UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

MAESTRÍA EN CIENCIAS DE LA INFORMACIÓN Y LAS TELECOMUNICACIONES

Herramientas matemáticas para el manejo de la información

Taller #1

Salazar Ortiz, Jaiver
jesalazaro@correo.udistrital.edu.co
20221495012

Forero Castro, Diego
ddforefoc@correo.udistrital.edu.co
20221495005

FECHA: *16 de enero de 2024*
FECHA DE ENTREGA: *19 de marzo de 2022*

1. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:*"Dado el conjunto de datos de una variable X cuantitativa, hacer un estudio basado estadística exploratoria y describir el comportamiento de la variable X.:

X			
8,65	24,53	13,25	9,84
4,03	4,55	14,31	4,54
8,57	4,72	3,95	23,56
7,41	1,86	11,78	1,49
1,89	10,98	7,10	18,99

RESPUESTA: Para un primer estudio se importan los datos a R luego se procede a su debido ordenamiento y se procede a calcular diversos datos mediante ecuaciones en especial la regla de Sturges, estos se pueden observar en la tabla 1.1.

Mínimo	1.49
Primer Cuartil	4.41
Mediana	7.99
Media	9.30
Tercer Cuartil	12.14
Máximo	24.53
Longitud	4.33
Intervalo	5.33
Desviación Estándar	6.81
Coeficiente de Variación	73.27
Asimetría	0.88

Tabla 1.1: Datos calculados

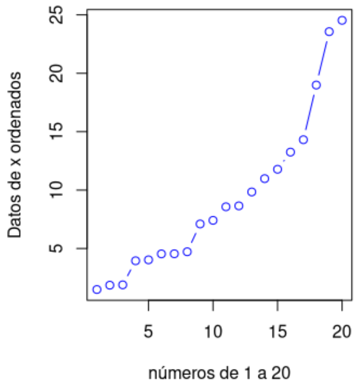
Debido a los datos de la tabla 1.1 es conveniente organizar una tabla con los valores de frecuencias basados en el intervalo(véase la tabla 1.2).

Marca de clase	Intervalo	Frec. absoluta	Frec. relativa	Frec. relativa acumulada
2,5	0-5	8	0.40	0.40
7,5	5-10	5	0.25	0.65
12,5	10-15	4	0.20	0.85
17,5	15-20	1	0.05	0.90
22,5	20-25	2	0.10	1
	Total	20	1	

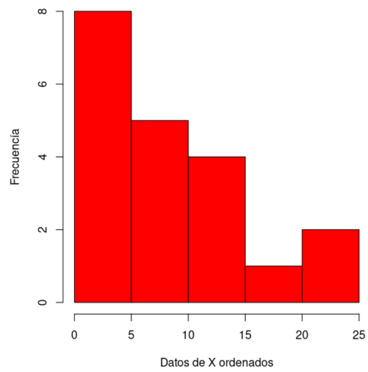
Tabla 1.2: frecuencia

Por consecuente es importante realizar gráficos para mostrar el comportamiento de los datos se realizaron los gráficos de comportamiento en orden(figura 1.1-a) , histograma(figura 1.2-b), diagrama de caja(figura 1.3-c), y curva de distribución(figura 1.1-d).

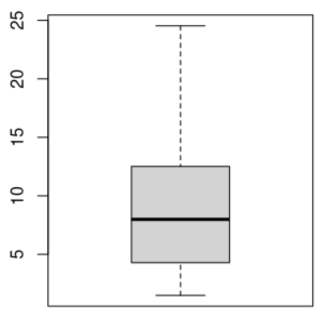
En la figura 1.1-a se observa el comportamiento de los datos en orden, este primer paso da una idea de en que zonas se encuentran la mayoría de los datos, se puede decir que gran parte de los datos se concentran de 0 a 15, y en menor medida de 15 a 25.



