2015

Universidad Politécnica Salesiana

ESTADÍSTICA Y PROBABILIDAD CON R.

ING. WILLAM CAIZA

Copyright © 2015 por Ing. William Caiza. Todos los derechos reservados.

DEDICATORIA

Quiero dedicar el presente trabajo a mis padres Yola Taco y Alejandro Caiza, cuyos esfuerzos diarios me han dado ejemplos de superación, dedicación y ante todo de amor.

PRÓLOGO

El presente trabajo consta de seis capítulos, en el capítulo uno se hace una introducción a conjuntos e inducción matemática en el mismo que se describe brevemente las leyes y operaciones con conjuntos, en inducción matemática se hace una descripción de uso y aplicación de la inducción; En el capítulo dos se trata sobre la estadística descriptiva, poniendo énfasis en los conceptos fundamentales como es la muestra, la media, la desviación estándar, etc... En el capítulo tres se trata sobre la probabilidad, se hace énfasis en la probabilidad clásica y probabilidad frecuentista se hace un desarrollo minucioso de las propiedades de la probabilidad hasta el tratamiento de la probabilidad de Bayes. En el capítulo cuatro se trata sobre las variables aleatorias, función de probabilidad, función de distribución, se hace un estudio completo de las esperanzas y varianzas de las variables aleatorias.

ÍNDICE

Contenido

CAPÍTULO 1: CONJUNTOS E INDUCCIÓN MATEMÁTICA 1

1.1 Conjuntos 2

- 1.1.1 Operaciones con conjuntos 4
- 1.1.2 Propiedades de las operaciones con conjuntos 4

1.2 Inducción matemática 12

CAPÍTULO 2: ESTADÍSTICA DESCRIPTIVA 16

2.1 Estadística descriptiva ..... ¡Error! Marcador no definido.

- 2.1.1 Definiciones Preliminares 17
- 2.1.2 Descripción de datos 36

2.2 Medidas de tendencia central 19

- 2.2.1 Media muestral 19
- 2.2.2 Moda muestral 22
- 2.2.3 Mediana muestral 23

2.3 Medidas de dispersión 25

- 2.3.1 Rango 25
- 2.3.2 Varianza muestral 26
- 2.3.3 Desviación estándar 28

2.4 Medidas de posición 31

- 2.4.1 Cuartiles 31
- 2.4.2 Deciles 32
- 2.4.3 Percentil 32

2.5 Medidas de forma 45

- 2.5.1 Asimetría 45
- 2.5.2 Curtosis 46

2.6 Análisis de datos agrupados ..... ¡Error! Marcador no definido.

2.7 Diagrama de caja 49

CAPÍTULO 3: PROBABILIDAD 53

3.1 Fundamentos de la teoría de probabilidad 55

- 3.1.1 Permutaciones 55
- 3.1.2 Combinaciones 57

3.2 Espacio muestral (Ω) **58**
- 3.2.1 Conjunto Partes de Ω (subconjuntos) 59

3.3 Eventos **59**
3.4 Probabilidad de eventos **60**
- 3.3.4 Propiedades de los eventos simples 63

3.5 Axiomas de la probabilidad **64**
3.6 Propiedades de la probabilidad **64**
- 3.6.1 Probabilidad de un evento nulo ..... **¡Error! Marcador no definido.**
- 3.6.2 Probabilidad del evento complemento 64
- 3.6.3 Probabilidad de eventos incluidos 65
- 3.6.4 La probabilidad de un evento está entre 0 y 1 65
- 3.6.5 Probabilidad de la diferencia de eventos 65
- 3.6.6 Regla aditiva de probabilidad de Eventos 65

3.7 Independencia de eventos **67**
3.8 Condicionalidad de eventos **67**
3.9 Probabilidad total **74**
- 3.9.1 Teorema de Bayes 75

CAPÍTULO 4: VARIABLES ALEATORIA **80**
4.1 Definición de Variable aleatoria **81**
- 4.1.1 Función de distribución v.a.d (variable aleatoria discreta) 82

4.2 ESPERANZAS, VARIANZAS Y CORRELACIONES DE UNA Y DOS VARIABLES ALEATORIAS **85**
- 4.2.1 CORRELACIÓN 91
- 4.2.2 ESPERANZA MATEMÁTICA 92
- 4.2.3 ESPERANZA MATEMÁTICA DE UNA VARIABLE ALEATORIA CONTINUA 93
- 4.2.4. PROPIEDADES DE LA ESPERANZA MATEMÁTICA 94
- 4.2.5 LA VARIANZA 95

4.3 FUNCIÓN GENERADORA DE MOMENTOS **100**
- 4.3.1. DISTRIBUCIÓN DE POISSON ..... **¡Error! Marcador no definido.**
- 4.3.2 DISTRIBUCIÓN UNIFORME DISCRETA ..... **¡Error! Marcador no definido.**
- 4.3.3. DISTRIBUCIÓN HIPERGEOMÉTRICA ..... **¡Error! Marcador no definido.**
- 4.3.4. DISTRIBUCIÓN GEOMÉTRICA ..... **¡Error! Marcador no definido.**
- 4.3.5. DISTRIBUCIÓN BINOMIAL NEGATIVA ..... **¡Error! Marcador no definido.**
- 4.3.6. DISTRIBUCIÓN DE BERNOULLI ..... **¡Error! Marcador no definido.**
- 4.3.7. DISTRIBUCIÓN BINOMIAL ..... **¡Error! Marcador no definido.**

CAPÍTULO 5: DISTRIBUCIONES DE VARIABLES ALEATORIAS 109

5.1 Distribución variable aleatoria discretas 110

- 5.1.1 Distribución Binomial y Bernoulli ..... **¡Error! Marcador no definido.**
- 5.1.3 Distribución hipergeométrica 115
- 5.1.4 Distribución geométrica 119
- 5.1.5 Distribución de poisson 121
- 5.1.6 Distribución binomial negativa 125

5.2 Distribución Variable Aleatoria Continuas 130

- 5.2.1 Distribución Uniforme 130
- 5.2.2 Distribución exponencial 133
- 5.2.3 Distribución normal 138
- 5.2.4. Teorema del Límite Central 153
- 5.2.5 Distribución gamma ..... **¡Error! Marcador no definido.**

5.3 EJERCICIOS PROPUESTOS DISTRIBUCION WEIBULL ..... ¡Error! Marcador no definido.

6. EJERCICIOS RESUELTOS ..... ¡Error! Marcador no definido.

6.1 INDUCCIÓN MATEMÁTICA ..... ¡Error! Marcador no definido.

6.2 ANÁLISIS EXPLORATORIO DE DATOS ..... ¡Error! Marcador no definido.

6.3 TABLA DE DISTRIBUCIONES ..... ¡Error! Marcador no definido.

6.4 CURTOSIS Y ASIMETRIA ..... ¡Error! Marcador no definido.

6.4 DISTRIBUCIÓN BINOMIAL NEGATIVA ..... ¡Error! Marcador no definido.

6.5 VARIABLES ALEATORIAS CONTINUAS (ESPERANZA Y VARIANZA) .... ¡Error! Marcador no definido.

6.6 DISTRIBUCION DE WEIBULL ..... ¡Error! Marcador no definido.

6.7 PROBABILIDAD CONDICIONAL ..... ¡Error! Marcador no definido.

6.8 BAYES ..... ¡Error! Marcador no definido.

7. ANEXOS ..... ¡Error! Marcador no definido.

ANEXO 1 ..... ¡Error! Marcador no definido.

- R CÓDIGO ABIERTO ..... **¡Error! Marcador no definido.**

Anexo 2 ..... ¡Error! Marcador no definido.

- INTERFASE GRAFICA EN R ..... **¡Error! Marcador no definido.**

Anexo 3 ..... ¡Error! Marcador no definido.

- EJEMPLOS ..... **¡Error! Marcador no definido.**

Índice de Tips R

Tips R 1. Unión de Conjuntos 5
Tips R 2. Intersección de Conjuntos 6
Tips R 3. Diferencia de Conjuntos 7
Tips R 4. Inducción Matemática ..... **¡Error! Marcador no definido.**
Tips R 5. Estadística descriptiva 19
Tips R 6. Promedio 21
Tips R 7. Promedio 21
Tips R 8. Moda 23
Tips R 9. Frecuencia ..... **¡Error! Marcador no definido.**
Tips R 10. Mediana 25
Tips R 11. Rango 26
Tips R 12. Rango ..... **¡Error! Marcador no definido.**
Tips R 13. Varianza 27
Tips R 14. Varianza 27
Tips R 15. Desviación estandar ..... **¡Error! Marcador no definido.**
Tips R 16. Medidas de posición 33
Tips R 17. Permutación 57
Tips R 18 Combinación. 58
Tips R 19 Distribución Binomial 113
Tips R 20. Distribución Binomial 114
Tips R 21. Distribución Binomial ..... **¡Error! Marcador no definido.**
Tips R 22. Distribución Hipergeométrica. 117
lustración 24Tips R 23. Distribución Hipergeométrica. 118
Tips R 25. Distribución Geométrica. 121
Tips R 26. Distribución Poisson. 123
Tips R 27. Distribución Poisson. 123
Tips R 28. Distribución Poisson. 124
Tips R 29. Distribución Binomial Negativa. ..... **¡Error! Marcador no definido.**
Tips R 30. Distribución Uniforme. 133
Tips R 31. Distribución Exponencial. 136

