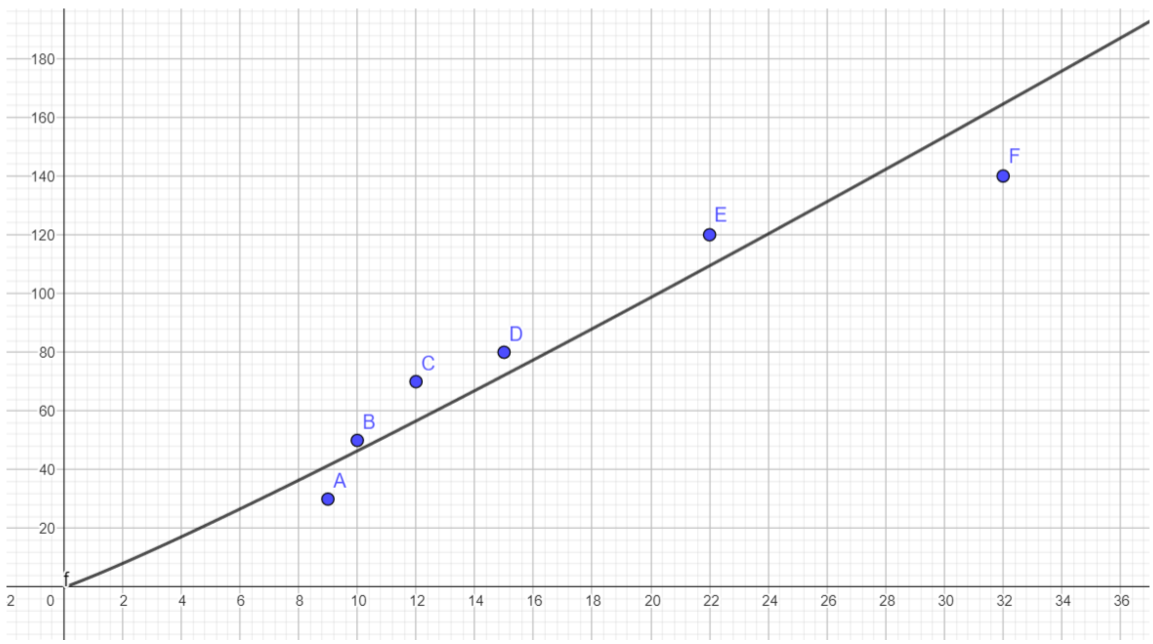
$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{6}\sum_{j=1}^{6} n_{ij}(X_i - \bar{X})(Y_j - \bar{Y}) = 1.3$$

Por tanto, los coeficientes de la recta de regresión $y = a + bx$ son:

$$b = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{1.3}{1.196} = 1.088$$

$$A = \bar{Y} - b\bar{X} = 4.28 - 4.28 \times 2.71 = 1.33$$

Teniendo en cuenta que $a = e^A = e^1.33 = 3.79$ el ajuste queda así:



Ajuste potencial ($y = 3.79\ x^{1.088}$)

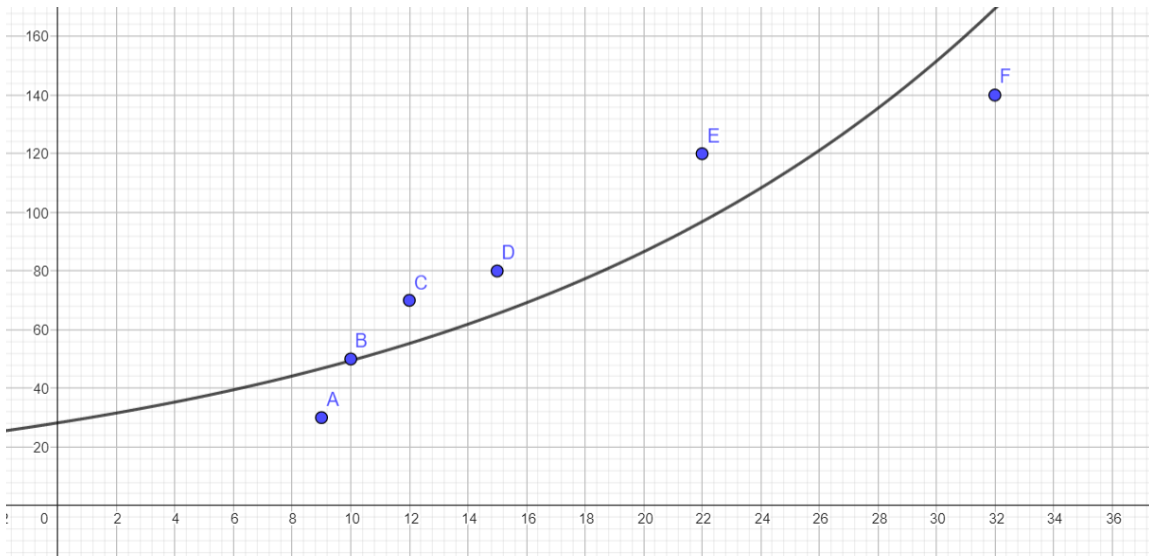
d) Al igual que para el ajuste potencial, para la curva exponencial tomaremos logaritmo a ambos lados de la ecuación de manera que nos queda $Y = A + bx$, donde $Y = ln(y)$ y $A = ln(a)$. Todos los coeficientes necesarios ya están calculados para este ajuste, salvo σ_{xY}.

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^{6} \sum_{j=1}^{6} n_{ij}(x_i - \bar{X})(Y_j - \bar{Y}) = 3.67$$

$$b = \frac{\sigma_{xY}}{\sigma_x^2} = \frac{3.67}{65.22} = 0.056$$

$$A = \bar{Y} - b\bar{x} = 4.28 - 0.056 \times 16.67 = 3.34$$

Teniendo en cuenta que $a = e^A = e^{3.34} = 28.25$ el ajuste queda así:



Ajuste exponencial ($y = 28.25\ e^{0.056x}$)

Para ver qué ajuste es mejor de todos, nos apoyaremos en los respectivos coeficientes de correlación.

$$Lineal: R^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{293.89^2}{65.22 \times 1447.22} = 0.915$$

$$Hiperbólico: R^2 = \frac{\sigma_{zy}^2}{\sigma_z^2 \sigma_y^2} = \frac{(-1.07)^2}{0.000805 \times 1447.22} = 0.983$$

Para los modelos potencial y exponencial no podemos usar la fórmula de R^2 pues no son lineales en los parámetros. En su lugar tendremos que usar las respectivas varianzas residuales:

$$Potencial: \sigma_{ry} = \frac{1}{n} \sum_{i=1}^{6} \sum_{j=1}^{6} n_{i.} (f(x_i) - y_j)^2 = 182.5$$

$$\eta^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = 1 - \frac{182.5}{1447.22} = 0.874$$

$$Exponencial: \sigma_{ry} = \frac{1}{n} \sum_{i=1}^{6} \sum_{j=1}^{6} n_{i.} (f(x_i) - y_j)^2 = 361.75$$

$$\eta^2 = 1 - \frac{\sigma_{ry}^2}{\sigma_y^2} = 1 - \frac{361.75}{1447.22} = 0.75$$

Con esto concluimos que el mejor ajuste de todos es el hiperbólico, con un coeficiente de correlación de 0.983, lo cual ya podíamos intuirlo al ver que la gráfica pasaba muy cerca de todos los puntos.

Por contra, el peor ajuste de todos es el exponencial, con un coeficiente de correlación de 0.75, que también podíamos intuirlo, pues la concavidad con la que están dispuestos es opuesta a la concavidad de una función exponencial.

3. Relación 3

3.1. Ejercicio 1:

Durante un año, las personas de una ciudad utilizan 3 tipos de transportes: metro (M), autobús (A), y coche particular (C). Las probabilidades de que durante el año hayan usado unos u otros transportes son:
M: 0.3 A: 0.2 C: 0.15 M y A: 0.1 M y C: 0.05 A y C: 0.06 M, A y C: 0.01
Calcular las probabilidades siguientes:
a) que una persona viaje en metro y no en autobús.
b) que una persona tome al menos dos medios de transporte;
c) que una persona viaje en metro o en coche, pero no en autobús;
d) que viaje en metro, o bien en autobús y en coche;
e) que una persona vaya a pie.
Solución:

a) $P(M-A) = P(M) - P(M\cap A) = 0.3 - 0.1 = 0.2$

b) $P((A\cap C)\cup(A\cap M)\cup(C\cap M)\cup(A\cap C\cap M)) =$
$P(((A\cap C)\cup(A\cap M)) + P((C\cap M)\cup(A\cap C\cap M)) - P(((A\cap C)\cup(A\cap M))\cap((C\cap M)\cup(A\cap C\cap M))) =$
$P(A\cap C) + P(A\cap M) - P(((A\cap C)\cap(A\cap M)) + P((C\cap M) + P(A\cap C\cap M)) - P((C\cap M)\cap$
$(A\cap C\cap M) - P(A\cap C\cap M) =$
$\underbrace{P(A\cap C)}_{0.06} + \underbrace{P(A\cap M)}_{0.1} - \underbrace{P(A\cap C\cap M)}_{0.01} + \underbrace{P((C\cap M)}_{0.05} + \underbrace{P(A\cap C\cap M)}_{0.01} - \underbrace{P(A\cap C\cap M)}_{0.01} - \underbrace{P(A\cap C\cap M)}_{0.01} =$
0.19

c) $P((C\cup M)-A) = P(C\cup M) - P((C\cup M)\cup A) = P(C) + P(M) - P(C\cap M) - P((C\cup M)\cap A) \overset{*}{=}$
$\underbrace{P(C)}_{0.15} + \underbrace{P(A)}_{0.3} - \underbrace{P(C\cap M)}_{0.05} - \underbrace{P((C\cup M)\cap A)}_{0.15} = 0.25$
$(*) = P((C\cup M)\cap A) = P((C\cap A)\cup(M\cap A)) = P(C\cap A) + P((M\cap A) - P((C\cap A)\cap(M\cap A)) =$
$\underbrace{P(C\cap A)}_{0.06} + \underbrace{P((M\cap A)}_{0.1} - \underbrace{P(C\cap M\cap A)}_{0.01} = 0.15$

d) $P(M\cup(A\cap C)) = \underbrace{P(M)}_{0.3} + \underbrace{P(A\cap C)}_{0.06} - \underbrace{P(M\cup A\cup C)}_{0.01} = 0,35$

e) $P(A\cup M\cup C) =$
$\underbrace{P(A)}_{0.2} + \underbrace{P(M)}_{0.3} + \underbrace{P(C)}_{0.15} - \underbrace{P(A\cap M)}_{0.1} - \underbrace{P(A\cap C)}_{0.06} - \underbrace{P(M\cap C)}_{0.05} + \underbrace{P(M\cup A\cup C)}_{0.01} = 0.45$, entonces
$\underbrace{P(\Omega)}_{1} - \underbrace{P(A\cup M\cup C)}_{0.45} = 0,55$

3.2. Ejercicio 2:

Sean A, B y C tres sucesos de un espacio probabilístico (ω, A, P), tales que P(A)=0,4, P(B) = 0,2, P(C)=0,3, $(A \cap B) = 0,1$ y $(A \cup B) \cap C = \emptyset$. Calcular las probabilidades de los siguientes sucesos:

a) sólo ocurre A,
Esto será $A \cap \bar{B} \cap \bar{C}$ por lo que tendré que calcular $P(A \cap \bar{B} \cap \bar{C})$. Por la definición de complementario y con los axiomas tenemos que

$$P(A \cap \bar{B} \cap \bar{C}) = P((A - B) - C) = P(A - B) - P(C \cap (A - B))$$

$$P(A - B) = P(A) - P(A \cap B) = 0,4 - 0,1 = 0,3$$
$$P(C \cap (A - B)) = P((C \cap A) - (C \cap B)) = P(C \cap A) - P((C \cap A) \cap (C \cap B))$$