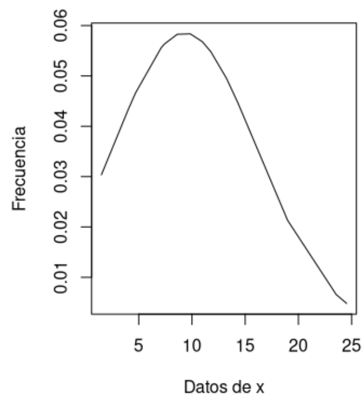
(a) Comportamiento de X



(b) Histograma



(c) Diagrama de caja



(d) Distribución normal

Figura 1.1: Gráficos realizados para los datos

El histograma de la figura 1.1-b muestra los datos de frecuencia e intervalo de la tabla 1.2, como se observa la mayoría de datos se encuentran en el intervalo de 0 a 5, el intervalo con menos datos es él de 15 a 20.

El diagrama de caja de la figura 1.1-c ilustra el comportamiento de los datos, como ya se mencionaba muchos datos se concentran en los primeros intervalos.

Debido a que se calculo el valor de asimetría es necesario el realizar un gráfico del mismo el cual se puede observar en la figura 1.1-d, esto indica que los datos tienden hacia la izquierda y concuerda con el valor de asimetría encontrado.

Con todo lo mencionado se puede concluir que el comportamiento de la variable x tiende a los primeros valores de 0 a 15.

2. PROBLEMA

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:*"Dado lo encontrado en el punto 1, proponga un modelo probabilístico para X y calcular lo siguiente:
>
> - Probabilidad que X esté entre 7 y 14.
> - Probabilidad que X sea menos de 10
> - Probabilidad que X es mayor que 17 y menor que 19
> - La probabilidad que esté entre 10 y 15, sujeto o condicionado a que X sea más de 7

RESPUESTA: Debido a la poca cantidad de datos y de información referente a ellos es conveniente usar el modelo frecuentista de probabilidad el cual establece que entre más se presente un evento es más probable, por tal razón es conveniente realizar un gráfico entre la frecuencia relativa tomada como probabilidad y los eventos véase figura 2.1.

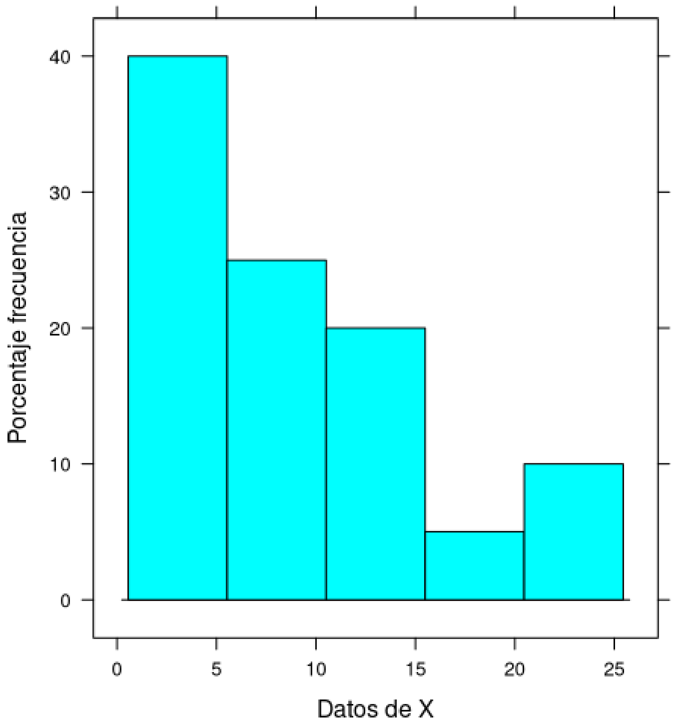


Figura 2.1: Frecuencia relativa en porcentaje

Probabilidad que X esté entre 7 y 14: Para esta probabilidad se cuentan las coincidencias que se encuentran entre los valores 7 y 14 luego se procede a dividir en el número total de valores, estas coincidencias son 8 por tanto se calcula la probabilidad de manera que:

$$P[7 < X < 14] = \frac{8}{20} = \frac{2}{5} = 0{,}4 \tag{2.1}$$

Probabilidad que X sea menos de 10: Dado que ya se tiene el porcentaje y el valor de frecuencia relativa simplemente sumamos los intervalos menores a 10, es decir:

$$P[X < 10] = P[5] + P[10] = 0{,}4 + 0{,}25 = 0{,}65 \tag{2.2}$$

Probabilidad que X es mayor que 17 y menos que 19: Se procede a contar las ocurrencias entre 17 y 19 para posteriormente dividirlas en el número total de datos, solo se encuentra un valor por tanto:

$$P[17 < X < 19] = \frac{1}{20} = 0{,}05 \tag{2.3}$$

Probabilidad que esté entre 10 y 15 condicionado a que X sea más que 7: Se parte de la probabilidad de 10 a 15 la cual es la frecuencia relativa con un valor de 0.2(20 %) luego se procede a calcular la probabilidad de que X sea mayor que 7, se cuentan el número de ocurrencias mayores a 7 y se dividen en el número total de datos, como se tienen 12 datos mayores a 7 se puede decir que:

$$P[X > 7] = \frac{12}{20} = \frac{3}{5} = 0{,}6 \tag{2.4}$$

Por tanto la probabilidad condicional:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}2 \times 0{,}6}{0{,}6} = 0{,}2 \tag{2.5}$$

3. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:*"De un estudio de ergonomía, se necesita caracterizar probabilísticamente la edad de un grupo de trabajadores y si T una variable aleatoria continua que representa la edad de ese grupo de trabajadores, entonces la función de densidad de probabilidad es:
>
> $$\begin{aligned} f(t) &= \frac{3}{20},\ si\ 20 < t < 25 \\ f(t) &= \frac{1}{100},\ si\ 25 < t < 50 \end{aligned} \tag{3.1}$$
>
> y es 0 en otro caso Calcular:
>
> - E[T]
> - Desv[T]
> - P[T>22]
> - P[52$\leq$ T $<$ 60]

Respuesta **Para** $E[T]$ **:**

$$\begin{aligned}
E[T] &= \int_{-\infty}^{\infty} tf(t)\,dt \\
&= \int_{20}^{25} t(\frac{3}{20})\,dt \ + \int_{25}^{50} t(\frac{1}{100})\,dt \\
&=(\frac{t^2}{2})(\frac{3}{20})\Big|_{20}^{25} + (\frac{t^2}{2})(\frac{1}{100})\Big|_{25}^{50} \\
&=(16,875) + (9,375) \\
&=26,25
\end{aligned} \tag{3.2}$$

Para $E[T^2]$ **:**

$$\begin{aligned}
E[T^2] &= \int_{-\infty}^{\infty} t^2 f(t)\,dt \\
&= \int_{20}^{25} t^2(\frac{3}{20})\,dt \ + \int_{25}^{50} t^2(\frac{1}{100})\,dt \\
&=(\frac{t^3}{3})(\frac{3}{20})\Big|_{20}^{25} + (\frac{t^3}{3})(\frac{1}{100})\Big|_{25}^{50} \\
&=(381,25) + (364,5) \\
&=745,75
\end{aligned} \tag{3.3}$$

Para $Desv[T]$ **:**

$$\begin{aligned}
Var[T] &=E[T^2] - E[T]^2 \\
&=745,75 - (26,25)^2 \\
&=56,65
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
Desv[T] &=\sqrt{Var[T]} \\
&=\sqrt{56{,}65} \\
&=7{,}5
\end{aligned} \tag{3.5}$$

Para $P[T > 22]$ **:**

$$\begin{aligned}
P[T > 22] &= \int_{-\infty}^{\infty} f(t)\,dt \\
&= \int_{22}^{\infty} f(t)\,dt \\
&= \int_{22}^{25} (\frac{3}{20})\,dt + \int_{25}^{50} (\frac{1}{100})\,dt \\
&=t(\frac{3}{20})\Big|_{22}^{25} + t(\frac{1}{100})\Big|_{25}^{50} \\
&=(\frac{9}{20}) + (\frac{25}{100}) \\
&=0,7
\end{aligned} \tag{3.6}$$

Para $P[52 \leq T < 60]$ **:**

$$
\begin{aligned}
P[52 \leq T < 60] =& \int_{-\infty}^{\infty} f(t)\,dt \\
=& \int_{52}^{60} 0\,dt \\
=& 0
\end{aligned}
\tag{3.7}
$$

4. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:*"También para un estudio de ergonomía, se necesita caracterizar la estatura en centímetros de un grupo de trabajadores. Un especialista encuentra que la estatura se comporta como variable aleatoria cuya función de densidad de probabilidad es f (x) = 0.00625exp(-0.00625x), para x>0, y en otro caso es 0.
>
> - ¿Estatura esperada?
> - ¿Probabilidad P[X <160]?
> - Con la probabilidad anterior cada 10000 trabajadores se va a encontrar que:

RESPUESTA: **Para hallar la altura:** Se procede a usar la esperanza matemática reemplazando la función dada y se reescribe la constante 0.00625 como c:

$$
\begin{aligned}
E(x) &= \int_{-\infty}^{\infty} xf(x)dx \\
&= \int_{-\infty}^{\infty} x(0{,}00625)e^{-0{,}00625x}dx \\
&= \int_{-\infty}^{\infty} cxe^{-cx}dx \\
&= \lim_{b\to\infty} \int_{0}^{b} cxe^{-cx}dx
\end{aligned}
\tag{4.1}
$$

Es necesario definir parámetros:

$$
\begin{aligned}
f = x \quad g &= \frac{1}{c}e^{-cx} \\
df = dx \quad dg &= e^{-cx}
\end{aligned}
\tag{4.2}
$$

Solucionando la integral y aplicando los parámetros:

$$
\begin{aligned}
&= \lim_{b\to\infty} c\left[x\left(\frac{1}{c}e^{-cx}\right) - \int\left(\frac{1}{c}e^{-cx}\right)\right]\\
&= \lim_{b\to\infty} c\left[\frac{x}{c}e^{-cx} - \frac{1}{c}\left(\frac{1}{c}e^{-cx}\right)\right]_0^b\\
&= \lim_{b\to\infty} \left[\frac{1}{c}(e^{-cx}(cx+1))\right]_0^b\\
&= \lim_{b\to\infty} \left[\frac{1}{c}(e^{-c(\infty)}(c(\infty)+1))\right] - \lim_{b\to\infty}\left[\frac{1}{c}(e^{-c(0)}(c(0)+1))\right]\\
&=\frac{1}{c}\\
&=\frac{1}{0{,}00625}\\
&=160
\end{aligned} \tag{4.3}
$$

Probabilidad P[X <160]: , se usa la ecuación inicial integrando de 0 a 160

$$
\begin{aligned}
P[X<160] &= \int_0^{160} (0{,}00625)e^{-0{,}00625x}dx\\
&= \int_0^{160} ce^{-cx}dx\\
&=c\int_0^{160} e^{-cx}dx\\
&=c\int_0^{-1} -\frac{1}{c}e^{u}du\\
&=c\left(-\int_{-1}^{0} -\frac{1}{c}e^{u}du\right)\\
&=\frac{-c}{c}\left(-\int_{-1}^{0} e^{u}du\right)\\
&=[e^u]_{-1}^0\\
&=1-\frac{1}{e}\\
&=1-0{,}367\\
&=0{,}632
\end{aligned} \tag{4.4}
$$

Por tanto la probabilidad de P[X <160] es de 0.632

5. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:*"Basado en el punto 4, tomando el evento y la probabilidad encontrada, completar la siguiente tabla:
>
> - Al tomar 7 trabajadores, ¿Cuál es la probabilidad que 3 midan menos de 160 centímetros?

Valores Dicotómicos	Proporción
La persona mide menos de 160 centímetros	

- Al tomar 7 trabajadores, ¿Cuál es la probabilidad que entre 1 y 4 trabajadores tengan más de 160 centímetros?
- ¿Cuántos trabajadores de 10000 se esperan encontrar que midan más de 160 centímetros?
- ¿Cuántos trabajadores de los 10000 se esperan encontrar que midan menos de 160 centímetros?