Índice de Gráficas

Gráfica 1 Clasificación de los Números 1
Gráfica 2 Ejercicio de conjuntos ..... **¡Error! Marcador no definido.**
Gráfica 3 Unión de dos intervalos 3
Gráfica 4 Unión Exclusiva 3
Gráfica 5 Unión de Conjuntos. 4
Gráfica 6 Intersección de Conjuntos 5
Gráfica 7 Diferencia de Conjuntos 6
Gráfica 8 Estadística Descriptiva e Inferencial 17
Gráfica 9 Tabla de variables Cualitativas 18
Gráfica 10 Tabla de variables Cuantitativas 18
Gráfica 11 Tabla simbología ..... **¡Error! Marcador no definido.**
Gráfica 12 Asimetría. 45
Gráfica 13 Curtosis 46
Gráfica 14 Diagrama de cuartiles 50
Gráfica 15 Probabilidad de n=1 54
Gráfica 16 Probabilidad de n=100 54
Gráfica 17 Demostración Probabilidad total. 74
Gráfica 18 Función Distribución. 82
Gráfica 19 Correlación lineal negativa. ..... **¡Error! Marcador no definido.**
Gráfica 20 Correlación lineal positiva ..... **¡Error! Marcador no definido.**
Gráfica 21 Variables no correlacionadas. ..... **¡Error! Marcador no definido.**
Gráfica 22 Variables no lineal. 91

CAPÍTULO UNO

Conjuntos e Inducción Matemática

CONTENIDO

1.0 Clasificación de los números

1.1 Conjuntos

- **1.1.1 Operaciones con conjuntos**
- **1.1.2 Propiedades de las operaciones con conjuntos**

1.2 Inducción matemática

CAPÍTULO 1

CONJUNTOS E INDUCCIÓN MATEMÁTICA

1.0 Clasificación de los Números

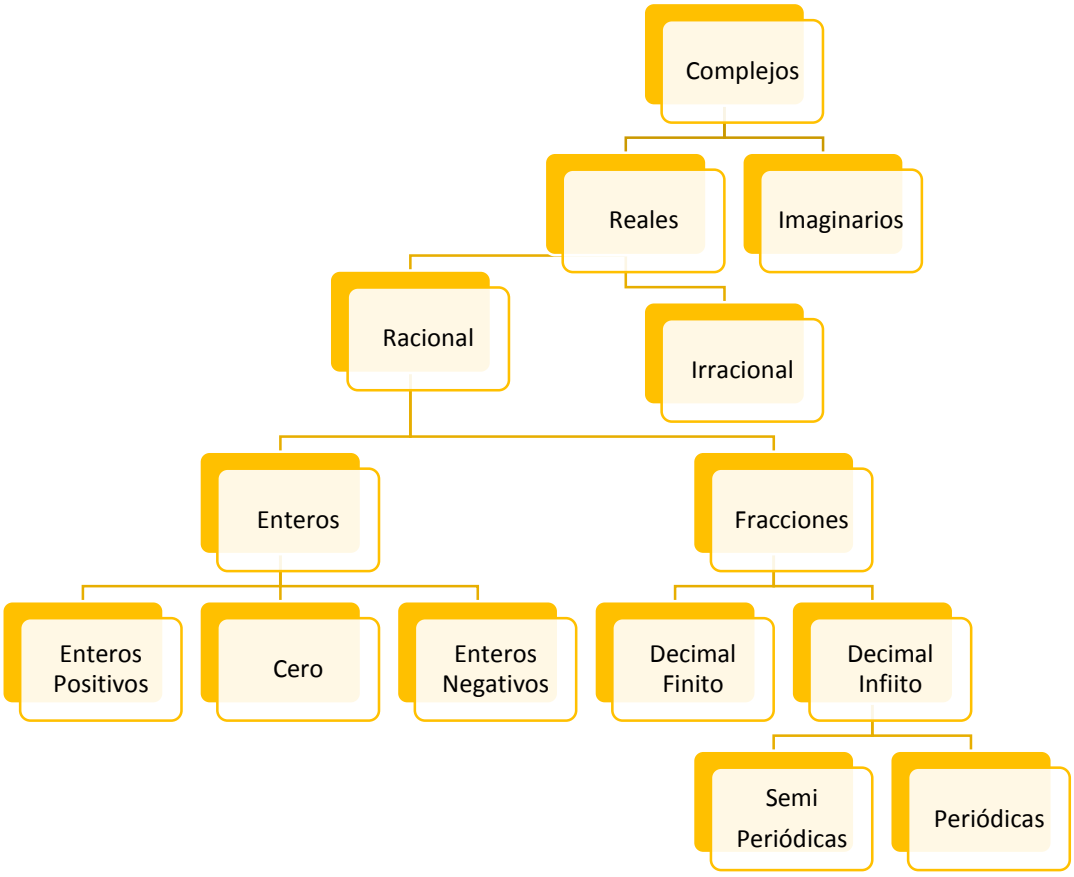


Gráfico 1 Clasificación de los Números

En el nivel superior se encuentra los números complejos, un ejemplo podría ser $3 + 4\,i$, donde la parte real es 3 y la parte imaginaria es 4.

Todo número real se puede clasificar en un número racional e irracional, un número racional es de la forma $\frac{\boldsymbol{m}}{\boldsymbol{n}}, \boldsymbol{m}, \boldsymbol{n} \in \boldsymbol{Z}$ y un número es irracional cuando no se puede escribir de la forma $\frac{m}{n}$, ejemplo $\pi = 3.141516$ ….

Todo número racional se puede clasificar en entero y fraccionario, los números enteros se clasifican en enteros positivos o naturales, cero y enteros negativos; los fraccionarios se clasifican en decimal finito e infinito, de decimal finito si el residuo es cero.

Las fracciones de decimal infinito se clasifican en periódicas y semi-periódicas.

Las fracciones periódicas como $3{,}3333 \ldots = 3,\hat{3}$, se puede obtener su fracción que es igual al cociente cuyo numerador es igual al número menos la parte periódica (33-3), y el denominador es tantos nueves como cifras (una) tenga la parte periódica (9).

$$3,\hat{3} = \frac{33-3}{9} = \frac{30}{9}.$$

Las fracciones semi-periódicas como $3{,}2455555 \ldots = 3{,}24\hat{5}$, se puede obtener su forma fraccionaria, cuyo numerador es el número menos el número sin la parte periódica (3245-324), el denominador es tantos nueves como cifras tenga la parte periódica y tantos ceros como cifras tenga la parte no periódica (dos).

$$3{,}24\hat{5} = \frac{3245 - 324}{900} = \frac{2921}{900}$$

1.1 Conjuntos

Un conjunto es la agrupación de ciertos objetos unidos por un criterio en común.

Como por ejemplo, tenemos:

a) $A_1 = \{a, e, i, o, u\}\,;\ A = \{las\ vocales\}$

b) $B_1 = \left\{\frac{1}{x}\ x \in N,\ x < 5\right\}\,;\ B = \left\{\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\right\}$

Los conjuntos se pueden definir por extensión y comprensión.

- Un conjunto se define **por extensión** cuando se enumera cada uno de los elementos del conjunto.
- Un conjunto se define **por comprensión** cuando se describe la característica que debe cumplir dicho conjunto

Ejercicio 1:

Dado el conjunto $A = \{x \ / \ x \in \mathrm{R}, \mathrm{x}^3 - \mathrm{x} = 0\}$, definido por comprensión, expréselo por extensión.

Desarrollo

$x^3 - x = 0$

$x\,(x^2 - 1) = 0$; factorando

$x\,(x+1)(x-1) = 0$; factorando

$x = 0 \quad \vee \quad x + 1 = 0 \quad \vee \quad x - 1 = 0$; relación lógica

$x = 0 \qquad x = -1 \qquad x = 1$

La solución del sistema es

A= {0, -1, 1}.

Ejercicio 2:

Dados los intervalos A= [3,5]; B= [4,7] encontrar:

a) **A ∪ B**

Desarrollo

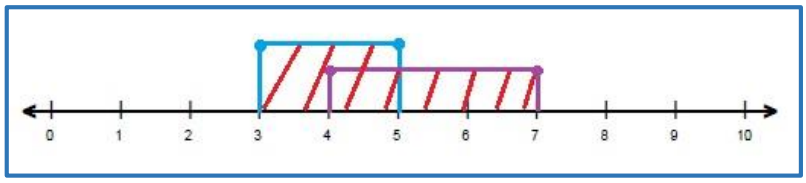


Gráfico 2: Unión de dos intervalos

A U B = $[3,7]$

b) **A Ú B**

Desarrollo

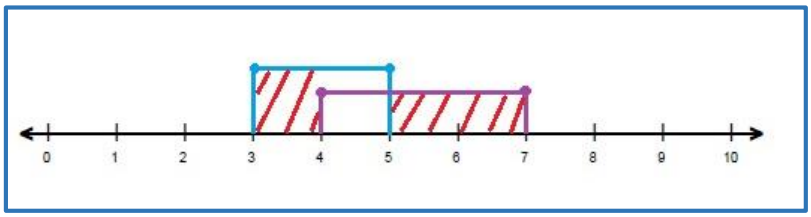


Gráfico 3: Unión Exclusiva

A Ú B = $[3,4) \cup (5,7]$

Nota: En el ejercicio se realiza una unión con tilde o lógicamente es una disyunción exclusiva, que indica lo uno o lo otro pero no los dos a la vez.

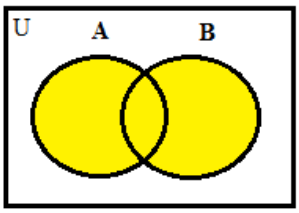
1.1.1 Operaciones con conjuntos

a) Unión de conjuntos

La Unión de dos conjuntos es el conjunto formado por todos los elementos que pertenecen a ambos conjuntos. La unión de A y B se denota A∪B.

$$A \cup B = \{x/x \in A \, \sigma \, x \in B\}$$

$$A \cup B$$



Gráfica 4: Unión de Conjuntos.

TIPS DE "R" CONJUNTOS

Para este tema necesitaremos dos paquetes, los mismo que pueden ser instalados al digitar las siguientes expresiones:

install.packages('gplots')

library(gplots)

Se utiliza el comando ***union (conjunto 1, conjunto2)*** para obtener la unión de los conjuntos, como se indica a continuación.