Aplicando algunas propiedades tenemos

$$(A \cup B) \cap C = \emptyset \Rightarrow (A \cap C) \cup (B \cap C) = \emptyset \Rightarrow (A \cap C) = \emptyset \text{ , } (B \cap C) = \emptyset$$

Y por tanto
$$P(C \cap A) - P((C \cap A) \cap (C \cap B)) = P(\emptyset) - P(\emptyset \cap \emptyset) = 0$$

Concluyendo

$$P(A \cap \bar{B} \cap \bar{C}) = 0,3 - 0 = 0,3 \Rightarrow \text{La probabilidad es del } 30\,\%$$

b) ocurren los 3 sucesos,
Si esto ocurre estaremos calculando la probabilidad relativa a $A \cap B \cap C$. Por el apartado anterior sabemos que $B \cap C = \emptyset$ y por lo tanto $A \cap B \cap C = A \cap \emptyset = \emptyset$ y, por los axiomas tenemos $P(\emptyset) = 0$ y la probabilidad es del 0 % (es un suceso imposible).

c) ocurren A y B pero no C, Es decir $A \cap B \cap \bar{C}$

$$P(A \cap B \cap \bar{C}) = P(A \cap (B - C)) = P((A \cap B) - (A \cap C)) = P(A \cap B) - P((A \cap B) \cap \underbrace{(A \cap C)}_{\emptyset}) =$$

$$P(A \cap B) - P(\emptyset) = 0,1 - 0 = 0,1 \Rightarrow \text{La probabilidad es del } 10\,\%$$

d) por lo menos dos ocurren,
Las combinaciones de que ocurran al menos 2 son:

a) A y B
b) A y C
c) B y C
d) A y B y C

Esto nos lleva a calcular

$$P((A \cap B) \cup \underbrace{(A \cap C)}_{\emptyset} \cup \underbrace{(B \cap C)}_{\emptyset} \cup \underbrace{(A \cap B \cap C)}_{\emptyset})) = P(A \cap B) = 0,1 \Rightarrow \text{La probabilidad es del } 10\,\%$$

e) ocurren dos y no más
Ahora tenemos las mismas combinaciones de antes excepto la *d)* y por tanto calculamos

$$P((A \cap B) \cup \underbrace{(A \cap C)}_{\emptyset} \cup \underbrace{(B \cap C)}_{\emptyset}) = P(A \cap B) = 0,1 \Rightarrow \text{La probabilidad es del } 10\,\%$$

f) no ocurren más de dos,
Esta opción incluye a todos los casos a excepción de que ocurran A,B y C simultáneamente, es decir, la probabilidad se podría calcular como:

$$P(\overline{A \cap B \cap C}) = 1 - \underbrace{P(A \cap B \cap C)}_{0} = 1$$

ya que al ser C disjunto con $A \cup B$ lo será tanto con A como con B como con $A \cap B$.

g) ocurre por lo menos uno,
Esto incluye:

a) A
b) B
c) C
d) A y B
e) A y C
f) B y C
g) A y B y C

Lo cual son las posibilidades del apartado anterior sumado a las posibilidades de obtener los tres que, como ya hemos visto, son nulas por lo que

$$P(A \cup B \cup C \cup \underbrace{(A \cap B)}_{(A \cap B) \subset A} \cup \underbrace{(A \cap C)}_{\emptyset} \cup \underbrace{(B \cap C)}_{\emptyset} + \underbrace{(A \cap B \cap C)}_{\emptyset}) = P(A \cup B \cup C) =$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - \underbrace{P(A \cap C)}_{0} - \underbrace{P(B \cap C)}_{0} + \underbrace{P(A \cap B \cap C)}_{0} =$$

$$= 0,4 + 0,2 + 0,3 - 0,1 - 0 - 0 - 0 = 0,8 \Rightarrow \text{La probabilidad es del } 80\,\%$$

h) ocurre sólo uno,
Esto es $(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)$. Aplicando las propiedades:

$$P((A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)) = P((A \cap (\bar{B} \cap \bar{C})) \cup (B \cap (\bar{A} \cap \bar{C})) \cup (C \cap (\bar{A} \cap \bar{B}))) =$$

$$P((A \cap (\overline{B \cup C})) \cup (B \cap (\overline{A \cup C})) \cup (C \cap (\overline{A \cup B}))) = P((A - (B \cup C)) \cup (B - (A \cup C)) \cup (C - (A \cup B))) =$$

$$P(((A - B) \cap \underbrace{(A - C)}_{A}) \cup ((B - A) \cap \underbrace{(B - C)}_{B}) \cup (\underbrace{(C - A)}_{C} \cap \underbrace{(C - B)}_{C})) =$$

$$= P((A - B) \cup (B - A) \cup (C)) = P(A - B) + P(B - A) + P(C)$$

Calculando cada sumando por separado

$$P(A - B) = P(A) - P(A \cap B) = 0,4 - 0,1 = 0,3$$

$$P(B - A) = P(B) - P(A \cap B) = 0,2 - 0,1 = 0,1$$

Y nos queda

$$0,3 + 0,1 + 0,3 = 0,7 \Rightarrow \text{La probabilidad es del } 70\,\%$$

i) no ocurre ninguno.

$$P(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - P(A \cup B \cup C) =$$

$$= 1 - (P(A) + P(B) + P(C) - P(A \cap B) - \underbrace{P(A \cap C)}_{0} - \underbrace{P(B \cap C)}_{0} + \underbrace{P(A \cap B \cap C)}_{0}) =$$

$$= 1 - (0,4 + 0,2 + 0,3 - 0,1) = 1 - 0,8 = 0,2 \Rightarrow \text{La probabilidad es del } 20\,\%$$

3.3. Ejercicio 3:

Se sacan dos bolas sucesivamente sin devolución de una urna que contiene 3 bolas rojas distinguibles y 2 blancas distinguibles.

a) Describir el espacio de probabilidad asociado a este experimento.

b) Descomponer en sucesos elementales los sucesos: *la primera bola es roja, la segunda bola es blanca* y calcular la probabilidad de cada uno de ellos.

c) ¿Cuál es la probabilidad de que ocurra alguno de los sucesos considerados en el apartado anterior?

Solución:

a) El espacio muestral sería:

$\Omega = \{b_i b_j;\ i,j = 1,2;\ i \neq j\} \cup \{r_i b_j;\ i = 1,2,3;\ j = 1,2\} \cup \{b_i r_j;\ i = 1,2;\ j = 1,2,3\} \cup \{r_i r_j;\ i,j = 1,2,3\ i \neq j\}$
$|\Omega| = 2 + 6 + 6 + 6 = 20$

b) La primera bola es roja:

$R1 = \{r_i b_j;\ i = 1,2,3;\ j = 1,2\} \cup \{r_i r_j;\ i,j = 1,2,3;\ i \neq j\}$

$$P(R1) = \frac{|R1|}{|\Omega|} = \frac{12}{20} = \frac{3}{5}$$

La segunda bola es blanca:

$B2 = \{r_i b_j;\ i = 1,2,3;\ j = 1,2\} \cup \{b_i b_j;\ i,j = 1,2;\ i \neq j\}$

$$P(B2) = \frac{|B2|}{|\Omega|} = \frac{8}{20} = \frac{2}{5}$$

c) La probabilidad sería 1 menos la probabilidad del suceso complementario, es decir, de que la primera sea blanca y la segunda roja:

$R1 \cup B2 = \{r_i b_j;\ i = 1,2,3;\ j = 1,2\} \cup \{r_i r_j;\ i,j = 1,2,3;\ i \neq j\} \cup \{b_i b_j;\ i,j = 1,2;\ i \neq j\}$

$$P(R1 \cup B2) = \frac{|R1 \cup B2|}{|\Omega|} = \frac{14}{20} = \frac{7}{10}$$

3.4. Ejercicio 4:

Una urna contiene a bolas blancas y b bolas negras. ¿Cuál es la probabilidad de que al extraer dos bolas simultáneamente sean de distinto color?

Sean a (bolas blancas) y b (bolas negras), entonces sean $A, B \in \mathit{\Lambda}$, los sucesos referidos a que la bola sea blanca o negra, respectivamente.

$P(A) = \frac{casos\ favorables}{casos\ posibles} = \frac{a}{a+b}$ y, análogamente: $P(B) = \frac{casos\ favorables}{casos\ posibles} = \frac{b}{a+b}$

Por tanto, como las dos bolas se sacan simultáneamente, podemos afirmar que la probabilidad de que salgan ambas es la multiplicación de las mismas: $P(AB) = \frac{a}{a+b} * \frac{b}{a+b-1}$ ya que hay que considerar que en la segunda extracción (aunque diga simultánea se consideran secuenciales en los cálculos), hay una bola menos.

$$= \frac{ab}{a^2 + b^2 + 2ab - a - b}$$

Análogamente, para el caso contrario, que salgan negra y blanca el resultado es $P(BA) = \frac{b}{a+b} * \frac{a}{a+b-1} = \frac{ba}{a^2+b^2+2ab-a-b}$ y por tanto la probabilidad de que suceda alguno de los dos es:

$$P(AB) + P(BA) = \frac{ab}{a^2 + b^2 + 2ab - a - b} + \frac{ba}{a^2 + b^2 + 2ab - a - b} = \frac{2ab}{a^2 + b^2 + 2ab - a - b}$$

3.5. Ejercicio 5:

Una urna contiene 5 bolas blancas y 3 rojas. Se extraen 2 bolas simultáneamente. Calcular la probabilidad de obtener:

a) dos bolas rojas,
Los distintos casos que pueden resultar de sacar 2 bolas rojas se calculan como $\binom{3}{2}$. Las distintas posibilidades de sacar 2 bolas de cualquier color son $\binom{8}{3}$. La probabilidad la puedo calcular como los casos favorables partido de los casos totales por lo que:

$$P(\text{Roja , Roja}) = \frac{\binom{3}{2}}{\binom{8}{2}} = \frac{3}{28} = 0.107$$

b) dos bolas blancas,
Si seguimos el razonamiento anterior los posibles casos de sacar 2 bolas blancas de entre todas las blancas son $\binom{5}{2}$. Por tanto:

$$P(\text{Blanca , Blanca}) = \frac{\binom{5}{2}}{\binom{8}{2}} = \frac{20}{56} = 0.357$$

c) una blanca y otra roja.
Los posibles casos de sacar una bola blanca son $\binom{5}{1}$ y las de sacar una roja son $\binom{3}{1}$. El número total de casos no varía y por tanto:

$$P(\text{Blanca , Roja}) = \frac{\binom{3}{1}\binom{3}{1}\ 2!}{\binom{8}{3}} = \frac{30}{56} = 0.5357$$

3.6. Ejercicio 6:

En una lotería de 100 billetes hay 2 que tienen premio.

a) ¿Cuál es la probabilidad de ganar al menos un premio si se compran 12 billetes?

b) ¿Cuantos billetes habrá que comprar para que la probabilidad de ganar al menos un premio sea mayor que 4/5?