RESPUESTA: Primero se completan los valores de la tabla, la proporción para que una persona mida menos de 160 centímetros es 0.632, de tal forma la proporción para que una persona mida más de 160 centímetros debe ser:

$$1 - 0{,}632 = 0{,}368 \tag{5.1}$$

Valores Dicotómicos	Proporción
La persona mide más de 160 centímetros	0.368
La persona mide menos de 160 centímetros	0.632

Tabla 5.1: Valores Dicotómicos

Al tomar 7 trabajadores, ¿Cuál es la probabilidad que 3 midan menos de 160 centímetros?
Haciendo uso de la distribución binomial , y tomando x como el número de trabajadores, con P=0.632 y n=7 se tiene que:

$$P(x) = \frac{n!}{(n-x)!x!} P^x (1-P)^{n-x} \tag{5.2}$$

$$P(x=3) = \binom{7}{3} (0{,}632)^3 (0{,}368)^4 = 0{,}1620 \tag{5.3}$$

Al tomar 7 trabajadores, ¿Cuál es la probabilidad que entre 1 y 4 trabajadores tengan más de 160 centímetros?
Debido a que tienen que ser más de 160 centímetros se tiene que tomar la probabilidad como p = 0.368:

$$\begin{aligned} P(1 \leq x \leq 4) =& P(x=1) + P(x=2) + P(x=3) + P(x=4) \\ =& \binom{7}{1}(0{,}368)^1(0{,}632)^6 + \binom{7}{2}(0{,}368)^2(0{,}632)^5 + \binom{7}{3}(0{,}368)^3(0{,}632)^4 + \\ & \binom{7}{4}(0{,}368)^4(0{,}632)^3 \\ =& 0{,}1641 + 0{,}2867 + 0{,}2782 + 0{,}1620 \\ =& 0{,}891 \end{aligned} \tag{5.4}$$

¿Cuántos trabajadores de 10000 se esperan encontrar que midan más de 160 centímetros?
Para una distribución binomial la esperanza es:

$$E(x) = np \tag{5.5}$$

Despejando n:

$$n = \frac{E(x)}{p} \tag{5.6}$$

Debido a que la probabilidad de que midan más de 160 cm es 0.368 se soluciona para el número de trabajadores:

$$\begin{aligned} n =& \frac{E(x)}{p} \\ =& \frac{160}{0{,}368} \\ \approx& 435 \end{aligned} \tag{5.7}$$

¿Cuántos trabajadores de los 10000 se esperan encontrar que midan menos de 160 centímetros?
Como la probabilidad de que midan menos de 160 cm es 0.632 se tiene que:

$$\begin{aligned} n =& \frac{E(x)}{p} \\ =& \frac{160}{0{,}632} \\ \approx& 253 \end{aligned} \tag{5.8}$$

6. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:* ün Ingeniero sabe que 2 trabajadores salen a cumplir una cita médica en horas laborales por un mes. ¿Cuál es la probabilidad de encontrar más de 25 trabajadores que salen a cumplir una cita médica en hora laborales por un año?

RESPUESTA: Es necesario el uso de la distribución de probabilidad de Poisson:

$$P(x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{6.1}$$

Primero se tiene que calcular λ para un año, por tanto:

$$\lambda = 2\frac{trab.}{mes} = 2\frac{trab.}{Mes} \times 12\frac{mes}{año} = 24\frac{trab.}{año} \tag{6.2}$$

Dado que el problema solicita que sean más de 25 empleados, se tiene que calcular la suma de la probabilidad de 0 a 24 y luego se tiene que restar del total 1, por conveniencia se uso R con el siguiente codigo:

```
1 - sum(dpois(1:24, lambda = 24))
```

Lo cual da como resultado lo siguiente:

$$1 - 0{,}5540012 = 0{,}4459988 \tag{6.3}$$

Por tanto la probabilidad de encontrar más de 25 empleados en una cita médica al año es de 0.4459988.

7. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

> "*Tarea:* ". Si el tiempo que toma un estudiante de especialización de llegar de su trabajo es modelo normal con media 45 minutos y desviación 5 minutos, entonces, de cada 100 veces que va el estudiante del trabajo a la universidad __ veces tarda entre 42 y 56 minutos.