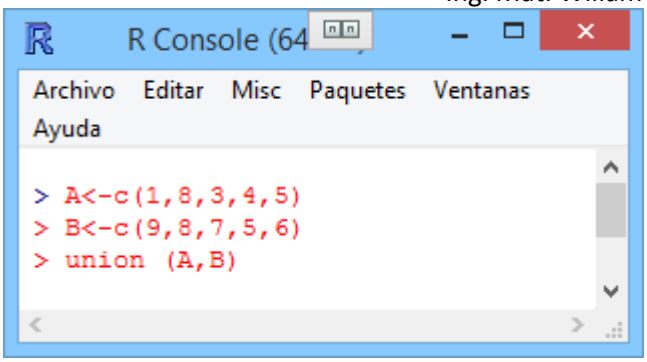


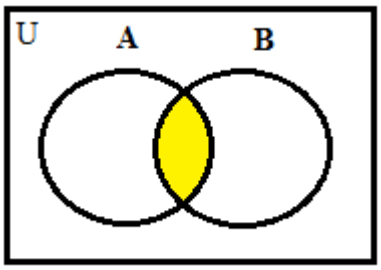
Ilustración 1 de R: Unión de Conjuntos

b) Intersección de conjuntos

La intersección de dos conjuntos es el conjunto formado por los elementos que tienen en común. La intersección de A y B se denota $A \cap B$

$$A \cap B = \{x/x \in A \text{ y } x \in B\}$$

$$A \cap B$$



Gráfica 5: Intersección de Conjuntos

TIPS DE "R" CONJUNTOS

Intercesión de conjuntos.

intersect(conjunto 1,conjunto2)

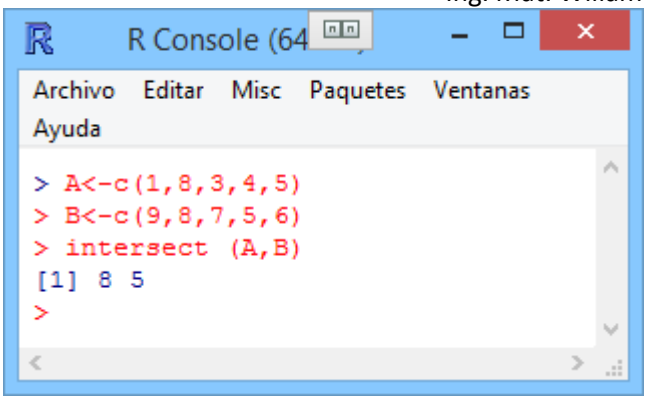


Ilustración 2 de R: Intersección de Conjuntos

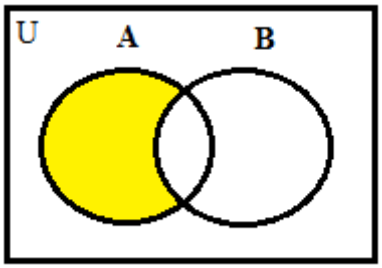
Utilizamos la función ***intersect (conjunto 1, conjunto2),*** con la cual obtenemos la intersección entre los conjuntos especificados.

c) Diferencia entre dos conjuntos

Dados dos conjuntos A y B, se llama diferencia entre el conjunto A y el conjunto B al conjunto formado por todos los elementos pertenecientes al conjunto A que no pertenecen al conjunto B.

$$A/B = \{x/x \in A \text{ y } x \text{ no es } \notin B\}$$

$$A/B$$



Gráfica 6: Diferencia de Conjuntos

TIPS DE "R" CONJUNTOS

Diferencia de conjuntos.

setdiff(conjunto 1,conjunto2)

setdiff(conjunto 2,conjunto1)

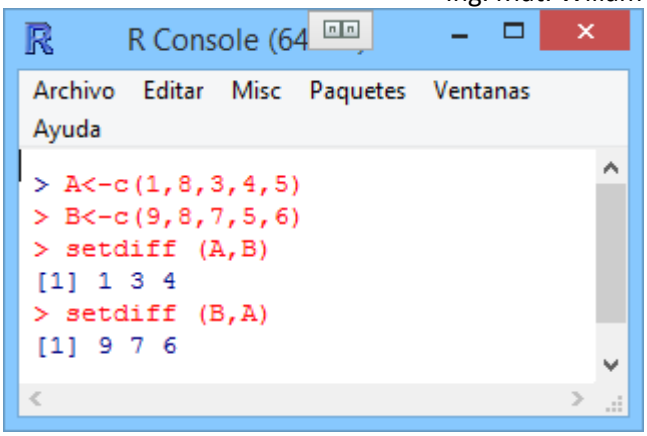


Ilustración 3 de R: Diferencia de Conjuntos

Utilizamos la función **setdiff (conjunto 1, conjunto2)** obtendremos la diferencia entre los conjuntos especificados.

Ejercicio 3:

Dado los conjuntos $A = \{1,3,5\}$; $B = \{2,9\}$ encontrar: A/B

Desarrollo

$A/B = \{1,3,5\}$

d) Complemento de un conjunto

El complemento de A, está formado por todos aquellos elementos que no pertenecen a A, es decir: $A^C = U/A$

e) Diferencia simétrica de dos conjuntos

La diferencia simétrica notada por $A \Delta B$, se define por $C = A \Delta B = \{x/x \in (A/B) \cup (B/A)\}$.

Dado los conjuntos A y B, el conjunto C diferencia simétrica está constituido por todos los elementos del conjunto A que no pertenecen al conjunto B, unidos a todos los elementos del conjunto B que no pertenecen al conjunto A.

Ejercicio 4:

Dados los conjuntos $M = \{1, 3, 6, 8, 9\}$; $N = \{2, 4, 6, 8\}$ encontrar:

$M \Delta N$

Desarrollo

$$M/N = \{1, 3, 9\}$$
$$N/M = \{2, 4\}$$
$$M \Delta N = \{1, 2, 3, 4, 9\}$$

Nota: Para encontrar M/N , se coloca todos los elementos de M y se va retirando los elementos de N , de la misma forma para N/M se coloca todos los elementos de N y se va retirando los elementos de M, por ultimo uno los dos conjuntos y se obtiene la diferencia simétrica de A y B.

1.1.2 Propiedades de las operaciones con conjuntos

1) **Idempotencia**

$$A \cup A = A$$
$$A \cap A = A$$

2) **Conmutativa**

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

3) **Asociativa**

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

4) **Distributiva**

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

5) Identidad

$A \cup \emptyset = A$ $\quad A \cup U = U$

$A \cap \emptyset = \emptyset$ $\quad A \cap U = A$

6) Complemento

$A \cup A^C = U$ $\quad (A^C)^C$= A $\quad \emptyset^C = U$

$A \cap A^C = \emptyset$ $\quad U^C = \emptyset$

7) Morgan

$(A \cup B)^C = A^C \cap B^C$

$(A \cap B)^C = A^C \cup B^C$

8) Diferencia

A / B = A $\cap$ B^C

Ejercicio 5:

a) Demostrar que (AUB)/C = (AUBUC)/C

Desarrollo (1ra Forma)

(AUBUC)/C	Hipótesis
(AUBUC) ∩ C’	Definición de diferencia
((AUB) UC) ∩ C’	Asociativa
((AUB) ∩C’) U (C ∩ C’)	Distributiva
((AUB) ∩C’) U ϕ	Complemento
(AUB) ∩C ‘	Identidad
(AUB)/C	Diferencia

Desarrollo (2da Forma)

(AUB)/C	Hipótesis

(AUB) ∩ C’	Diferencia
((AUB) ∩ C’) U (C ∩ C’)	Identidad
((AUB) UC) ∩ C’	Distributiva
(AUBUC) ∩ C’	Asociativa
(AUBUC) /C	Diferencia

a) Demuestre que [(AUC)/B]U[(BUC)/A]=(AUBUC)/(A∩ B)

Desarrollo

(AUBUC)/(A ∩ B)	Hipótesis
(AUBUC) ∩ (A∩B)’	Diferencia
(AUBUC) ∩ (A’UB’)	Ley de Morgan
[(AUBUC) ∩A’]U [(AUBUC) ∩ B’]	Distributiva
[(A U (BUC)) ∩ A’]U [(B U (AUC)) ∩ B’]	Conmutativa, Asociativa
[(A ∩ A’) U ((BUC) ∩ A’)] U [(B ∩ B’) U ((AUC) ∩ B’)]	Distributiva
[ϕ U ((BUC) ∩ A’)] U [ϕ U ((AUC) ∩B’)]	Complemento
[(BUC) ∩A’]U [(AUC) ∩B’]	Identidad
[(BUC)/A] U [(AUC)/B]	Diferencia
[(AUC)/B] U [(BUC)/A]	Conmutativa

EJERCICIOS PROPUESTOS

1) Determinar por comprensión los siguientes conjuntos:

a) A ={ lunes, martes, miércoles, jueves, viernes}

b) B= { b, a, n ,d ,e ,r }

R: a) A= {x/x días laborales}

c) B= { x/x letras de la palabra bandera}

2) Determine por extensión los siguientes conjuntos:

a) U= { x ∈ ℕ / x es impar }

b) P = { x/x es un numero par; 2 ≤ x ≤ 8 }

c) O= { x/x ∈ ℕ/; x < 7}

d) W = { números de una cifra}

R: a) U= {1,3; 5; 7…}

b) P = {2; 4; 6; 8}

c) O= {6, 5,4; 3; 2; 1; 0}

d) W = {0,1; 2; 3; 4,5; 6; 7; 8; 9}

3) Sean los conjuntos A y B hallar la gráfica y simbólicamente A U B y A ∩ B

A= {x/x las vocales}

B= {x/x letras de las palabras nadar**}**

4) Dado los conjuntos A= {1,3,5,6} y B= { 3,5,7} encuentre A Δ B

R: A Δ B = {1, 6,7}

5) Demostrar:

a) A ∩ (B \ A) = Ø

b) A ∪ (B \ A) = A ∪ B

a) A ∩ (B \ A) = A ∩ (B ∩ A^c)	Diferencia de conjuntos
= A ∩ (A^c ∩ B)	Conmutatividad de la unión
= (A ∩ A^c) ∩ B	Asociativa de la intersección
= Ø ∩ B	Leyes del complementario
= Ø	Leyes de identidad

b) A ∪ (B \ A) = A ∪ (B ∩ Ac)	Diferencia de conjuntos
= (A ∪ B) ∩ (A ∪ Ac)	Distributiva
= (A ∪ B) ∩ U	Leyes del complementario
= A ∪ B	Leyes de identidad