Solución:

a) Empecemos calculando la probabilidad de no ganar ningún premio. La probabilidad de ganar será 1 menos el resultado que obtengamos. Con 1 billete, la probabilidad es de $\frac{98}{100}$. Con 2 la probabilidad será de $\frac{98}{100} \times \frac{97}{99}$. Con 3, $\frac{98}{100} \times \frac{97}{99} \times \frac{96}{98}$ y así sucesivamente obtenemos que la probabilidad de no ganar ningún premio comprando n billetes será de:

$$\frac{98!(100-n)!}{100!(98-n)!} = \prod_{k=0}^{n-1} \frac{98-k}{100-k}$$

Por tanto la probabilidad de ganar un premio comprando n billetes será:

$$P(n) = 1 - \frac{98!(100-n)!}{100!(98-n)!} = 1 - \prod_{k=0}^{n-1} \frac{98-k}{100-k}$$

Entonces comprando 12 billetes la probabilidad será:

$$P(12) = 1 - \prod_{k=0}^{11} \frac{98-k}{100-k} = 0.22667$$

b) Para que la probabilidad sea mayor que 4/5, como P(n) es creciente, buscamos un n tal que:

$$P(n-1) < 4/5 \leq P(n)$$

Vemos que para n=55, se cumple la condición:

$$P(54) = 0.790909 < 0.8 \leq 0.8 = P(55)$$

3.7. Ejercicio 7:

Se consideran los 100 primeros números naturales. Se sacan 3 al azar.

a) Calcular la probabilidad de que en los 3 números obtenidos no exista ningún cuadrado perfecto
b) Calcular la probabilidad de que exista al menos un cuadrado perfecto.
c) Calcular la probabilidad de que exista un sólo cuadrado perfecto, de que existan dos, y la de que los tres lo sean.

Para el transcurso del ejercicio vamos a suponer que no se pueden repetir números. El hecho de que si el 0 es par no influye en ningún apartado así que nos da igual.
Sea $\square$ el suceso referido a sacar un cuadrado perfecto, y sea $P_i(\square)$ la probabilidad asociada al suceso referido a no sacar cuadrado perfecto en la i-ésima sacada de bola.

a) Los cuadrados perfectos son 10 en total, (los cuadrados de los primeros 10 números naturales)
Entonces $P_1(\square) = \frac{90}{100}$
$P_2(\square) = \frac{89}{99}$
$P_3(\square) = \frac{88}{98}$
Podemos afirmar entonces que la probabilidad de que salga algún cuadrado al menos es de
$P(\square) = \frac{90}{100} \times \frac{89}{99} \times \frac{88}{98} = 0.727$

b) $P(\Omega - \square) = P(\Omega) - P(\square) = 1 - 0.727 = 0.273$

c) Empiezo por la Probabilidad de que solo salga 1, calcularé qué probabilidad hay de que salgan 1, 2, o 3 cuadrados perfectos, $P(1\square), P(2\square), P(3\square)$, respectivamente. :

$$P(1\square) = \frac{\binom{90}{2} \times \binom{10}{1}}{\binom{100}{3}} = 0.2477$$

$$P(2\square) = \frac{\binom{90}{1} \times \binom{10}{2}}{\binom{100}{3}} = 0.025$$

$$P(3\square) = \frac{\binom{90}{0} \times \binom{10}{3}}{\binom{100}{3}} = 0.00742$$

3.8. Ejercicio 8:

En una carrera de relevos cada equipo se compone de 4 atletas. La sociedad deportiva de un colegio cuenta con 10 corredores y su entrenador debe formar un equipo de relevos que disputará el campeonato, y el orden en que participarán los seleccionados.

a) ¿Entre cuántos equipos distintos habrá de elegir el entrenador si los 10 corredores son de igual valía? (Dos equipos con los mismos atletas en orden distinto se consideran diferentes)

   Dado que sí importa el orden, no se pueden repetir corredores y que hay 10 corredores posibles a rellenar 4 huecos se trata de un problema de variación sin repetición.

   $$V_{4,10} = \frac{m!}{(m-n)!} = \frac{10!}{(10-4)!} = 5040$$

   Por tanto hay 5040 posibles equipos.

b) Calcular la probabilidad de que un alumno cualquiera sea seleccionado.

   Sabiendo que hay 10 alumnos y solo 4 plazas, las posibilidades de que un alumno cualquiera sea seleccionado serían de

   $$\frac{4}{10} = 0,4 \Rightarrow \text{ Tiene un } 40\,\% \text{ de posibilidades}$$

3.9. Ejercicio 9:

Una tienda compra bombillas en lotes de 300 unidades. Cuando un lote llega, se comprueban 60 unidades elegidas al azar, rechazándose el envío si se supera la cifra de 5 defectuosas. ¿Cuál es la probabilidad de aceptar un lote en el que haya 10 defectuosas?

Solución:

Para calcular la probabilidad de que un lote sea aceptado, calcularemos las probabilidades de que se encuentren 0, 1, 2, 3, 4 y 5 bombillas defectuosas, llamando a dichos sucesos A, B, C, D, E y F respectivamente.

Para el cálculo de cada una de las probabilidades, haremos uso de combinatoria y la regla de Laplace. El número de maneras posibles de seleccionar 60 de entre 300 bombillas es de $\binom{300}{60}$, esos serán los casos posibles para cada suceso.
Los casos favorables del suceso A serán el número de formas posibles de seleccionar 60 bombillas perfectas de entre las 290 que hay, por lo que su probabilidad quedaría:

$$P(A) = \frac{\binom{290}{60}}{\binom{300}{60}} = \frac{290!\ 60!\ 240!}{300!\ 60!\ 230!} = \frac{240 \times 239 \times ... \times 231}{300 \times 299 \times ... \times 291} = \prod_{k=0}^{9} \frac{240-k}{300-k} = 0.103328$$

Para la probabilidad del suceso B hay que tener en cuenta que hay 10 maneras de seleccionar la bombilla defectuosa, para el suceso C hay $\binom{10}{2}$ formas de seleccionarla y así hasta el suceso F, de manera que las probabilidades quedan así:

$$P(B) = \frac{\binom{10}{1}\binom{290}{59}}{\binom{300}{60}} = 10 \times \frac{290!\ 60!\ 240!}{300!\ 59!\ 231!} = 10 \times \frac{60}{291} \prod_{k=0}^{8} \frac{240-k}{300-k} = 0.268383$$

Y de manera similar:

$$P(C) = \frac{\binom{10}{2}\binom{290}{58}}{\binom{300}{60}} = \frac{10 \times 9 \times 60 \times 59}{2 \times 292 \times 291} \prod_{k=0}^{7} \frac{240-k}{300-k} = 0.307137$$

$$P(D) = \frac{\binom{10}{3}\binom{290}{57}}{\binom{300}{60}} = \binom{10}{3} \left(\prod_{k=0}^{2} \frac{60-k}{291+k} \right) \left(\prod_{k=0}^{6} \frac{240-k}{300-k} \right) = 0.203879$$

$$P(E) = \frac{\binom{10}{4}\binom{290}{56}}{\binom{300}{60}} = \binom{10}{4} \left(\prod_{k=0}^{3} \frac{60-k}{291+k} \right) \left(\prod_{k=0}^{5} \frac{240-k}{300-k} \right) = 0.086910$$

$$P(F) = \frac{\binom{10}{5}\binom{290}{55}}{\binom{300}{60}} = \binom{10}{5} \left(\prod_{k=0}^{4} \frac{60-k}{291+k} \right) \left(\prod_{k=0}^{4} \frac{240-k}{300-k} \right) = 0.024852$$

Y la suma quedaría:

$$P(X) = P(A) + P(B) + P(C) + P(D) + P(E) + P(F) = 0.994489$$

Se podría condensar todo en una sola fórmula:

$$P(X) = \sum_{n=0}^{5} \frac{\binom{10}{n}\binom{290}{60-n}}{\binom{300}{60}} = \sum_{n=0}^{5} \left[\binom{10}{n} \left(\prod_{k=0}^{n-1} \frac{60-k}{291+k} \right) \left(\prod_{k=0}^{10-n-1} \frac{240-k}{300-k} \right) \right] = 0.994489$$

3.10. Ejercicio 10:

Una secretaria debe echar al correo 3 cartas; para ello, introduce cada carta en un sobre y escribe las direcciones al azar. ¿Cuál es la probabilidad de que al menos una carta llegue a su destino?

Pongamos el suceso de que alguna carta llegue a su destino, como A_x para cada x variando desde 1 hasta 3. Como todos los sucesos tienen las mismas condiciones, podemos afirmar para todos que:

$P(A_x) = \frac{1}{3}, \forall x = 1, 2, 3$
$P(A_x \cap A_y) = \frac{1}{3} * \frac{1}{2} = \frac{1}{6}, \forall x, y = 1, 2, 3,\ x \neq y$
$P(A_1 \cap A_2 \cap A_3) = \frac{1}{3} * \frac{1}{2} * \frac{1}{1} = \frac{1}{6}$, podemos destacar que si los dos primeros se dan, el tercero es trivial.

Por tanto tenemos que, como hemos visto en clase.

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_3 \cap A_2) + P(A_1 \cap A_2 \cap A_3) =$$

$$= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$$

4. Relación 4

4.1. Ejercicio 1:

En una batalla naval, tres destructores localizan y disparan simultáneamente a un submarino. La probabilidad de que el primer destructor acierte el disparo es 0.6, la de que lo acierte el segundo es 0.3 y la de que lo acierte el tercero es 0.1. ¿Cuál es la probabilidad de que el submarino sea alcanzado por algún disparo?

Cabe destacar que dichos sucesos son independientes, entonces tenemos las siguientes igualdades:

$P(A \cap B) = P(A) * P(B), \forall A, B, C$
$P(A \cap B \cap C) = P(A) * P(B) * P(C)$

Entonces

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) =$
$0.6 + 0.3 + 0.1 - 0.6 * 0.3 - 0.6 * 0.1 - 0.3 * 0.1 + 0.6 * 0.3 * 0.1 = 0,748$

Por tanto podemos comprobar que la probabilidad de que algún barco acierte es de $0,748$

4.2. Ejercicio 2:

Un estudiante debe pasar durante el curso 5 pruebas selectivas. La probabilidad de pasar la primera es 1/6. La probabilidad de pasar la i-ésima, habiendo pasado las anteriores es $1/(7-i)$. Determinar la probabilidad de que el alumno apruebe el curso.

Si denoto por E_i al suceso de aprobar el examen i-ésimo tengo que:

$$P(E_1) = \frac{1}{6}$$

$$P(E_i/E_{i-1} \cap E_{i-2} \cap ... \cap E_1) = \frac{1}{7-i} \ \ \forall i \in 2,3,4,5$$

Con esta notación y aplicando el Teorema de la Probabilidad Compuesta tengo:

$$P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5) = P(E_1)P(E_2/E_1)P(E_3/E_2 \cap E_1)P(E_4/E_3 \cap E_2 \cap E_1)P(E_5/E_4 \cap E_3 \cap E_2 \cap E_1) =$$

$$= \frac{1}{6} \times \frac{1}{5} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{2} = 0.001389 \ \Rightarrow \ \text{Tiene un } 0.1389\,\% \text{ de posibilidades de aprobar}$$

4.3. Ejercicio 3:

En una ciudad, el 40 % de las personas tienen pelo rubio, el 25 % tienen ojos azules y el 5 % el pelo rubio y los ojos azules. Se selecciona una persona al azar. Calcular la probabilidad de los siguientes sucesos:

a) tener el pelo rubio si se tiene los ojos azules,

b) tener los ojos azules si se tiene el pelo rubio,

c) no tener pelo rubio ni ojos azules,

d) tener exactamente una de estas características.