RESPUESTA: Se tienen los siguientes datos:

$$\mu = 45m \ \ \sigma = 5m \ \ e = 100 \rightarrow 1 \tag{7.1}$$

Y se tiene que encontrar:

$$P(42m \leq t \leq 56m) \tag{7.2}$$

Dado que es un intervalo de una distribución normal se rige por la ecuación :

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < t < \frac{b-\mu}{\sigma}\right)$$

Con a = 42 y b = 56 Por simplicidad de los cálculos se uso R con el siguiente código:

```
a = pnorm(56, mean = 45, sd = 5, lower.tail=TRUE)
b = pnorm(42, mean = 45, sd = 5, lower.tail=TRUE)
c = a - b
```

Lo cual da como resultado c = 0.7118434, bien esto es solo la probabilidad, por tanto se tiene que multiplicar por el número de eventos e:

$$c \times e = 0{,}7118434 \times 100 \approx 71 \tag{7.3}$$

En conclusión el estudiante llegara 71 veces entre 42 y 56 minutos.

8. Problema

OBJETIVO El siguiente enunciado ha sido dado por el taller:

"*Tarea:* "sea una variable aleatoria discreta cuya función de probabilidad sea:

x	-3	-2	-1	0	1
P(x)	0,13	0,24	0,15	0,1	Y

O en otros casos, donde Y es un valor entre 0 y 1. Calcular

- E[x]
- Var[x]
- P[-2<x<4]
- P[x $\geq \frac{1}{2}$]
- P[x = -3 o x = -1 o x = 1]

RESPUESTA:

Para determinar el valor de (Y), se debe tener en cuenta que la probabilidad total es igual a uno.

$$P(1) = 1 - (0,13 + 024 + 0,15 + 0,1) = 0,38 \tag{8.1}$$

De esta manera se pude completar la tabla anterior, determinando los valores de las variables discretas.

x	-3	-2	-1	0	1
P(x)	0,13	0,24	0,15	0,1	0,38

A partir de los valores de tabla procedemos a determinar la esperanza matemática.

$$\mathrm{E}[X] = \sum_{i=1}^{n} x_i \, \mathrm{P}[X = x_i] \tag{8.2}$$

Calculando la esperanza matemática a partir del siguiente código en R.

```r
#DEFINICION DE VALORES
VALORES <- c(-3, -2, -1, 0, 1)
#DEFINICION DE PROBABILIDADES
PROBABILIDAD <- c(.13, .24, .15, .1, .38)
#CALCULO DEL VALOR DE ESPERANZA MATEMATICA
weighted.mean(VALORES, PROBABILIDAD)
[-64]
```

La varianza esta derminada por:

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] && (8.3)\\
&= \mathrm{E}[(X^2 - 2X\mu + \mu^2)] && (8.4)\\
&= \mathrm{E}[X^2] - 2\mu\,\mathrm{E}[X] + \mu^2 && (8.5)\\
&= \mathrm{E}[X^2] - 2\mu^2 + \mu^2 && (8.6)\\
&= \mathrm{E}[X^2] - \mu^2 && (8.7)\\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2 && (8.8)\\
& && (8.9)
\end{aligned}
$$

A partir de

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 \tag{8.10}$$

Se calcula E[x^2]

```
#DEFINICION DE VALORES
VALORES <- c(-3, -2, -1, 0, 1)
VALORES_AL_CUADRADO<-VALORES**2
#DEFINICION DE PROBABILIDADES
PROBABILIDAD <- c(.13, .24, .15, .1, .38)
#CALCULO DEL VALOR DE ESPERANZA MATEMATICA
weighted.mean(VALORES_AL_CUADRADO, PROBABILIDAD)
[2,66]
```

De los valores obtenidos

$$E[X]E[X^2] \tag{8.11}$$

Se procede a calcular la varianza para las variables aleatorias.

$$
\begin{aligned}
\mathrm{Var}(X) &= 2{,}66 - 0{,}4096\\
&= 2{,}2504
\end{aligned} \tag{8.12}
$$

- P[-2<X<4]

$$P[-2 < X < 4] = P[X = -2] + \cdots + P[X = 4] \tag{8.13}$$

$$
\begin{aligned}
&= 0,24 + 0,15 + 0,1 + 0,38 + 0 + 0 + 0\\
&= 0,87
\end{aligned} \tag{8.14}
$$