1.2 Inducción Matemática

Dado una expresión matemática, que sea dependiente de un número natural n, su valor de verdad puede ser demostrada mediante el principio de inducción matemática, es decir: la inducción matemática nos permite verificar si la expresión es verdadera, siempre y cuando dicha expresión sea función de números naturales, los pasos para la demostración son:

1) Se comprueba la validez de la afirmación para n=1

2) Se supone la validez de esta afirmación para n=k, nuestra Hipótesis Inductiva

3) Se demuestra la validez de esta afirmación para n=k+1, nuestra Tesis Inductiva, tomando en consideración su validez para n= k, después de lo cual se concluye que la expresión es válida para cualquier número natural n.

Ejercicio 6:

Demostrar que para todo número natural se verifica la siguiente igualdad

$$\boldsymbol{1+2+3+4\ldots+n=\frac{n(n+1)}{2}=S_n}$$

Desarrollo

1. Se verifica el cumplimiento para n=1 (n= número de términos)

$$1=\frac{(1+1)}{2}=\frac{2}{2}=1$$

Se podría verificar, para más de un término (n=3 términos)

$$1+2+3=\frac{3(3+1)}{2}=\frac{12}{2}=6$$

$$6=6$$

2. Suponer que para **n=k** es verdadera (Hipótesis Inductiva)

$$1+2+3+4\ldots+k=\frac{k(k+1)}{2}\quad hipotesis\ Inductiva$$

3. Se debe demostrar para **n=k+1**, la expresión sea verdadera (Tesis Inductiva)

$$1+2+3+4\ldots+k+(k+1)=\frac{(k+1)(k+2)}{2}\quad TI$$

Se debe empezar la demostración con la hipótesis inductiva

$$1+2+3+4\ldots+k=\frac{k(k+1)}{2}\quad H.I$$

$$1+2+3+4\ldots+k+(k+1)=\frac{k(k+1)}{2}+(k+1)$$

$$=\frac{k(k+1)+2(k+1)}{2}\quad=\quad\frac{(k+2)(k+1)}{2}\quad LQQD$$

Nota: Por lo tanto se demuestra que la expresión anterior es válida para **n=k+1**, concluyendo que la misma es verdadera para todos los números naturales.

Ejercicio 7:

Verifique que para todo número natural n se cumple n≤2^{n-1}

Desarrollo

1. Se verifica el cumplimiento para n=1, reemplace

$$1 \leq 2^{1-1}$$

$$1 \leq 2^{0}$$

$$1 \leq 1$$

2. Suponemos que para n=k es verdadero (Hipótesis Inductiva)

$$k \leq 2^{k-1}$$

3. Se debe demostrar que para n=k+1 (Tesis Inductiva), la expresión es verdadera

$$k+1 \leq 2^{(k+1)-1}$$
$$\mathrm{k}+1 \leq 2^{\mathrm{k}} \quad \mathrm{T.I}$$

i. Se comienza la demostración utilizando la hipótesis inductiva, detallada a continuación:

$$k \leq 2^{k-1} \quad (1)\ H.I$$

ii. Conocemos que k+1 es menor o igual a 2k

$$k+1 \leq 2k \qquad (2)$$

iii. Multiplicando la expresión (1) y (2) tenemos que:

$$k(k+1) \leq 2k(2^{k-1});\ k > 0$$

$$k+1 \leq 2(2^{k-1})$$

$$k+1 \leq 2^{k} \qquad LQQD.$$

Nota: Por lo tanto se demuestra que para **n=k+1** es verdadera concluyendo, que la expresión anterior es válida para todos los números naturales.

EJERCICIOS PROPUESTOS

Fuente: SÁEZ Eduardo, SZANTÓ Iván, inducción matemática, Ejercicios 3 y 4, literales 1, 4 y 1 respectivamente. pág. 7 y 8.

1) **Pruebe que la fórmula $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \ldots + n \cdot (n+1) = \frac{n(n+1)(n+2)}{3}$ es válida para todo natural n.**

2) **Demuestre que para todo natural n, $n^5 - n$ es divisible por 5.**

R: $n^5 - n$ es divisible por 5

3) **Demuestre por inducción las siguientes igualdades**

a) $1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(n+2)}{6}$

b) $(n+1)(n+2)(n+3)\ldots(n+n) = 2^n * 1 * 3 * 5 * (2n-1)$

c) $1^2 - 2^2 + 3^2 - 4^2 + \ldots + (-1)^{n-1} n^2 = (-1)^{n-1} \frac{n(n+1)}{2}$

d) $(1 - \frac{1}{4})(1 - \frac{1}{9})\ldots(1 - \frac{1}{(n+1)^2}) = \frac{n+2}{2n+2}$

SEGUNDO CAPÍTULO

Estadística Descriptiva

CONTENIDO

2.1 Estadística descriptiva

- **2.1.1 Definiciones Preliminares**
- **2.1.2 Descripción de datos**

2.2 Medidas de tendencia central

- **2.2.1 Media muestral**
- **2.2.2 Moda muestral**
- **2.2.3 Mediana muestral**

2.3 Medidas de dispersión

- **2.3.1 Rango**
- **2.3.2 Varianza muestral**
- **2.3.3 Desviación estándar**

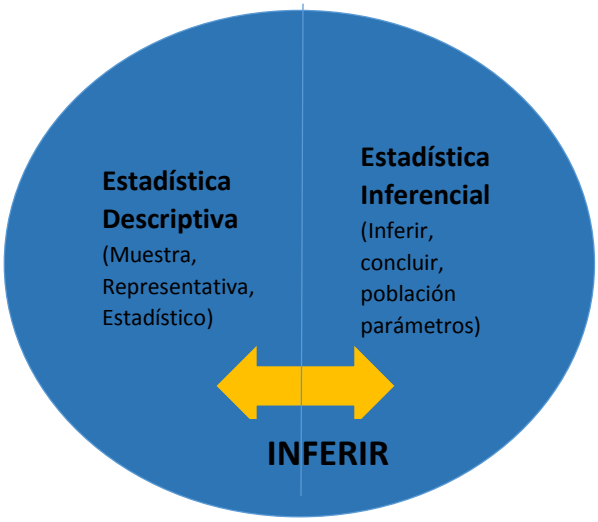
2.4 Medidas de posición

- **2.4.1 Cuartiles**
- **2.4.2 Deciles**
- **2.4.3 Percentil**

2.5 Medidas de forma

- **2.5.1 Asimetría**

CAPÍTULO 2
ESTADÍSTICA DESCRIPTIVA



Gráfica 7: Estadística Descriptiva e Inferencial

2.1 Definiciones Preliminares

1. **Población:** Es el conjunto de todos los elementos correspondientes al experimento analizado, población no es sinónimo de un conjunto inmenso, si no que depende del experimento analizado.
2. **Muestra:** Es una parte de la población y esta debe ser representativa, la misma que se escogerá según una metodología, que asegure la propiedad.
3. **Variable:** Son características de un experimento, las variables se clasifican en:

CUALITATIVAS.- Describen características que no son medibles (cualidades).	
Se clasifica según una escala de medición.	
Nominales.- Los datos pueden ser contados pero no pueden ser ordenados o medidos.	**Ordinales**.-Los datos pueden ser contados y ordenados pero no pueden ser medidos. Los datos deben tener un orden natural.
Ningún atributo	**Atributo: orden**
Ejemplo **Género:** Masculino, femenino **Religión:** Católica, Testigos de Jehová	**Ejemplo** **Estudios:** Primaria, secundaria, universidad **Estaciones del año:** primavera, verano, otoño, invierno
Dicotómicas: Cuando tenemos dos categorías.	
Politómicas: Más de dos categorías.	

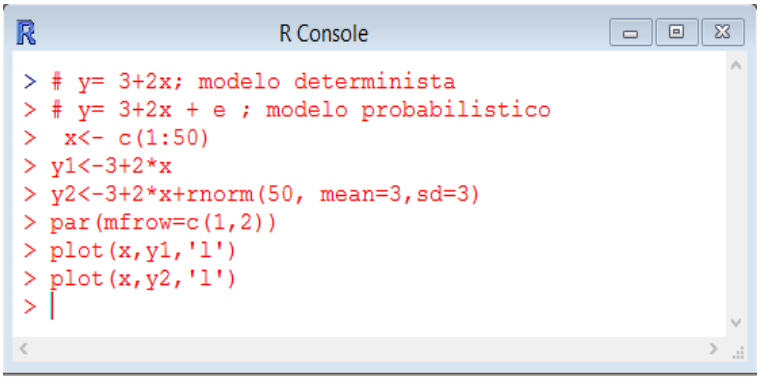
Gráfica 8: Tabla de variables Cualitativas

CUANTITATIVAS (NUMÉRICAS).- Describen características que son medibles.	
Se clasifican según una escala de medición.	
Intervalo.- Tienen la particularidad de tener "cero relativo", no significa carencia de la característica, depende de la escala de medición.	**Razón**.- Tiene la particularidad de tener "cero absoluto", significa carencia de la característica.
Atributo: Orden y distancia	**Atributo: Orden, distancia y origen.**
Ejemplo **temperatura:** -10°, 0°,10°,20°	**Ejemplo** **Peso:** 0kg, 150 kg, 200 kg. **Número de hijos:** 0, 1, 2
Continuas (medibles): Se refiere a variables que se pueda medir.	
Discretas (contar): Se refiere a variables que se pueda contar.	