Solución:

Llamemos A al suceso 'tener el pelo rubio' y B al suceso 'tener ojos azules'. En nuestra población, se tiene que $P(A) = 0.4,\ P(B) = 0.25,\ P(A \cap B) = 0.05$. Entonces podemos calcular las probabilidades:

a)
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.05}{0.25} = 0.2$$

b)
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.05}{0.4} = 0.125$$

c)
$$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B) = 1 - 0.4 - 0.25 + 0.05 = 0.4$$

d)
$$P(A \Delta B) = P(A \cup B) - P(A \cap B) = P(A) + P(B) - 2 \times P(A \cap B) = 0.4 + 0.25 - 2 \times 0.05 = 0.55$$

4.4. Ejercicio 4:

En una población de moscas, el 25 % presentan mutación en los ojos, el 50 % presentan mutación en las alas, y el 40 % de las que presentan mutación en los ojos presentan mutación en las alas.

a) ¿Cuál es la probabilidad de que una mosca elegida al azar presente al menos una de las mutaciones?
b) ¿Cuál es la probabilidad de que presente mutación en los ojos pero no en las alas?

Sean A, B los sucesos referidos a tener mutación en los ojos y en las alas, respectivamente, entonces $P(A) = 0.25\ y\ P(B) = 0.5$

Para empezar esto es un problema de probabilidad condicionada, donde $P(B/A) = 0.4$
Y según la definición de probabilidad condicionada:$P(B/A) = \frac{P(A \cap B)}{P(A)}$

Entonces $P(B/A) = 0.4 = \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{0.25} \longrightarrow P(A \cap B) = 0.4 * 0.25 = 0.1$

Esto nos lleva a que: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.25 + 0.5 - 0.1 = \underline{0.65}$

Para el apartado b), tomaremos $\bar{B}$ para negar la mutación en las alas, entonces
$P(\bar{B}) = 1 - P(B) = 1 - 0.5 = 0.5$ y $P(A/\bar{B}) = 1 - P(B/A) = 1 - 0.4 = 0.6$

Por el mismo razonamiento que antes:
$P(A/\bar{B}) = 0.4 = \frac{P(A \cap \bar{B})}{P(A)} = \frac{P(A \cap \bar{B})}{0.25} \longrightarrow P(A \cap \bar{B}) = 0.6 * 0.25 = \underline{0.15}$

4.5. Ejercicio 5:

Una empresa utiliza dos sistemas alternativos, A y B, en la fabricación de un artículo, fabricando por el sistema A el 20 % de su producción. Cuando a un cliente se le ofrece dicho artículo, la probabilidad de que lo compre es 2/3 si éste se fabricó por el sistema A y 2/5 si se fabricó por el sistema B. Calcular la probabilidad de vender el producto.

Si denoto por S_A al suceso de que el producto haya sido producido por el sistema A y S_B al suceso de de que haya sido producido por el sistema B tengo que:

$$P(S_A) = 0,2 \Rightarrow P(S_B) = 1 - P(S_A) = 1 - 0,2 = 0,8$$

Denotando ahora por C el suceso de que el cliente compre el producto tengo que:

$$P(C/S_A) = \frac{2}{3} \qquad P(C/S_B) = \frac{2}{5}$$

La probabilidad de vender el producto es la probabilidad del suceso C que, aplicando el Teorema de la Probabilidad total se calcula como:

$$P(C) = P(S_A)P(C/S_A) + P(S_B)P(C/S_B) = 0,2 \times \frac{2}{3} + 0,8 \times \frac{2}{5} = \frac{34}{75} = 0.453$$

Por tanto hay una probabilidad del 45.3 % de posibilidades de vender el producto.

4.6. Ejercicio 6:

Se consideran dos urnas: la primera con 20 bolas, de las cuales 18 son blancas, y la segunda con 10 bolas, de las cuales 9 son blancas. Se extrae una bola de la segunda urna y se deposita en la primera; si a continuación, se extrae una bola de esta, calcular la probabilidad de que sea blanca.

Solución:

Denotamos los sucesos:

B_2: ”sacar bola blanca de la segunda urna”
B_1: ”sacar bola blanca de la primera urna”
X_1: ”sacar bola de otro color de la primera urna”

La probabilidad la podemos calcular aplicando el Teorema de la probabilidad total:

$$P(B_1) = P(B_1|B_2)P(B_2) + P(B_1|X_1)P(X_1) = \frac{19}{21} \times \frac{9}{10} + \frac{18}{21} \times \frac{1}{10} = \frac{9}{10}$$

4.7. Ejercicio 7:

7. Se dispone de tres urnas con la siguiente composición de bolas blancas y negras:

U1: 5B y 5N U2: 6B y 4N U3: 7B y 3N.

Se elige una urna al azar y se sacan cuatro bolas sin reemplazamiento.

a) Calcular la probabilidad de que las cuatro sean blancas.
b) Si en las bolas extraídas solo hay una negra, ¿cual es la probabilidad de que la urna elegida haya sido U2?

Sea A el suceso 'sacar exactamente 4 bolas blancas', sea B el suceso 'sacar exactamente 1 bola negra' y sea U_i el suceso 'escoger la urna i' $i = 1, 2, 3$

Aplicando el Teorema de Probabilidad Total y combinatoria:

$$P(A) = P(U_1)P(A|U_1) + P(U_2)P(A|U_2) + P(U_3)P(A|U_3) = \frac{1}{3}\left(\frac{\binom{5}{4}\binom{5}{0}}{\binom{10}{4}} + \frac{\binom{6}{4}\binom{4}{0}}{\binom{10}{4}} + \frac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}}\right) =$$

$$= \frac{\binom{5}{4} + \binom{6}{4} + \binom{7}{4}}{3\binom{10}{4}} = 0.087302$$

Aplicando el Teorema de Bayes y combinatoria:

$$P(U_2|B) = \frac{P(U_2 \cap B)}{P(B)} = \frac{P(U_2)P(N|U_2)}{P(U_1)P(N|U_1) + P(U_2)P(N|U_2) + P(U_3)P(N|U_3)} =$$

$$= \frac{\frac{1}{3}\frac{\binom{4}{1}\binom{6}{3}}{\binom{10}{4}}}{\frac{1}{3}\left(\frac{\binom{5}{1}\binom{5}{3}}{\binom{10}{4}} + \frac{\binom{4}{1}\binom{6}{3}}{\binom{10}{4}} + \frac{\binom{3}{1}\binom{7}{3}}{\binom{10}{4}}\right)} = \frac{4\binom{6}{3}}{5\binom{5}{3} + 4\binom{6}{3} + 3\binom{7}{3}} = 0.340426$$

4.8. Ejercicio 8:

La probabilidad de que se olvide inyectar el suero a un enfermo durante la ausencia del doctor es 2/3. Si se le inyecta el suero, el enfermo tiene igual probabilidad de mejorar que de empeorar, pero si no se le inyecta, la probabilidad de mejorar se reduce a 0.25. Al regreso, el doctor encuentra que el enfermo ha empeorado. ¿Cuál es la probabilidad de que no se le haya inyectado el suero?

Si denoto por I al suceso de inyectar el suero y M al suceso de mejorar tengo que:

$$P(\bar{I}) = \frac{2}{3} \ \Rightarrow \ P(I) = 1 - P(\bar{I}) = 1 - \frac{2}{3} = \frac{1}{3}$$

$$P(M/I) = 0,5 \qquad\qquad P(M/\bar{I}) = 0,25$$

Me piden calcular $P(\bar{I}/\bar{M})$. Aplicando propiedades y definiciones,

$$P(\bar{I}/\bar{M}) = \frac{P(\bar{I} \cap \bar{M})}{P(\bar{M})} = \frac{P(\overline{I \cup M})}{P(\bar{M})} = \frac{1 - P(I \cup M)}{1 - P(M)} = \frac{1 - (P(I) + P(M) - P(I \cap M))}{1 - P(M)} \ (*)$$

$$P(M/I) = \frac{P(M \cap I)}{P(I)} \ \Rightarrow \ P(M \cap I) = P(I)P(M/I) = \frac{1}{3} \times 0,5 = \frac{1}{6}$$

$$P(M) = P(I)P(M/I) + P(\bar{I})P(M/\bar{I}) = \frac{1}{6} + \frac{2}{3} \times 0,25 = \frac{1}{3}$$

Sustituyendo los valores obtenidos en la expresión (*) tengo:

$$P(\bar{I}/\bar{M}) = \frac{1 - (\frac{1}{3} + \frac{1}{3} - \frac{1}{6})}{1 - \frac{1}{3}} = \frac{3}{4} = 0,75$$

4.9. Ejercicio 9:

N urnas contienen cada una 4 bolas blancas y 6 negras, mientras otra urna contiene 5 blancas y 5 negras. De las N + 1 urnas se elige una al azar y se extraen dos bolas sucesivamente, sin reemplazamiento, resultando ser ambas negras. Sabiendo que la probabilidad de que queden 5 blancas y 3 negras en la urna elegida es 1/7, encontrar N.

Solución:

Sean los siguientes sucesos:

A: escoger la urna mayoritaria
B: escoger la urna especial
N: sacar dos bolas negras

Entonces la probabilidad de que queden 5 bolas blancas y 3 negras será:

$$P(B|N) = \frac{P(B \cap N)}{P(N)}$$

Se proceden a calcular las probabilidades necesarias:

$$P(A \cap N) = P(A)P(N) = \frac{N}{N+1} \times \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{N}{N+1} \times \frac{6 \times 5}{10 \times 9} = \frac{N}{3(N+1)}$$

$$P(B \cap N) = P(B)P(N) = \frac{1}{N+1} \times \frac{\binom{5}{2}}{\binom{10}{2}} = \frac{1}{N+1} \times \frac{5 \times 4}{10 \times 9} = \frac{2}{9(N+1)}$$

$$P(N) = P(A \cap N) + P(B \cap N) = \frac{N}{3(N+1)} + \frac{2}{9(N+1)} = \frac{3N+2}{9(N+1)}$$

Entonces quedaría:

$$P(B|N) = \frac{P(B \cap N)}{P(N)} = \frac{\frac{2}{9(N+1)}}{\frac{3N+2}{9(N+1)}} = \frac{2}{3N+2} = \frac{1}{7} \Rightarrow 3N + 2 = 14 \Rightarrow \underline{N = 4}$$

4.10. Ejercicio 10:

Se dispone de 6 cajas, cada una con 12 tornillos; una caja tiene 8 buenos y 4 defectuosos; dos cajas tienen 6 buenos y 6 defectuosos y tres cajas tienen 4 buenos y 8 defectuosos. Se elige al azar una caja y se extraen 3 tornillos con reemplazamiento, de los cuales 2 son buenos y 1 es defectuoso. ¿Cual es la probabilidad de que la caja elegida contuviera 6 buenos y 6 defectuosos?
Sean C_1, C_2, C_3 los sucesos referidos a elegir cada una de las tres opciones de cajas.

- (caso C_1) = 8 buenos y 4 defectuosos, $P(C_1) = 1/6$
- (caso C_2) = 6 buenos y 6 defectuosos, $P(C_2) = 2/6 = 1/3$
- (caso C_3) = 4 buenos y 8 defectuosos, $P(C_3) = 3/6 = 1/2$

se extraen 3 tornillos, 2 buenos y 1 defectuoso, sea este suceso B_2D_1, buscamos
$P(C_2/B_2D_1) = \frac{P(C_2 \cap B_2D_1)}{P(B_2D_1)}$

- $P(C_2 \cap B_2D_1)$ Esto es una permutación con repetición (puesto que los tornillos se reemplazan) cuya probabilidad es: $P(C_2 \cap B_2D_1) = P(C_2) * P(B_2D_1) = \frac{1}{3} * (\frac{1}{2})^3 = \frac{1}{24}$
- $P(B_2D_1) = P(C_1) * P(B_2D_1/C_1) + P(C_2) * P(B_2D_1/C_2) + P(C_3) * P(B_2D_1/C_3) =$
$\frac{1}{6} * (\frac{8*8*4}{12*12*12}) + \frac{1}{3} * (\frac{2}{2})^3 + \frac{1}{2} * (\frac{4*4*8}{12*12*12}) = 0.10339506172$
- $P(C_2/B_2D_1) = \frac{P(C_2 \cap B_2D_1)}{P(B_2D_1)} = \frac{\frac{1}{24}}{0.10339506172} = \frac{1}{2.48148148148} = \underline{0.40298507462}$

4.11. Ejercicio 11:

Se seleccionan n dados con probabilidad $p_n = 1/2^n, n \in \mathbb{N}$. Si se lanzan estos n dados y se obtiene una suma de 4 puntos, ¿cuál es la probabilidad de haber seleccionado 4 dados?

Denotando por E_n a la probabilidad de elegir n dados tal que $P(E_n) = p_n = \frac{1}{2^n}$ y denotando por V_i la probabilidad de que salga el valor i en el dado sabiendo que $P(V_i) = \frac{1}{6}$ $(\forall i \in \mathbb{N} \cap [1,6])$.
Como máximo puede haber 4 dados.

$$P(E_1) = \frac{1}{2} \qquad P(E_2) = \frac{1}{4} \qquad P(E_3) = \frac{1}{8} \qquad P(E_4) = \frac{1}{16}$$

Si S es el suceso de sumar 4 puntos se me pide calcular $P(E_4/S)$.
Teniendo en cuenta

a) Con un dado la posibilidad de sumar 4 es única que es un 4 $\Rightarrow P(S/E_1) = P(V_4) = \frac{1}{6}$

b) Con dos dados los posibles casos de sumar 4 son 2:

- 1 dado de 3 + 1 dado de 1
- 1 dado de 1 + 1 dado de 3
- 1 dado de 2 + 1 dado de 2

$$P(S/E_2) = P(V_1)P(V_3/V_1) + P(V_3)P(V_1/V_3) + P(V_2)P(V_2|V_2)$$

Como los sucesos V_i son independientes:

$$P(V_i/V_j) = \frac{P(V_i \cap V_j)}{P(V_j)} = \frac{P(V_i)P(V_j)}{P(V_j)} = P(V_i)$$

$$P(S/E_2) = P(V_1)P(V_3)+P(V_3)P(V_1)+P(V_2)P(V_2) = 2P(V_1)P(V_3)+P(V_2)(V_2) = 3\times\frac{1}{6^2} = \frac{1}{12}$$

c) Con 3 dados los casos son 3:

- 1 dado de 1 + 1 dado de 1 + 1 dado de 2
- 1 dado de 1 + 1 dado de 2 + 1 dado de 1
- 1 dado de 2 + 1 dado de 1 + 1 dado de 1

Con el razonamiento seguido anteriormente:

$$P(S/E_3) = P(V_1)P(V_1/V_1)P(V_2/V_1 \cap V_1) + P(V_1)P(V_2/V_1)P(V_1/V_2 \cap V_1)+$$
$$+P(V_2)P(V_1/V_2)P(V_1/V_1 \cap V_2) = P(V_2)P(V_1)P(V_1)+P(V_1)P(V_2)P(V_1)+P(V_2)P(V_1)P(V_1) =$$
$$= 3P(V_1)P(V_1)P(V_2) = 3 \times \frac{1}{6^3} = \frac{1}{72}$$

d) Con 4 dados que suman 4 solo cabe la posibilidad de 4 dados de 1

$$P(S/E_4) = P(V_1)P(V_1)P(V_1)P(V_1) = \frac{1}{6^4} = \frac{1}{1296}$$

e) Con más de 4 dados es imposible que sumen 4

Con los datos calculados puedo aplicar el Teorema de Bayes:

$$P(E_4/S) = \frac{P(E_4 \cap S)}{P(S)} = \frac{P(E_4)P(S/E_4)}{P(S)}$$

$$P(S) = P(E_1)P(S/E_1) + P(E_2)P(S/E_2) + P(E_3)P(S/E_3) + P(E_4)P(S/E_4) =$$
$$= \frac{1}{2} \times \frac{1}{6} + \frac{1}{4} \times \frac{1}{12} + \frac{1}{8} \times \frac{1}{72} + \frac{1}{16} \times \frac{1}{1296} = 0.105951$$

$$P(E_4/S) = \frac{\frac{1}{16} \times \frac{1}{1296}}{0.105951} = 0.000455$$

4.12. Ejercicio 12:

Se lanza una moneda; si sale cara, se introducen k bolas blancas en una urna y si sale cruz, se introducen 2k bolas blancas. Se hace una segunda tirada, poniendo en la urna h bolas negras si sale cara y 2h si sale cruz. De la urna así compuesta se toma una bola al azar. ¿Cuál es la probabilidad de que sea negra?

Solución:

La urna puede tener 4 composiciones diferentes, cada una con la misma probabilidad 0.25:

- k bolas blancas y h bolas negras
$$P(N_1) = \frac{h}{k+h}$$
- 2k bolas blancas y h bolas negras
$$P(N_2) = \frac{h}{2k+h}$$
- k bolas blancas y 2h bolas negras
$$P(N_3) = \frac{2h}{k+2h}$$
- 2k bolas blancas y 2h bolas negras
$$P(N_4) = \frac{2h}{2k+2h} = \frac{h}{k+h} = P(N_1)$$

Entonces, por el Teorema de Probabilidad Total, la probabilidad quedaría como:

$$P(N) = \frac{1}{4}\sum_{i=1}^{4} P(N_i) = \frac{h}{4}\left(\frac{2}{k+h} + \frac{1}{2k+h} + \frac{2}{k+2h}\right)$$

Nótese que si $k = h$, queda $P(N) = \frac{k}{4}\left(\frac{1}{k} + \frac{1}{3k} + \frac{2}{3k}\right) = \frac{1}{2}\ \forall k \in \mathbb{N}$.

5. Relación 5

5.1. Ejercicio 1:

1. Sea X una variable aleatoria con función masa de probabilidad P (X = i) = k_i; i = 1, ..., 20.
a) Determinar el valor de k, la función de distribución y las siguientes probabilidades:
$P(X = 4), P(X < 4), P(3 \leq X \leq 10), P(3 < X \leq 10), P(3 < X < 10)$.
b) Supongamos que un jugador gana 20 monedas si al observar esta variable obtiene un valor menor que 4, gana 24 monedas si obtiene el valor 4 y, en caso contrario, pierde una moneda. Calcular la ganancia esperada del jugador y decir si el juego le es favorable.
$\sum_{i=1}^{20} P(X = i) = \sum_{i=1}^{20} ki = 1 \xrightarrow{\sum_{i=1}^{20} i=210} 210k = 1 \rightarrow k = \frac{1}{210}$

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ \sum_{i=0}^{E(x)} P(X = j) = \sum_{i=1}^{E(x)} \frac{1}{210} * i & \text{si } 1 \leq x < 20 \\ 1 & \text{si } x \geq 20 \end{cases}$$

Siendo E(x) la parte entera de x

a)

- $P(X = 4) = \frac{4}{210} = \frac{2}{105} = 0.019048$
- $P(X < 4) = P(X \leq 3) = \sum_{i=1}^{3} \frac{1}{210} * i = \frac{1+2+3}{210} = \frac{6}{210} = \frac{1}{35}$
- $P(3 \leq X \leq 10) = \sum_{i=3}^{10} \frac{1}{210} * i = \frac{3+4+5+6+7+8+9+10}{210} = \frac{26}{105} = 0.24761904761$
- $P(3 < X \leq 10) = P(4 \leq X \leq 10) = \sum_{i=4}^{10} \frac{1}{210} * i = \frac{4+5+6+7+8+9+10}{210} = \frac{7}{30} = 0.23333333333$
- $P(3 \leq X < 10) = P(4 \leq X \leq 9) = \sum_{i=4}^{9} \frac{1}{210} * i = \frac{4+5+6+7+8+9}{210} = \frac{39}{210} = 0.18571428571$

b)
Tomemos una nueva variable aleatoria Y, que será las monedas ganadas en función de X

$$y = g(x) = \begin{cases} \text{-1} & \text{si } 4 < x \leq 20 \\ 24 & \text{si } x = 4 \\ 20 & \text{si } 1 < x < 4 \end{cases}$$ y calculamos los valores de que Y en tome todos sus posibles valores en funcion de X

- $P(Y = -1) = P(X > 4) = 1 - P(X \leq 4) = 1 - P(X < 4) - P(X = 4) =$
  $1 - 0.02857142857 - 0.01904761904 = 0.95238095239$
- $P(Y = 24) = P(X = 4) = \frac{4}{210} = \frac{2}{105} = 0,01904761904$
- $P(Y = 20) = P(X < 4) = P(X \leq 3) = F_X(3) = \frac{1+2+3}{210} = \frac{1}{35} = 0.02857142857$

calculemos entonces la esperanza matemática:$(y_1 = -1, y_2 = 24, y_3 = 20)$

$$E[Y] = \sum_{i=1}^{n=3} y_i P(Y = y_i) = -1 * 0.95238095239 + \frac{24 * 2}{105} + \frac{20 * 1}{35} = 0.07619047618 \tag{1}$$

y como tenemos una esperanza de Y > 0, el juego es favorable al jugador.

5.2. Ejercicio 2:

Sea X el número de bolas blancas obtenidas al sacar dos de una urna con 10 bolas de las que 8 son blancas. Calcular:

a) Función masa de probabilidad y función de distribución.

La probabilidad de sacar x bolas blancas se calcula mediante la siguiente expresión:

$$P[X = x] = \frac{\binom{8}{x}\binom{2}{2-x}}{\binom{10}{2}} \qquad x = 0, 1, 2$$

Que es la función masa de probabilidad (en el tema 6 se ve que es una distribución hipergeométrica de parámetros $N = 10,\ N_1 = 8,\ n = 2$).

De la función anterior obtenemos:

$$P[X = 0] = \frac{\binom{8}{0}\binom{2}{2}}{\binom{10}{2}} = \frac{1}{45}$$

$$P[X = 1] = \frac{\binom{8}{1}\binom{2}{1}}{\binom{10}{2}} = \frac{16}{45}$$

$$P[X = 2] = \frac{\binom{8}{2}\binom{2}{0}}{\binom{10}{2}} = \frac{28}{45}$$

Y entonces la función de distribución $F_X(x) = P[X \leq x]$ queda:

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{45} & \text{si } 0 \leq x < 1 \\ \frac{17}{45} & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

b) Media, mediana y moda, dando la interpretación de cada una de estas medidas.

$Mo = 2 \Rightarrow$ La mayoría de veces se sacarán 2 bolas blancas
$Me = 2 \Rightarrow$ La probabilidad de sacar 2 o menos como de sacar 2 o más (2) es mayor o igual a 1/2

$$E[X] = \sum_{i=0}^{2} x_i P[X = x_i] = 0 \times \frac{1}{45} + 1 \times \frac{16}{45} + 2 \times \frac{28}{45} = 1.6$$

Al repetir el experimento varias veces, de media se sacarán aproximadamente 1.6 bolas blancas, por ejemplo si lo repetimos 10 veces lo esperado será sacar 16 bolas blancas en total.

c) Intervalo intercuartílico, especificando su interpretación.