Gráfica 9: Tabla de variables Cuantitativas

4. **Modelo Matemático:** Es la representación de una realidad, dicha realidad se representa mediante expresiones matemáticas.
5. **Modelo Determinístico:** Es aquel donde los parámetros están determinados o son fijos, **Ejemplo**: x = v t, se puede observar en la expresión anterior, no depende o no es función de un error, por lo tanto es determinista.

TIPS DE "R" CONJUNTOS



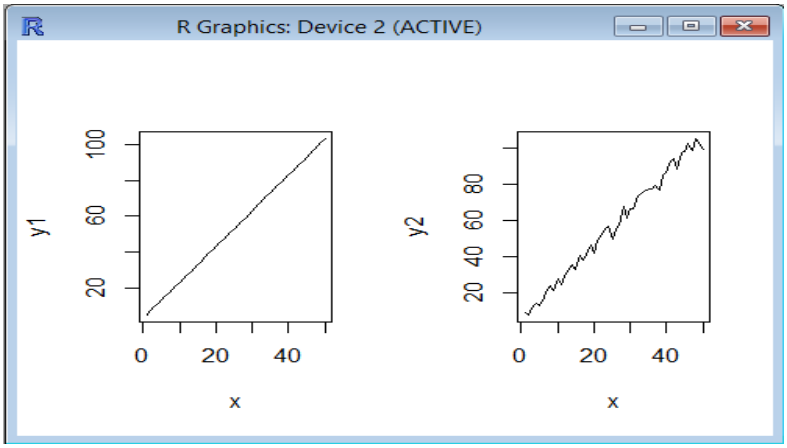


Ilustración 4 de R: Modelo probabilístico.

6. **Modelo Probabilístico:** Es un modelo de la realidad, la cual es aleatoria o estocástica, **Ejemplo**: y = a + bx + error, por lo tanto el modelo es probabilístico.
7. **Estadística Descriptiva:** Es la ciencia que recolecta datos, los analiza, los ordena, los procesa (mediante herramientas matemáticas), finalmente obteniendo información para una toma de decisiones.
8. **Estadística Inferencial:** Se basa en una muestra, la misma que debe ser representativa, obtenida mediante metodologías, en base a la muestra se puede inferir (concluir) las características de la población.
9. **Estadístico:** Son las características de una muestra.
10. **Parámetro:** Son las características de la población.

2.2 Medidas de tendencia central

2.2.1 Media muestral

Matemáticamente la media muestral se expresa:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \dots \dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} xi$$

La media muestral, se podría considerar como el valor representativo de un conjunto de datos analizado.

Ventajas de la media muestral

- Es fácil de calcular.
- En la gráfica de frecuencia representa el centro de gravedad, el mismo que equidista de los demás puntos.

Desventajas de la Media Muestral

- Sensibilidad a valores extremos (muy altos o muy bajos).
- No es recomendable emplearla en distribuciones muy asimétricas

Ejercicio 8:

Dados los siguientes datos: 2, 6, 11, 8, 11, 4, 7, 5, 90, encuentre la media muestral**.**

Desarrollo

$$\bar{x} = \frac{2+6+11+8+11+4+5+5+90}{9} = 16$$

Se podría decir que 16, es el valor representativo del conjunto de datos.

En el conjunto de datos se observa valores atípicos por lo que se va a calcular una media al 30%, podríamos asumir que los valores atípicos son **2, 4** y **90** del conjunto de datos.

Calculando la media al 30%, se tiene lo siguiente:

$100\% \quad \rightarrow \quad 9$

$30\% \quad \rightarrow \quad x \qquad \rightarrow x = 3$

Lo que indica que tres valores no se tomaran en cuenta en el cálculo.

$$\bar{x}_{30\%} = \frac{6+11+8+11+5+7}{6} = 8$$

Lo que indica que el valor representativo, sin los valores atípicos es 8.

Observación: Cuando se tiene valores atípicos es recomendable utilizar una media cortada, generalmente a un 5 y 10%.

TIPS DE "R" MEDIDAS DE TENDENCIA CENTRAL

Promedio.

Variable <-mean(datos)

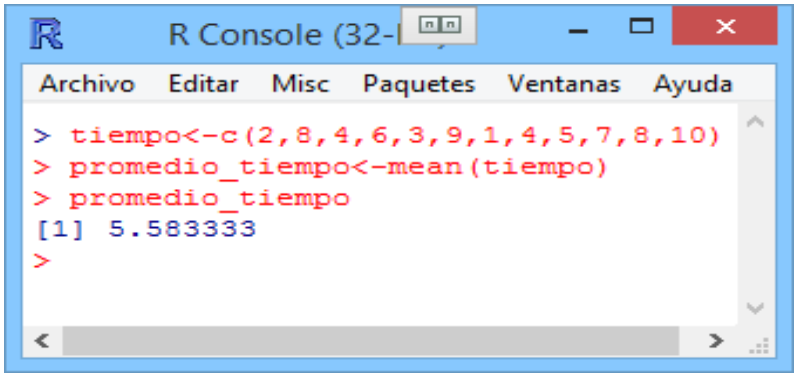


Ilustración 5 de R: Promedio

Utilizamos la función **mean (datos),** para calcular el promedio de los datos ingresados.

TIPS DE "R" MEDIDAS DE TENDENCIA CENTRAL

Función promedio

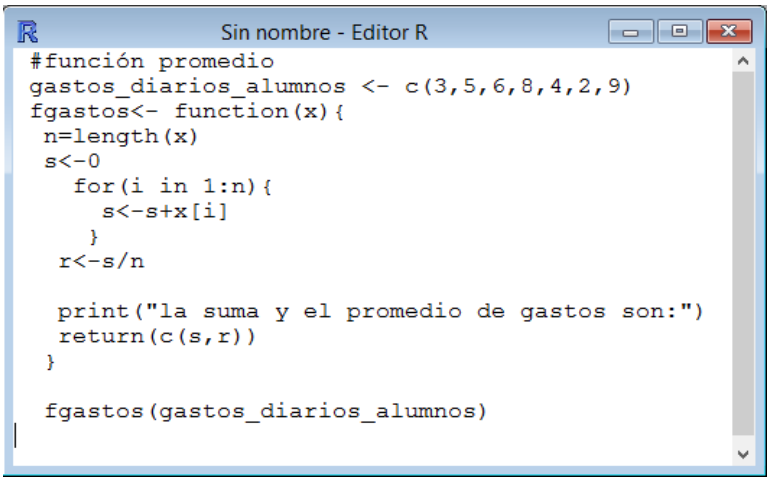


Ilustración 6 de R: Función de promedio en R.

2.2.2 Moda muestral

La moda muestral es una medida de tendencia central, es el dato que más se repite, más de uno en el conjunto observaciones. Representado por Mo, el cual no es único.

Ventajas de la moda muestral:

- Es útil cuando hay agrupaciones con diferentes valores
- Fácil de reconocer.

Desventajas de la moda muestral:

- Puede no existir en algunos datos
- Puede existir más de una moda
- Puede estar demasiado lejos del centro de gravedad de los datos
- En distribuciones muy asimétricas suele ser un dato muy poco representativo

Ejercicio 9:

Se tiene el siguiente conjunto de edades de una escuela: 7, 7, 5, 4, 6

Desarrollo

La moda del conjunto de datos es Mo=7.

Nota: Con los datos anteriores se puede verificar que la moda de ese conjunto es la edad 7, ya que es la que tiene mayor frecuencia. A la vez si se tuviese el siguiente conjunto de edades, 7, 7, 6, 6, 4, 5, 8, tendríamos como moda las edades 7 y 6 ya que son las que tienen mayor frecuencia e iguales.

TIPS DE "R" MEDIDAS DE TENDENCIA CENTRAL

Para este tema necesitaremos dos paquetes, los mismo que pueden ser instalados al digitar las siguientes expresiones indicadas abajo:

```
install.packages('modeest')

library(modeest)
```

La función **"mfv(datos)"** es la que se ha de utilizar para encontrar la moda de un conjunto de datos.

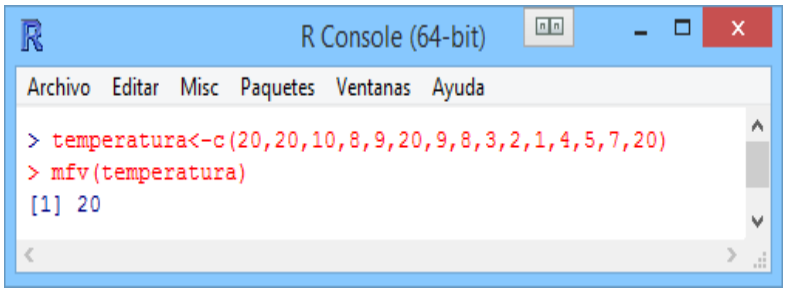


Ilustración 7 de R: Moda de un conjunto de datos.

2.2.3 Mediana muestral

Es el valor a partir del cual el 50% de datos están sobre ese valor y el otro 50% de datos están bajo ese valor, no es sensible a datos atípicos (Md).

La mediana se puede calcular a partir de datos que deben estar ordenados de menor a mayor, cuyo cálculo se puede realizar mediante la siguiente expresión.

$$Md = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & Si\ n\ (\#\ de\ terminos) es\ impar \\ \frac{1}{2}\left(X_{\frac{n}{2}} + X_{\left(\frac{n}{2}+1\right)}\right) & Si\ n\ es\ par \end{cases}$$

Ventajas de la mediana muestral:

- No es sensible a los valores extremos.
- Es fácil de interpretar.
- Es recomendable para distribuciones muy asimétricas.

Desventajas de la mediana muestral:

- Se deben ordenar los datos para el cálculo.
- Los valores extremos pueden ser importantes.
- Se emplea solo en variables cuantitativas.

Ejercicio 10:

Para datos impares

Dado los siguientes datos: 2, 2, 3, 4, 5 encuentre la mediana muestral.

Desarrollo

$$\boldsymbol{X_1 = 2}$$

$$\boldsymbol{X_2 = 2}$$

$$\boldsymbol{X_3 = 3}$$

$$\boldsymbol{X_4 = 4}$$

$$\boldsymbol{X_5 = 5}$$

$Md = X_{\left(\frac{n+1}{2}\right)}$

$Md = X_{\left(\frac{5+1}{2}\right)} = X_3 = 3$

Para datos pares

Dado los siguientes datos: 3, 3, 4, 4, 4, 5, 6, 7, 8, 23, encuentre la media muestral.

Desarrollo

3, 3, 4, 4, 4, 5, 6, 7, 8, 23

$Md = \frac{1}{2}(x_5 + x_6)$

$Md = \frac{1}{2}(4 + 5) = 4{,}5$

Ejercicio 11:

Se tiene 20 datos, obtener el valor del conjunto de datos en la posición indicada.

Desarrollo

Se ordena de menor a mayor como se indica a continuación, y se procede a realizar el cálculo:

$X_{25\%} = X_{0,25*20} = X_5$ = es el valor que está en la posición quinta del conjunto de datos.

$X_{60\%} = X_{0,60*20} = X_{12}$

$X_{33\%} = X_{0,33*20} = X_7$

$$X_{75\%} = X_{0,75*20} = X_{15}$$

2.3 Medidas de dispersión

TIPS DE "R" MEDIDAS DE TENDENCIA CENTRAL

Mediana.

En el cálculo de la mediana de un conjunto de datos, se utiliza la función **"median(datos)".**

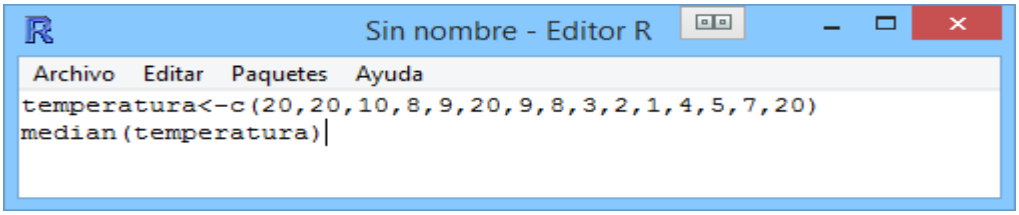


Ilustración 8 de R: Mediana en R.

2.3.1 Rango

Es una medida de dispersión (diferencia) es igual a su valor máximo menos su valor mínimo, representa la máxima dispersión existente en el conjunto de datos.

Habitualmente el rango se utiliza para el cálculo de la desviación empírica, que es igual a:

$$s\ empírica \approx \frac{R}{4}$$

Se utiliza la desviación empírica cuando inicialmente en una investigación no se posee ninguna información.

Ventajas del rango

- Útil cuando se quiere conocer la extensión de las variaciones extremas (valor máximo de la dispersión).
- Fácil de calcular.

Desventajas del rango

- No es una medida de dispersión con respecto al centro de la distribución.
- Solo emplea dos valores en su cálculo.

TIPS DE "R" MEDIDAS DE DISPERSIÓN

Rango.

Para el cálculo del rango se utiliza la función " **range (datos) "**

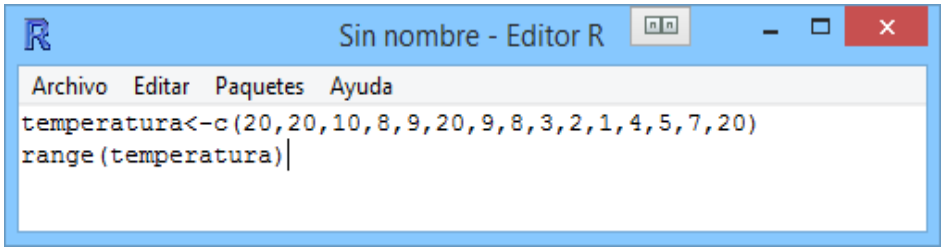


Ilustración 9 de R: Rango en R

2.3.2 Varianza muestral

Es un valor representativo de la dispersión de un conjunto de datos con respecto a la media, su cálculo se puede realizar:

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

Ventajas varianza muestral

- Es útil cuando se compara la variabilidad de dos o más conjuntos de datos.
- Utiliza toda la información disponible

Desventajas varianza muestral

- No proporciona ayuda inmediata cuando se estudia la dispersión de un solo conjunto de datos, no hay con quien comparar.
- Difícil de interpretar por tener sus unidades elevadas al cuadrado.

Ejercicio 12:

Dados los siguientes datos: 11, 2, 6, 8, 11, 4, 7, 5, encuentre la desviación estándar.
Desarrollo

$$\bar{x} = \frac{11+2+6+8+11+4+7+5}{8} = 6{,}75$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_1 - \bar{x})^2}{n-1}$$

$$s^2 = \frac{(11-6{,}75)^2 + (2-6{,}75)^2 + (6-6{,}75)^2 + (11-6{,}75)^2 + (8-6{,}75)^2 + (4-6{,}75)^2 + (7-6{,}75)^2 + (5-6{,}75)^2}{7}$$

$s^2 = 10{,}21 \rightarrow$ *Es el valor representativo de la dispersión de ese conjunto de datos.*

TIPS DE "R" MEDIDAS DE DISPERSIÓN

Varianza.

La función "**var (datos)"** se utiliza para calcular la varianza de un conjunto de datos.

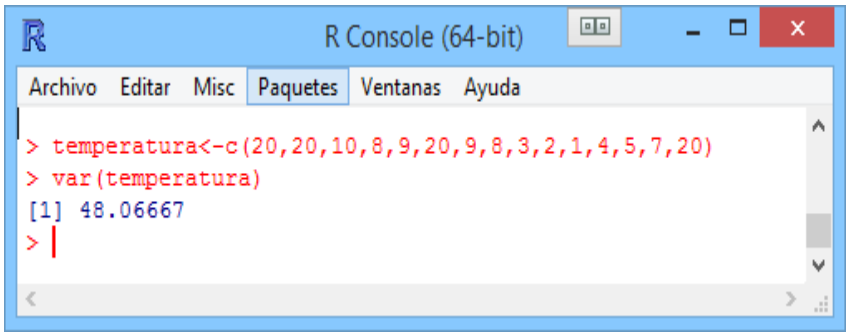


Ilustración 10 de R: Varianza en R

TIPS DE "R" MEDIDAS DE DISPERSIÓN

Programa de R, para calcular la varianza.

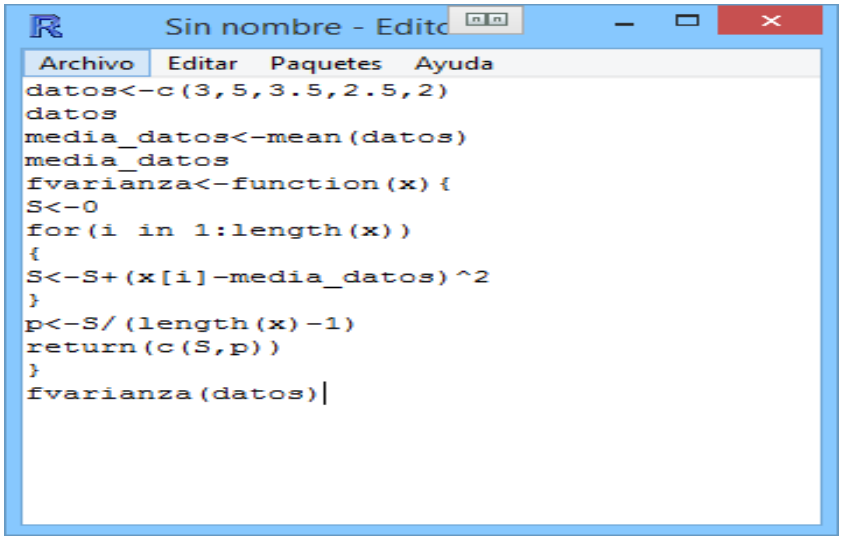


Ilustración 11 de R: Función propia en R. de la varianza

2.3.3 Desviación estándar

Es el valor representativo de la dispersión de un conjunto de datos con respecto a la media, la diferencia entre la varianza y la desviación estándar es que esta no tiene sus unidades elevadas al cuadrado, se la representa mediante **"s".**

Matemáticamente:

$$s^2 = Var(x)$$

$$s = \sqrt{Var(x)}$$

Observación: se puede notar que las unidades de la desviación estándar corresponden a las unidades de los datos.

Ventajas desviación estándar

- Esta expresada en las mismas unidades que la variable en estudio.
- Utiliza todas las observaciones en su cálculo.
- Fácil de interpretar.

Desventajas desviación estándar

- Es sensible a las unidades de medida.