$Q_1 = 1 \qquad Q_3 = 2 \Rightarrow R_I = 1$

El recorrido intercuartílico es 1 en comparación al recorrido que es 2. Esto parece indicar que la distribución es homogénea pero no es cierto, pues los cuartiles Q_1 y Q_3 están en la parte derecha de la distribución, lo que indica que la distribución está desplazada a la derecha.

5.3. Ejercicio 3:

El número de lanzamientos de una moneda hasta salir cara es una variable aleatoria con distribución $P(X = x) = 2^{-x};\ x = 1, 2, ...$

a) Probar que la función masa de probabilidad está bien definida.

b) Calcular la probabilidad de que el número de lanzamientos necesarios para salir cara esté entre 4 y 10.

c) Calcular los cuartiles y la moda de la distribución, interpretando los valores.

d) Calcular la función generatriz de momentos y, a partir de ella, el número medio de lanzamientos necesarios para salir cara y la desviación típica.

Solución:

a) Para que la función masa de probabilidad $P(X = x) = 2^{-x}$ esté bien definida basta con que se cumplan estas dos condiciones:

   - $0 \leq P(X = x) \leq 1\ \forall x \in \mathbb{N}$
   - $\sum_{x\in\mathbb{N}} P(X = x) = 1$

   Para la primera, como 2^{-x} es una función decreciente y positiva y $2^{-1} = \frac{1}{2} < 1$ podemos afirmar que $0 < 2^{-x} \leq \frac{1}{2}\ \forall x \in \mathbb{N}$

   Para la segunda tenemos que ver la siguiente serie converge a 1:

$$\sum_{x\in\mathbb{N}} 2^{-x} = \sum_{x\geq 0} 2^{-x} - 2^0 = \frac{1}{1-\frac{1}{2}} - 1 = 2 - 1 = 1$$

   En efecto, por ser una serie geométrica, podemos comprobar que converge a 1.

b) La probabilidad de que se necesiten x lanzamientos para que salga cara es la que nos da la función masa de probabilidad $P(X = x) = 2^{-x}$. Por lo que la probabilidad que nos piden es:

$$P(4 \leq X \leq 10) = \sum_{n=4}^{10} P(X = n) = \sum_{n=4}^{10} 2^{-x} = \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \frac{1}{2^{10}} = 0.124$$

c) Los 3 cuartiles $Q_1,\ Q_2,\ Q_3$ dividen la función masa de probabilidad en 4 partes, de manera que:

$$P(X \leq Q_1) \geq \frac{1}{4} \Rightarrow Q_1 = 1$$
$$P(X \leq Q_2) \geq \frac{1}{2} \Rightarrow Q_2 = 1$$
$$P(X \leq Q_3) \geq \frac{3}{4} \Rightarrow Q_3 = 2$$

   Que los cuartiles estén tan a la izquierda de la distribución significa que la distribución es muy asimétrica y que la probabilidad de que salga cara en alguna de las primeras tiradas es alta en comparación con que tengamos que tirar la moneda muchas veces hasta que salga cara.
   La moda en variable discreta se define como:

$$P(X = Mo) = max\{P(X = x) : x \in E\}$$

   Siendo E = $\mathbb{N}$ en este caso. Como la función masa de probabilidad es decreciente, $Mo = 1$.
   Esto significa que si hacemos un número muy elevado de experimentos, en la mayoría de ellos solo habríamos tirado la moneda una vez.

d) La función generatriz de momentos se calcula como:

$$M_X(t) = E[e^{tX}] = \sum_{x\in\mathbb{N}} e^{tx} P(X = x) = \sum_{x\in\mathbb{N}} \frac{e^{tx}}{2^x} = \sum_{x\in\mathbb{N}} \left(\frac{e^t}{2}\right)^x = \frac{1}{1-\frac{e^t}{2}} - \left(\frac{e^t}{2}\right)^0 = \frac{1}{1-\frac{e^t}{2}} - 1 =$$

$$= \frac{\frac{e^t}{2}}{1-\frac{e^t}{2}} = \frac{e^t}{2-e^t} \ \forall t < ln2$$

Derivando y evaluando en el 0, podemos obtener el momento de orden 1, que es la esperanza:

$$M_X'(t) = \frac{e^t(2-e^t)+e^{2t}}{(2-e^t)^2} = \frac{2e^t}{(2-e^t)^2}$$

$$\underline{E[X]} = M_X'(0) = \frac{2}{1} = \underline{2}$$

Derivando otra vez y evaluando en el 0, podemos obtener el momento de orden 2, con el que podemos obtener la varianza y la desviación típica:

$$M_X''(t) = \frac{2e^t(2-e^t)^2 + 4e^{2t}(2-e^t)}{(2-e^t)^4} = \frac{2e^t(2-e^t)+4e^{2t}}{(2-e^t)^3} = \frac{4e^t+2e^{2t}}{(2-e^t)^3} = \frac{2e^t(2+e^t)}{(2-e^t)^3}$$

$$E[X^2] = M_X''(0) = \frac{2\times 3}{1} = 6$$

$$\sigma_X^2 = E[X^2] - E[X]^2 = 6 - 2^2 = 2 \Rightarrow \underline{\sigma_X = \sqrt{2}}$$

5.4. Ejercicio 4:

Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} k_1(x+1) & \text{si } 0 \leq x \leq 4 \\ & & \forall x \in [0,6] \\ k_2 x^2 & \text{si } 4 < x \leq 6 \end{cases}$$

Sabiendo que $P(0 \leq X \leq 4) = 2/3$, determinar k_1, k_2, y deducir su función de distribución.

Como f(x) es una función de densidad continua:

$$P(0 \leq X \leq 4) = 2/3 = \int_0^4 f(x)dx = \int_0^4 k_1(x+1)dx = k_1[\frac{x^2}{2} + x]_0^4 = k_1 * 12 \Rightarrow k_1 = \frac{2/3}{12} = \frac{1}{18}$$

Además, se debe cumplir que $\sum_{x \in \mathbb{N}} P(X = x) = 1 \Rightarrow P(4 < X \leq 6) = 1/3$

$$P(4 < X \leq 6) = 1/3 = \int_4^6 f(x)dx = \int_0^4 k_2(x^2)dx = k_2[\frac{x^3}{3}]_4^6 = k_2 * \frac{152}{3} \Rightarrow k_2 = \frac{1/3}{152/3} = \frac{1}{152}$$

La función de distribución $F_X(x)$ se calcula como:

$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f(t)dt$$

En nuestro caso será:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{18}\int_0^x (t+1)dt & \text{si } 0 \leq x < 4 \\ \frac{2}{3} + \frac{1}{152}\int_4^x t^2 dt & \text{si } 4 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$$

Y resolviendo las integrales queda como:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x(x+2)}{36} & \text{si } 0 \leq x < 4 \\ \frac{x^3+240}{456} & \text{si } 4 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$$

5.5. Ejercicio 5:

La dimensión en centímetros de los tornillos que salen de cierta fábrica es una variable aleatoria, X, con función de densidad

$$f(x) = \frac{k}{x^2}, \quad 1 \leq x \leq 10$$

a) Determinar el valor de k, y obtener la función de distribución.

   Para ser una función de densidad debe verificar $\int_{-\infty}^{+\infty} \frac{k}{x^2} dx = 1$

$$\int_{-\infty}^{+\infty} \frac{k}{x^2} dx = \cancel{\int_{-\infty}^{1} \frac{k}{x^2}} + \int_{1}^{10} \frac{k}{x^2} + \cancel{\int_{10}^{+\infty} \frac{k}{x^2}} = -k \int_{1}^{10} \frac{-1}{x^2} dx = -k(\frac{1}{10} - \frac{1}{1}) = \frac{9k}{10}$$

   Por tanto

$$\frac{9k}{10} = 1 \Leftrightarrow k = \frac{10}{9}$$

   La función de distribución será:

$$F_X(x) = \begin{cases} 0 & si \quad x \leq 1 \\ \int_1^x \frac{\frac{10}{9}}{x^2} dx = \frac{-10}{9}(\frac{1}{x} - 1) & si \quad 1 \leq x < 10 \\ 1 & si \quad x \geq 10 \end{cases}$$

b) Hallar la probabilidad de que la dimensión de un tornillo esté entre 2 y 5 cm.

   Para hallar dicha probabilidad aplicaré que $F_X(B) = \int_B f(x)dx$ siendo $f(x)$ la función de densidad y B un intervalo contenido en Ω.

$$F_X(B) = \int_B f(x)dx = \int_2^5 \frac{\frac{10}{9}}{x^2} dx = \frac{-10}{9}(\frac{1}{5} - \frac{1}{2}) = \frac{1}{3}$$

   Por tanto la probabilidad será de $\frac{1}{3}$.

c) Determinar la dimensión máxima del 50 % de los tornillos con menor dimensión y la dimensión mínima del 5 % con mayor dimensión.

   La dimensión máxima del 50 % de los tornillos con menor dimensión va a coincidir con el valor de la mediana, es decir aquel $x_0 \in \Sigma / F_X(X) = 0,5$. Los cálculos quedan:

$$\frac{-10}{9}(\frac{1}{x} - 1) = \frac{1}{2} \;\Rightarrow\; \frac{1}{x} = \frac{-9}{20} + 1 \;\Rightarrow\; x = \frac{1}{\frac{-9}{20} + 1} = \frac{20}{11} = 1,81818$$

   Y por tanto la dimensión máxima será de 1,81818 cm (aproximadamente).

   La dimensión mínima del 5 % con mayor dimensión hace referencia al percentil 95 y se calculará de igual manera que antes pero cambiando 0,5 por 0,95:

$$\frac{-10}{9}(\frac{1}{x} - 1) = \frac{95}{100} \;\Rightarrow\; \frac{1}{x} = \frac{-855}{1000} + 1 \;\Rightarrow\; x = \frac{1}{\frac{-855}{100} + 1} = \frac{200}{29} = 6,89655$$

   Es decir, la dimensión mínima será de 6,89655 cm (aproximadamente).

d) Si Y denota la dimensión de los tornillos producidos en otra fábrica, con la misma media y desviación típica que X, dar un intervalo en el que tome valores la variable Y con una probabilidad mínima de 0.99.

   (Hace falta la desigualdad de Chebyshev: no la hemos dado)

5.6. Ejercicio 6:

Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{2x-1}{10} & \text{si } 1 < x \leq 2 \\ 0.4 & \text{si } 4 < x \leq 6 \end{cases}$$

a) Calcular $P(1.5 < X \leq 2),\ P(2.5 < X \leq 3.5),\ P(4.5 \leq X < 5.5),\ P(1.2 < X \leq 5.2)$.

b) Dar la expresión general de los momentos no centrados y deducir el valor medio de X.

c) Calcular la función generatriz de momentos de X.