Ejercicio 13:

Demostrar:

$$\boldsymbol{s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}}$$

Desarrollo

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n-1}$$

$$s^2 = \frac{\sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2}{n-1}$$

$$s^2 = \frac{\sum x_i^2 - 2\bar{x}\sum x_i + \sum \bar{x}^2}{n-1}$$

$$s^2 = \frac{\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}{n-1}$$

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

Nota: $\bar{x} = \frac{\sum x_i}{n}$

$\sum x_i$ = n$\bar{x}$

Ejercicio 14:

Encontrar la varianza y la desviación de las siguientes edades del salón de clases.

x = "edad"

$x_1 = 22$

$x_2 = 21$

$x_3 = 23$

$x_4 = 22$

$x_5 = 22$

$x_6 = 19$

$\bar{x} = 21.5$; 21.5 años es la edad representativa del salón de clases.

Desarrollo (Primera forma)

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{(22-21{,}25)^2 + (21-21{,}25)^2 + (23-21{,}25)^2 + (22-21{,}25)^2 + (22-21{,}25)^2 + (19-21{,}25)^2}{5}$$

$s^2 = 1{,}9$

$s = \sqrt{1{,}9} = 1.38$; 1.38 años es la dispersión existente entre cualquier par de edades del salón de clases.

Desarrollo (Segunda forma)

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

$$s^2 = \frac{[(22)^2 + (21)^2 + (23)^2 + (22)^2 + (22)^2 + (19)^2] - 6(21{,}5)^2}{5}$$

$s^2 = 1{,}9$

$s = \sqrt{1{,}9} = 1{,}38$

Nota: Indica que la diferencia promedio entre par de edades es aproximadamente de 1.38 años.

TIPS DE "R" MEDIDAS DE DISPERSIÓN

Desviación Estándar.

La función que utiliza R, para el cálculo de la desviación estándar es **"sd (datos)"**.

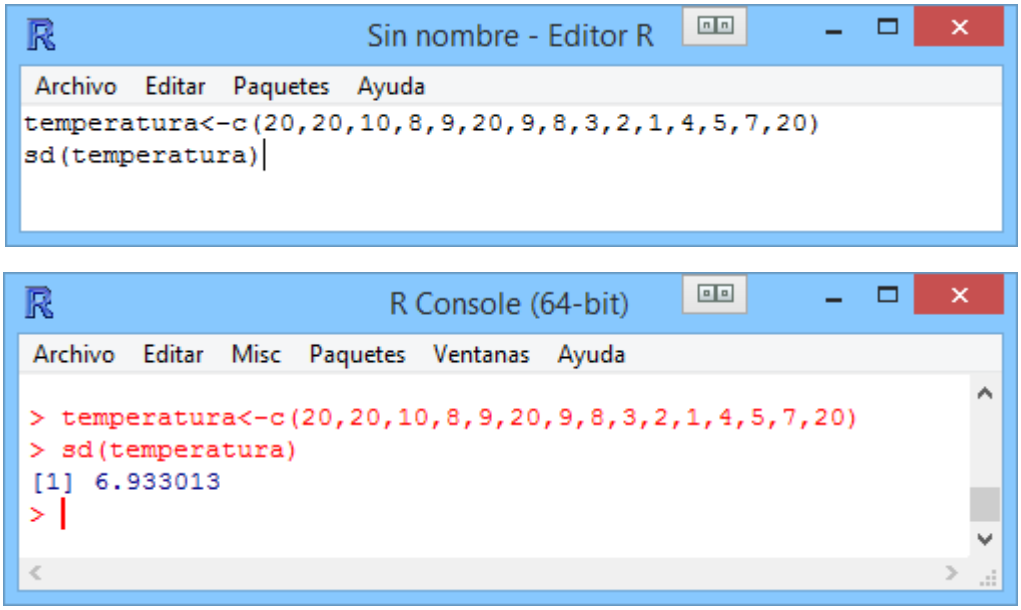


Ilustración 4 Desviación estándar en R

2.3.4 Intervalo de confianza

$$[\bar{x} \pm k\, S_{\bar{x}}] \qquad (1)$$

La expresión (1), representa el intervalo de confianza de las medias, k es el cuantil de la distribución normal y $S_{\bar{x}}$ es la desviación estándar de la media.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

$$\left[\bar{x} - k\frac{S}{\sqrt{n}} \; ; \; \bar{x} + k\frac{S}{\sqrt{n}}\right]$$

Ejercicio 15:

Calcular el intervalo de confianza de las edades del salón de clases, al 95% de confianza tenemos:

$$\left[21.5 - k\frac{1.38}{\sqrt{6}}S_{\bar{x}}\,;\ 21.5 + k\frac{S_{\bar{x}}}{\sqrt{6}}\right]$$

$$\left[21.5 - \left(1.96\frac{1.38}{\sqrt{6}}\right)\ ;\ 21.5 + \left(1.96\frac{1.38}{\sqrt{6}}\right)\right]$$

$$[20.4\ años\ \ ;\ \ 22.6\ años]$$

En un 95% las edades del salón de clases se encuentra entre 20.4 años a 22.6 años.

2.4 Medidas de Posición (Cuantiles)

Indica la posición del valor del dato analizado, dentro del grupo y debe estar ordenado de menor a mayor.

2.4.1 Cuartiles

Q1: Es el valor a partir del cual el 25% de datos están bajo ese valor, y el 75% de datos sobre ese valor.

Q2: es el valor a partir del cual el 50% de datos están bajo ese valor, y el 50% de datos sobre ese valor.

Q3: es el valor a partir del cual el 75% de datos están bajo ese valor, y el 25% de datos sobre ese valor.

Ventajas de los cuartiles

- Ayudan a describir la posición que tiene cada valor específico en relación al conjunto de datos.

Desventajas de los cuartiles

- Si la distribución es simétrica, los cuartiles deben estar a la misma distancia de la mediana.
- Cuando el número de observaciones es impar, la observación del medio es la mediana.
- El rango entre cuartiles no es muy conocido.
- Los cuartiles son sensibles a valores externos

2.4.2 Deciles

D1: es el valor a partir del cual el 10% de datos están bajo ese valor, y el 90% de datos sobre ese valor, existen diez deciles: D1, D2,…, D10.

Ventajas de los deciles

- Nos dice como está posicionado un valor
- Determina el valor que ocupa una posición

Desventajas de los deciles

- No reflejan ninguna tendencia central, sino una posición de la distribución de los datos.
- Debemos ordenar los datos antes de cualquier cálculo.

2.4.3 Percentil

Hay cien percentiles, y cada uno posiciona un valor dentro del rango.

P1: es el valor a partir del cual el 1% de datos están bajo ese valor, y el 99% de los datos sobre ese valor, así cada uno de los 99 percentiles restantes.

Ventajas de los percentiles

- Nos dice la posición de un valor en relación a una muestra.
- Está dividido en partes y cada una representa un percentil
- Es fácil y muy útil de determinar mediante la fórmula de posición
- Se utiliza para analizar datos

Desventajas de los percentiles

- Se debe organizar la muestra para poder calcular su posición

TIPS DE "R" MEDIDAS DE POSICION

Cuantiles

Para calcular el cuantil de un conjunto de datos utilizamos la función "**quantile(datos)**".

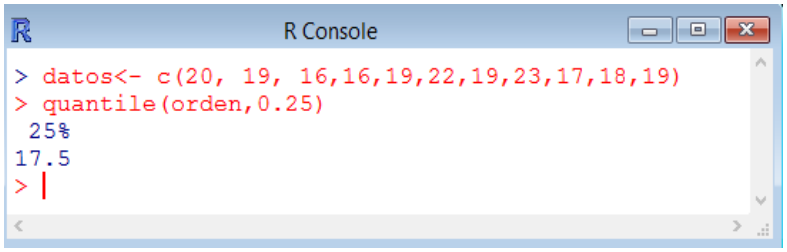


Ilustración 13 de R. Cálculo del cuantil en R.

EJERCICIOS PROPUESTOS

(Estadística - Métodos y aplicaciones, 2da Edición, Edwin Galindo, Ejercicios: Análisis Exploratorio de datos pag. 43, 44,45, literales: (5, 10, 13,14, 20, 21, 23, 25 ,26).

1) Un inversor tiene ahorros repartidos en 3 depósitos con 2000, 5000 y 10 000 dólares, respectivamente. Si el primero le rinde un 5% anual, el segundo un 4% anual y el tercero un 2% anual. ¿Cuál es el tipo de interés que recibe?

Respuesta: 2.94%

2) Dados los datos y sus frecuencias

X_i	1	3	6	9	10
f_i	8	20	25	10	2

Halle:

a) Q2

b) Media

c) S

d) RIQ

Respuesta: a) Q2=6; b) x=5.046; c) S=2.63; d) RIQ = 3

3) La tabla muestra la temperatura nocturna (en º C) durante 200 días.

Intervalo	Frecuencia	Intervalo	Frecuencias
2 – 4	21	12 – 14	14
4 – 6	16	14 – 16	20
6 – 8	15	16 – 18	22
8 – 10	26	18 – 20	18

10 – 12	23	20 – 22	25

Encuentre:

a) Media
b) Mediana
c) cuartiles inferior y Superior

Respuesta: x=12,29; Q1 = 7.73; Q2 = 11.91; Q3 = 17.36

4) Un automóvil ha recorrido los 832 km que separan Loja de Esmeraldas, permutando regularmente las 5 llantas (incluida la de emergencia) para que todas tengan igual desgaste. ¿Cuál es el recorrido promedio de cada llanta?

Respuesta: 665.6 km

5) EL kilometraje que marca un auto, luego de 4 años de uso, es 100 mil kilómetros, Si el dueño lo compro nuevo y lo hace descansar 1 día, luego de usarlos 4 días seguidos. ¿Cuál es el recorrido promedio diario de los días manejados, considerando años de 365 días?

Respuesta: 85.62 km/día

6) Se tiene 4 números. Al añadir el promedio de 3 de ellos al número restante, se obtienen los números 17, 21, 23, 29. Si se excluye al mayor de estos números, ¿Cuál es el promedio de los 3 restantes?