Solución:

a) Para calcular las respectivas probabilidades, se integra en los intervalos correspondientes de la función, teniendo en cuenta que la función vale 0 en los intervalos en los que no está definida explícitamente.

$$P(1.5 < X \leq 2) = \int_{1.5}^{2} f(x)dx = \int_{1.5}^{2} \frac{2x-1}{10}dx = \frac{x^2-x}{10}]_{1.5}^{2} = 0.125$$

$$P(2.5 < X \leq 3.5) = \int_{2.5}^{3.5} 0dx = 0$$

$$P(4.5 \leq X < 5.5) = \int_{4.5}^{5.5} f(x)dx = \int_{4.5}^{5.5} 0.4dx = 0.4(5.5-4.5) = 0.4$$

$$P(1.2 < X \leq 5.2) = \int_{1.2}^{2} f(x)dx + \int_{4}^{5.2} f(x)dx = \int_{1.2}^{2} \frac{2x-1}{10}dx + \int_{4}^{5.2} 0.4dx =$$

$$= \frac{x^2-x}{10}]_{1.2}^{2} + 0.4(5.2-4) = 0.176 + 0.48 = 0.656$$

b) La expresión general de los momentos no centrados m_k se calcula de la siguiente manera:

$$m_k = E[X^k] = \int_{-\infty}^{+\infty} x^k f(x)dx = \int_{1}^{2} x^k \frac{2x-1}{10}dx + \int_{4}^{6} 0.4x^k dx =$$

$$= \int_{1}^{2} \left(\frac{1}{5}x^{k+1} - \frac{1}{10}x^k\right) dx + \int_{4}^{6} 0.4x^k dx = \left(\frac{x^{k+2}}{5(k+2)} - \frac{x^{k+1}}{10(k+1)}\right)_1^2 + 0.4\left(\frac{x^{k+1}}{k+1}\right)_4^6$$

Desarrollando:

$$m_k = \frac{(k+1)(2^{k+3}-2) + (k+2)(2^{k+3}3^{k+1} - 2^{2k+4} - 2^{k+1} + 1)}{10(k+1)(k+2)}$$

Sustituyendo $k = 1$ se obtiene la esperanza de x:

$$E[X] = m_1 = \frac{2 \times (16-2) + 3 \times (16 \times 9 - 64 - 4 + 1)}{10 \times 2 \times 3} = \frac{259}{60} = 4.31667$$

c) La función generatriz de momentos la da la siguiente expresión:

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f(x)dx = \int_{1}^{2} e^{tx}\frac{2x-1}{10}dx + \int_{4}^{6} 0.4e^{tx}dx$$

Separando integrales queda:

$$M_X(t) = \frac{1}{5}\int_{1}^{2} xe^{tx}dx - \frac{1}{10}\int_{1}^{2} e^{tx}dx + 0.4\int_{4}^{6} e^{tx}dx$$

Donde las dos últimas son inmediatas, y la primera se resuelve por partes:

$$M_X(t) = \left(\frac{e^{tx}(tx-1)}{5t^2} - \frac{e^{tx}}{10t}\right)_1^2 + \left(\frac{0.4e^{tx}}{t}\right)_4^6$$

Y por último desarrollando:

$$M_X(t) = \frac{4te^{2t}(e^{4t} - e^{2t} + 1) - e^{2t}(t+2) - e^t(3t-2)}{10t^2}$$

5.7. Ejercicio 7:

Con objeto de establecer un plan de producción, una empresa ha estimado que la demanda de sus clientes, en miles de unidades del producto, se comporta semanalmente con arreglo a una ley de probabilidad dada por la función de densidad:

$$f(x) = \frac{3}{4}(2x - x^2), 0 \leq x \leq 2$$

a) ¿Qué cantidad deberá tener dispuesta a la venta al comienzo de cada semana para poder satisfacer plenamente la demanda con probabilidad 0.5?
b) Pasado cierto tiempo, se observa que la demanda ha crecido, estimándose que en ese momento se distribuye según la función de densidad:

$$f(y) = \frac{3}{4}(4y - y^2 - 3), 1 \leq y \leq 3$$

Los empresarios sospechan que este crecimiento no ha afectado a la dispersión de la demanda, ¿es cierta esta sospecha?
Damos por hecho que dicha función de probabilidad cumple las condiciones:

- $0 \leq P(X = x) \leq 1 \ \forall x \in \mathbb{N}$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

a) nos están pidiendo la Mediana de la distribución, por tanto buscamos

$$Me \in [0,2] tq \int_{-\infty}^{Me} f(x)dx = 0.5 = \int_{0}^{Me} \frac{3}{4}(2x - x^2)dx = \frac{3}{4}(x^2 - \frac{x^3}{3}]_0^{Me}) = \frac{3}{4}(Me^2 - \frac{Me^3}{3}) \Rightarrow Me = 1$$

De lo que podemos concluir que necesitaríamos 1000 unidades para satisfacer dicha demanda.

b)

Para comprobar la suposición, utilizamos la desviación típica en ambas distribuciones, ya que tienen las mismas unidades y se pueden comparar de esa manera.

$$E[X] = \int_0^2 \frac{3}{4}x(2x - x^2)dx = \left[\frac{x^3}{2} - \frac{3}{16}x^4\right]_0^2 = 1$$

$$E[X^2] = \int_0^2 \frac{3}{4}x^2(2x - x^2)dx = \left[\frac{3}{8}x^4 - \frac{3}{20}x^5\right]_0^2 = 1.2$$

$$E[Y] = \int_1^3 \frac{3}{4}y(4y - y^2 - 3)dy = \left[y^3 - \frac{3}{16}y^4 - \frac{9}{8}y^2\right]_1^3 = 2$$

$$E[Y^2] = \int_1^3 \frac{3}{4}y^2(4y - y^2 - 3)dy = \left[\frac{3}{4}y^4 - \frac{3}{20}y^5 - \frac{3}{4}y^3\right]_1^3 = 4.2$$

$$\sigma_X = \sqrt{Var(X)} = \sqrt{E[X^2] - E[X]^2} = \sqrt{0.2} = \frac{\sqrt{5}}{5}$$

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{E[Y^2] - E[Y]^2} = \sqrt{0.2} = \frac{\sqrt{5}}{5}$$

Como $\sigma_X = \sigma_Y$, la suposición era cierta, no ha variado la dispersión.

5.8. Ejercicio 8:

Calcular las funciones masa de probabilidad de las variables Y = X +2 y Z = X^2, siendo X una variable aleatoria con distribución:

$$P(X=-2)=\frac{1}{5} \quad P(X=-1)=\frac{1}{10} \quad P(X=0)=\frac{1}{5} \quad P(X=1)=\frac{2}{5} \quad P(X=2)=\frac{1}{10}$$

¿Cómo afecta el cambio de X a Y en el coeficiente de variación?

En este caso, si E_X es la función masa de probabilidad, tengo que $Img(E_X)=\{\frac{1}{5},\frac{1}{10},\frac{2}{5}\}$
Por el teorema General del Cambio de Variable

$$P_Y(B)=P_X(g^{-1}(B)) \text{ donde g(x) es una función medible}$$

Por tanto, si en este caso $g(X)=Y=X+2$, entonces $g^{-1}(X)=Y-2$

$$P(Y=y)=P(X=y-2)$$

$$-2=y-2 \Rightarrow y=0 \Rightarrow P(Y=0)=P(X=-2)=\frac{1}{5}$$
$$-1=y-2 \Rightarrow y=1 \Rightarrow P(Y=1)=P(X=-1)=\frac{1}{10}$$
$$0=y-2 \Rightarrow y=2 \Rightarrow P(Y=2)=P(X=0)=\frac{1}{5}$$
$$1=y-2 \Rightarrow y=3 \Rightarrow P(Y=3)=P(X=1)=\frac{2}{5}$$
$$2=y-2 \Rightarrow y=4 \Rightarrow P(Y=4)=P(X=2)=\frac{1}{10}$$

En el caso $h(x)=Z=X^2$ tendría $E_X=\{-2,-1,0,1,2\} \xrightarrow{h(x)=x^2} h(E_X)=\{0,1,4\}$

$$P(Z=z)=P(X^2=z)$$

$$z^2=0 \Rightarrow z=0 \Rightarrow P(Z=0)=P(X=0)=\frac{1}{5}$$
$$z^2=1 \Rightarrow z=\pm 1 \Rightarrow P(Z=1)=P(X=1)+P(x=-1)=\frac{2}{5}+\frac{1}{10}=\frac{1}{2}$$
$$z^2=4 \Rightarrow z=\pm 2 \Rightarrow P(Z=4)=P(X=2)+P(x=-2)=\frac{1}{10}+\frac{1}{5}=\frac{3}{10}$$

Para responder a la segunda cuestión deberé calcular los coeficientes de variación de Pearson para cada una de las variables:

$$CV_X=\frac{\sigma_X}{E[X]}=\frac{\sqrt{\sigma_X^2}}{E[X]}=\frac{\sqrt{E[(X-E[X])^2]}}{E[X]}$$

$$E[X]=(-2)\times\frac{1}{5}+(-1)\times\frac{1}{10}+0\times\frac{1}{5}+1\times\frac{2}{5}+2\times\frac{1}{10}=\frac{1}{10}$$

$$E[(X-E[X])^2]=E[X^2]-E[X]^2(-2)^2\times(\frac{1}{5})+(-1)^2\times(\frac{1}{10})+0^2\times(\frac{1}{5})+1^2\times(\frac{2}{5})+2^2\times(\frac{1}{10})-E[X]^2=$$
$$=\frac{17}{10}-(\frac{1}{10})^2=\frac{169}{100}=\frac{13^2}{10^2}$$

$$CV_X = \frac{\sqrt{\frac{13^2}{10^2}}}{\frac{1}{10}} = \frac{\frac{13}{\cancel{10}}}{\frac{1}{\cancel{10}}} = 13$$

$$CV_Y = \frac{\sigma_Y}{E[Y]} = \frac{\sqrt{\sigma_Y^2}}{E[Y]} = \frac{\sqrt{E[(Y-E[Y])^2]}}{E[Y]}$$

$$E[Y] = 0 \times \frac{1}{5} + 1 \times \frac{1}{10} + 2 \times \frac{1}{5} + 3 \times \frac{2}{5} + 4 \times \frac{1}{10} = \frac{21}{10}$$

$$E[(Y-E[Y])^2] = E[Y^2]-E[Y]^2 = 0^2\times\frac{1}{5}+1^2\times\frac{1}{10}+2^2\times\frac{1}{5}+3^2\times\frac{2}{5}+4^2\times\frac{1}{10}-E[Y]^2 = \frac{61}{10}-(\frac{21}{10})^2 = \frac{169}{100} = \frac{13^2}{10^2}$$

$$CV_Y = \frac{\sqrt{\frac{13^2}{10^2}}}{\frac{21}{10}} = \frac{\frac{13}{\cancel{10}}}{\frac{21}{\cancel{10}}} = \frac{13}{21}$$

$$\frac{CV_X}{CV_Y} = \frac{\cancel{13}}{\frac{\cancel{13}}{21}} = 21$$

Por tanto la desviación de la variable aleatoria Y es 21 veces menor que la desviación de X

Otra forma de resolver esta cuestión sería haciendo uso de las siguientes igualdades para transformaciones lineales:

$$E[aX+b] = aE[X]+b \qquad\qquad Var(aX+b) = a^2Var(X)$$

Entonces será:

$$E[Y] = E[X]+2$$

$$Var(Y) = Var(X) \Longleftrightarrow \sigma_Y = \sigma_X$$

Ahora calculamos la esperanza de X (de igual forma que antes) y de Y:

$$E[X] = (-2)\times\frac{1}{5} + (-1)\times\frac{1}{10} + 0\times\frac{1}{5} + 1\times\frac{2}{5} + 2\times\frac{1}{10} = \frac{1}{10}$$

$$E[Y] = E[X]+2 = \frac{1}{10}+2 = \frac{21}{10} \quad \text{(mismo resultado que el calculado anteriormente)}$$

Y por último comparamos los coeficientes de variación:

$$C.V.(X) = \frac{\sigma_X}{|E[X]|} = 10\sigma_X$$

$$C.V.(Y) = \frac{\sigma_Y}{|E[Y]|} = \frac{\sigma_X}{\frac{21}{10}} = \frac{10}{21}\sigma_X$$

$$\frac{C.V.(X)}{C.V.(Y)} = 21 \qquad (\sigma_X \neq 0)$$

Llegamos a la misma conclusión que antes confirmando su validez.

5.9. Ejercicio 9:

Calcular las funciones de densidad de las variables $Y = 2X + 3$ y $Z = |X|$, siendo X una variable continua con función de densidad:

$$f_X(x) = \frac{1}{4}, \qquad -2 < x < 2$$

Solución:

$Y = 2X + 3 = h(X)$ estrictamente monótona, medible y derivable en $]-2, 2[$

$h(]-2,2[) =]-1,7[$ $\quad h^{-1}(y) = \frac{y-3}{2}\ \forall y \in]-1,7[$ $\quad (h^{-1})'(y) = \frac{1}{2}\ \forall y \in]-1,7[$

Por el Teorema de Cambio de Variable de continua a continua:

$$f_Y(y) = f_X(h^{-1}(y)|(h^{-1})'(y)| = \frac{1}{4} \times |\frac{1}{2}| = \frac{1}{8}\ \forall y \in]-1,7[$$

Entonces la función de densidad queda:

$$f_Y(y) = \begin{cases} \frac{1}{8} & y \in]-1,7[\\ 0 & y \notin]-1,7[\end{cases}$$

$Z = |X| = h(X)$ medible, no derivable, no estrictamente monótona

$$h(x) = \begin{cases} x & \text{si } -2 < x < 0 \\ -x & \text{si } 0 \leq x < 2 \end{cases} \quad \forall z \in]-2,2[$$

$$h^{-1}(z) = \begin{cases} h_1^{-1}(z) = z \\ h_2^{-1}(z) = -z \end{cases} \quad \forall z \in [0,2[$$

$$(h^{-1})'(z) = \begin{cases} (h_1^{-1})'(z) = 1 \\ (h_2^{-1})'(z) = -1 \end{cases} \quad \forall z \in [0,2[$$

Se requiere aplicar la generalización del Teorema de Cambio de Variable continua a continua, ya que cada valor de $z \in [0,2[$ tiene 2 preimágenes:

$$f_Z(z) = f_X(h_1^{-1}(z))|(h_1^{-1})'(z)| + f_X(h_2^{-1}(z))|(h_2^{-1})'(z)| = \frac{1}{4} \times |1| + \frac{1}{4} \times |(-1)| = \frac{1}{2}\ \forall z \in [0,2[$$

Entonces la función de densidad queda:

$$f_Z(z) = \begin{cases} \frac{1}{2} & z \in [0,2[\\ 0 & z \notin [0,2[\end{cases}$$

5.10. Ejercicio 10:

Si X es una variable aleatoria con función de densidad $f(x) = \frac{e^{-|x|}}{2} si - \infty < x < \infty$,
hallar su función de distribución y las probabilidades de cada uno de los siguientes sucesos:
$a)|X| \leq 2$.
b) $|X| \leq 2$ o $X \geq 0$.
c) $|X| \leq 2$ y $X \leq -1$.
d) $X^3 - X^2 - X - 2 \leq 0$.
e) X es irracional.
Primero separamos el valor absoluto:

$$f(x) = \begin{cases} \frac{e^x}{2} & x \in]-\infty, 0[\\ \\ \frac{e^{-x}}{2} & x \in [0, \infty[\end{cases}$$

Veamos que dicha función de densidad cumple las condiciones:

- $0 \leq P(X = x) \leq 1 \ \forall x \in \mathbb{N}$

$$f'(x) = \begin{cases} \frac{e^x}{2} & x \in]-\infty, 0[\\ \\ \frac{-e^{-x}}{2} & x \in [0, \infty[\end{cases}$$

  es fácil ver que $f'(x) > 0 \ \forall x \in]-\infty, 0[$ y $f'(x) < 0 \ \forall x \in [0, \infty[$, por tanto f crece desde -infinito hasta el 0 y después decrece hasta el infinito. Además, como f(x) $> 0 \ \forall x \in \mathbb{R}$ podemos ver que f(0) = 1/2 en ambas ecuaciones por lo que sí, está acotada entre 0 y 1/2.

- $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} \frac{e^x}{2}dx + \int_{0}^{+\infty} \frac{e^{-x}}{2}dx = \frac{e^x}{2}]_{-\infty}^{0} + \frac{-e^x}{2}]_{0}^{\infty} = \frac{1}{2} + \frac{1}{2} = 1 \qquad (2)$$

  Por tanto podemos concluir que cumple ambas condiciones, y es una función de densidad.

Hallemos ahora la función de distribución:

$$F(X) = \begin{cases} \frac{e^x}{2} & x \in]-\infty, 0[\\ \\ 1 - \frac{e^{-x}}{2} & 0 \in [x, \infty[\end{cases}$$

Hallemos las probabilidades de cada uno de los siguientes sucesos:

- a) $P(|X| \leq 2) = P(-2 \leq X \leq 2) = F(2) - F(-2) = 1 - \frac{e^{-2}}{2} - \frac{e^{-2}}{2} = 1 - e^{-2}$
- b) $P(|X| \leq 2$ o $X \geq 0) = P(X \geq -2) = 1 - P(X \leq 2) = 1 - \frac{e^{-2}}{2} = 1 - \frac{1}{2e^2} = \frac{2e^2-1}{2e^2}$
- c) $P(|X| \leq 2$ y $X \leq -1) = P(-2 \leq X \leq -1) = \int_{-2}^{-1} \frac{e^x}{2} = \frac{e^{-1}}{2} - \frac{e^{-2}}{2} = \frac{e-1}{2e^2}$
- d) $P(X^3 - X^2 - X - 2 \leq 0)$
  $X^3 - X^2 - X - 2 \leq 0 \Leftrightarrow X \leq 2$
  $P(X \leq 2) = 1 - \frac{e^{-2}}{2} = 1 - \frac{1}{2e^2} = \frac{2e^2-1}{2e^2}$
- e) $P(X \in (R - Q)\ ) = 1$ (por que sabemos que el conjunto de los números racionales es numerable, o sea, es insignificante en comparación al de los irracionales, entonces la integral se hace igualmente.

5.11. Ejercicio 11:

Sea X una variable aleatoria con función de densidad

$$f(x) = 1, \ \ 0 \leq x \leq 1$$

Encontrar la distribución de las variables:

a) $Y = \frac{X}{1+X}$

Es claro que Y es una variable aleatoria de tipo continuo ya que para cada valor de Y existe un único X. Por tanto su función de densidad se obtiene aplicando la fórmula de Cambio de Variable a cada antiimagen y sumando las funciones resultantes:

$$y = g(x) = \frac{x}{1+x} \ \Rightarrow \ x = \frac{y}{1-y} \ \Rightarrow \ g^{-1}(y) = \frac{y}{1-y}$$

$$(g)'(x) = \frac{(1+x)-x}{(1+x)^2} = \frac{1}{(1+x)^2} > 0 \ \Rightarrow \ g \text{ es estrictamente creciente}$$

$$g(0) = 0 \qquad g(1) = \frac{1}{2} \ \Rightarrow \ Im(g(x)) = [0, \frac{1}{2}]$$

$$(g^{-1})'(x) = \frac{(1-y)+y}{(1-y)^2} = \frac{1}{(1-y)^2} > 0 \qquad \forall y \in [0, \frac{1}{2}]$$

Por el teorema de cambio de Variable de Continua a Continua tengo:

$$f_Y(Y) = f_X(g^{-1}(y))|\frac{dg^{-1(y)}}{dy}| = f_X(\frac{x}{1-x})|\frac{1}{(1-y)^2}| = \frac{1}{(1-y)^2} \qquad \forall y \in [0, \frac{1}{2}]$$

Una vez tengo calculada la función de densidad puedo calcular la función de distribución

$$F_Y(y) = \begin{cases} 0 & si \quad y < 0 \\ \int_{-\infty}^{y} f_Y(t)\ dt & si \quad 0 \leq y < \frac{1}{2} \\ 1 & si \quad y >= \frac{1}{2} \end{cases}$$

$$\int_{-\infty}^{y} f_Y(t)\ dt = \cancel{\int_{-\infty}^{0}} 0\ dt + \int_{0}^{y} \frac{1}{(1-t)^2}\ dt = (\frac{1}{1-y} - \frac{1}{1-0}) = \frac{1-(1-y)}{1-y} = \frac{y}{1-y}$$

Y finalmente queda

$$F_Y(y) = \begin{cases} 0 & si \quad y < 0 \\ \frac{y}{1-y} & si \quad 0 \leq y < \frac{1}{2} \\ 1 & si \quad y >= \frac{1}{2} \end{cases}$$

b) $Z = \begin{cases} -1 & si \quad X < 3/4 \\ 0 & si \quad X = 3/4 \\ 1 & si \quad X > 3/4 \end{cases}$

La variable Z es claramente discreta y aplicando el Teorema de Cambio de Variable de continua a discreta tengo que

$$\forall z \in E_Z, \ P\{Z = z\} = \int_{g^{-1}(y)} f_X(x)dx$$

En este caso tengo que

$$P(Z = 0) = P(X = \frac{3}{4}) = \int_{\frac{3}{4}}^{\frac{3}{4}} 1\, dx = 0$$

$$P(Z=-1)=P(X<\frac{3}{4})=\int_{-\infty}^{\frac{3}{4}}1\,dx=\cancel{\int_{-\infty}^{0}1\,dx}+\int_{0}^{\frac{3}{4}}1\,dx=\frac{3}{4}$$

$$P(Z=1)=P(X>\frac{3}{4})=\int_{\frac{3}{4}}^{+\infty}1\,dx=\int_{\frac{3}{4}}^{1}1\,dx+\cancel{\int_{1}^{+\infty}1\,dx}=1-\frac{3}{4}=\frac{1}{4}$$

Por tanto me queda que la función masa de probabilidad es

$$P_Z(z)=\begin{cases} \frac{3}{4} & si \quad z=-1 \\ 0 & si \quad z=0 \\ \frac{1}{4} & si \quad z=1 \end{cases}$$

Lo que me lleva a la función de distribución:

$$F_Z(z)=\begin{cases} 0 & si \quad z<-1 \\ \frac{3}{4} & si \quad -1\leq x<1 \\ 1 & si \quad z\geq 1 \end{cases}$$

5.12. Ejercicio 12:

Sea X una variable aleatoria simétrica con respecto al punto 2, y con coeficiente de variación 1. ¿Qué puede decirse acerca de las siguientes probabilidades?:

- $P(-8 < X < 12)$
- $P(-6 < X < 10)$

Solución:

Sabiendo que es simétrica con respecto al punto 2, se puede concluir que su esperanza vale 2.
Como se define el coeficiente de variación de Pearson como el cociente entre la desviación típica y la esperanza, al valer este 1, podemos concluir que la desviación típica vale 2, y por tanto, la varianza será 4. Como la varianza existe y es finita, se puede aplicar la desigualdad de Chebyshev:

$$P(|X - E[X]| < k) \geq 1 - \frac{Var(X)}{k^2}$$

En nuestro caso:

$$P(|X - 2| < k) \geq 1 - \frac{4}{k^2} \iff P(2 - k < X < 2 + k) \geq 1 - \frac{4}{k^2}$$

Tomando $k = 10$ y $k = 8$ se acotan las probabilidades del enunciado:

$$P(-8 < X < 12) \geq 1 - \frac{4}{100} = 0.96$$

$$P(-6 < X < 10) \geq 1 - \frac{4}{64} = 0.9375$$

Se puede concluir que dada una distribución con las características del enunciado, las probabilidades centrales serán muy altas.