Respuesta: 8

7) El promedio de 53 números es 600. Si se elimina 3 números consecutivos, se observa que el nuevo promedio aumenta en 5%. ¿Cuál es el mayor de dichos números consecutivos?

Respuesta: 101

2.5 Descripción de datos con datos individuales y agrupados

La descripción de los datos, se la realiza mediante una tabla de frecuencias, tanto para datos individuales como agrupados, la misma que consta de la siguiente información:

- **Frecuencia (f_i):** Es el número de veces que se repite un dato; también es el número de datos de contiene la clase i.
- **Frecuencia Relativa (f_{ri}):** Es el tanto por uno, se obtiene dividiendo la frecuencia del dato para el número total de datos; también es el número total de datos de la clase i, dividida para el número total de datos $f_{ri} = \frac{f_i}{n}$.
- **Frecuencia Acumulada (F_i):** Es igual al número de datos hasta un dato en particular; también es igual al número de datos hasta la clase i.
- **Frecuencia Relativa Acumulada (F_{ri}):** Es el tanto por uno, hasta la clase i, y es igual a frecuencia acumulada hasta la clase i dividido para el número de datos $F_{ri} = \frac{F_i}{n}$.
- **Tabla de frecuencia:** Es una representación agrupada de los datos para facilitar su interpretación, la tabla de frecuencias puede ser desarrollada mediante datos individuales o datos agrupados (clases).

Ejercicio 16:

Dados los siguientes datos, correspondientes a los tiempos de servicio en minutos por día, calcular e interpretar la tabla de frecuencia con datos individuales

Tiempo de servicio por persona	Personas atendidas	Tiempo total de servicio	frecuencia relativa del tiempo total de servicio	frecuencia relativa de las personas atendidas	Frecuencia Acumulada de tiempo total de servicio	Frecuencia Acumulada de las personas atendidas	Frecuencia Acumulada relativa del tiempo total de servicio	Frecuencia relativa acumulada del tiempo las personas atendidas
1,8	1	1,8	0,012	0,03	1,8	1	0,012	0,03
2,1	1	2,1	0,013	0,03	3,90	2	0,025	0,05
2,2	1	2,2	0,014	0,03	6,10	3	0,039	0,08
2,5	2	5	0,032	0,05	11,10	5	0,071	0,13
2,7	1	2,7	0,017	0,03	13,80	6	0,088	0,15
2,8	2	5,6	0,036	0,05	19,40	8	0,124	0,20
2,9	2	5,8	0,037	0,05	25,20	10	0,162	0,25
3,1	2	6,2	0,040	0,05	31,40	12	0,201	0,30

3,5	2	7	0,045	0,05	38,40	14	0,246	0,35
3,6	4	14,4	0,092	0,10	52,80	18	0,338	0,45
3,7	2	7,4	0,047	0,05	60,20	20	0,386	0,50
3,8	1	3,8	0,024	0,03	64,00	21	0,410	0,53
3,9	1	3,9	0,025	0,03	67,90	22	0,435	0,55
4,1	2	8,2	0,053	0,05	76,10	24	0,488	0,60
4,2	2	8,4	0,054	0,05	84,50	26	0,542	0,65
4,3	1	4,3	0,028	0,03	88,80	27	0,569	0,68
4,4	1	4,4	0,028	0,03	93,20	28	0,597	0,70
4,5	1	4,5	0,029	0,03	97,70	29	0,626	0,73
4,7	1	4,7	0,030	0,03	102,40	30	0,656	0,75
4,8	1	4,8	0,031	0,03	107,20	31	0,687	0,78
4,9	2	9,8	0,063	0,05	117,00	33	0,750	0,83
5,1	3	15,3	0,098	0,08	132,30	36	0,848	0,90
5,6	1	5,6	0,036	0,03	137,90	37	0,884	0,93
5,7	1	5,7	0,037	0,03	143,60	38	0,921	0,95
6,1	1	6,1	0,039	0,03	149,70	39	0,960	0,98
6,3	1	6,3	0,040	0,03	156,00	40	1,000	1,00
	40	**156**						

Análisis de la tabla

El total de tiempo de servicio de las tres persona es de 15.3 minutos, cada una de ellas fueron atendidas en 5.1 minutos, estos 15.3 minutos me representan el 9.8% del tiempo total ocupado del día de servicio, las tres personas representan el 8% del total de personas atendidas durante el día de servicio.

Hasta los 132.3 minutos del tiempo de atención fueron atendidas 36 personas, este tiempo representa el 84.8% del trabajo neto de atención, las 36 personas representan el 90% de las personas atendidas durante el día de trabajo.

Ejercicio 17:

Obtenga la Tabla de Frecuencia, mediante datos agrupados para los siguientes 40 datos de una muestra, correspondiente al tiempo (minutos) que se utilizó para atender a las personas en una estación de servicio durante un día.

Desarrollo

Datos:

3,1	4,9	2,8	3,6
4,5	3,5	2,8	4,1
2,9	2,1	3,7	4,1
2,7	4,2	3,5	3,7
3,8	2,2	4,4	2,9
5,1	1,8	2,5	6,3

2,5	3,6	5,6	4,8
3,6	6,1	5,1	3,9
4,3	5,7	4,7	3,6
5,1	4,9	4,2	3,1

Número de clase = $\sqrt{número\ de\ datos}$

Número de clase = $\sqrt{40}$ = 6,32455532 = 6

Nota: El número de clases también se puede calcular mediante la siguiente expresión:

NC = 1 + 3.3 Log (N)

Donde:

NC = Número de Intervalos
Log = Logaritmo base 10
N = Número de datos

Longitud del histograma = Valor máximo – Valor mínimo

Longitud del histograma= 6,3 – 1,8 = 4,5

Ancho de la clase= $\frac{\text{longitud del histograma}}{\#\ de\ clases}$

Ancho de la clase= $\frac{4,5}{6} = 0,75 = 0,8$

Tabla de frecuencia

Clase i	lim.inferior	lim.superior	fi	fri	F	Fri
1	1,8-0,05=1,75	1,75+0,8=2,55	5	0,13	5	0,12
2	2,55	3,35	7	0,18	12	0,30
3	3,35	4,15	12	0,3	24	0,6
4	4,15	4,95	9	0,22	33	0,82
5	4,95	5,75	5	0,12	38	0,95
6	5,75	5,75+0,8=6,55	2	0,05	40	1,00

Interpretación

Una de las interpretaciones es: 12 personas fueron atendidas en un intervalo de 3.35 a 4 .15 minutos, las 12 personas representan el 30% de las personas atendidas, hasta los 4,15 minutos fueron atendidas 24 personas, las cuales representan el 60% del total de personas atendidas

El análisis de los datos se las ha realizado mediante datos individuales y datos agrupados, generalmente el análisis que se realiza mediante datos agrupados ya que el proceso se realiza con una gran cantidad de datos.

Media de datos agrupados:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} m_i f_i$$

Varianza de datos agrupados:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i (m_i - \bar{x})^2$$

n número de datos

k número de clase

$\boldsymbol{m_i}$ Marca de la clase i (es el punto medio de la clase)

$\boldsymbol{f_i}$ Frecuencia de la clase i

Ejercicio 18:

La tabla de frecuencias siguiente contiene los datos agrupados en 6 clases del número de artículos vendidos por un almacén en 50 días.

m_i	15	25	35	45	55	65
f_i	2	10	12	14	9	3

Encuentre:

a) **Media**

Desarrollo

$$\bar{x} = \frac{1}{50}(m_1 f_1 + m_2 f_2 + m_3 f_3 + m_4 f_4 + m_5 f_5 + m_6 f_6)$$

$$\bar{x} = \frac{1}{50}(15 * 2 + 25 * 10 + 35 * 12 + 45 * 14 + 55 * 9 + 65 * 3)$$

$$\bar{x} = 40{,}4$$

Interpretación: Indica que diariamente se ha vendido en promedio 40 artículos.

b) Varianza

Desarrollo

$$s^2 = \frac{1}{49}[2(15 - 40{,}4)^2 + 10(25 - 40{,}4)^2 + 12(35 - 40{,}4)^2 + 14(45 - 40{,}4)^2$$
$$+ 9(55 - 40{,}4)^2 + 3(65 - 40{,}4)^2]$$

$$s^2 = 164{,}12$$

$$s = 12{,}81$$

Interpretación: Indica que la diferencia de ventas entre días fue aproximadamente de 13 artículos.

Ejercicio 19:

Se dispone de los siguientes datos incompletos en una tabla de frecuencia. Se conoce que la media calculada con los datos agrupados es 19.7 y n= 40.

Numero	Clase	Marca	f	F	f_r	F_r
1				**2**		**0,05**
2						**0,25**
3	**[15,20)**		**14**			**0,6**
4						
5				**36**		
6						**0,975**
7						

Desarrollo

$f_{r3} = \frac{f_3}{n} = \frac{14}{40} = 0.35$	$F_2 = F_{r2}(40) = 0.25(40) = 10$

$F_3 = F_{r3}(40) = 0.6(40) = 24$	$F_2 = F_1 + f_2$ $10 = 2 + f_2$
$F_6 = 0{,}975(40) = 39$	$F_6 = F_5 + f_6$ $39 = 36 + f_6$
$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} mi, fi$	$19{,}7 = \frac{1}{40}(7{,}5 * 2 + 12{,}5 * 8 + 17{,}5 * 14 + 22{,}5f_4 + 27{,}5f_5 + 32{,}5 * 3 + 37{,}5 * 1)$ $788 = 495 + 22{,}5f_4 + 27{,}5f_5$ $293 = 22{,}5f_4 + 27{,}5f_5 \quad (\mathbf{1}).$
$f_4 + f_5 = 12 \quad (\mathbf{2})$	Resolviendo (1) y (2) $293 = 22{,}5(12 - f_5) + 27{,}5f_5$ $293 = 270 - 22{,}5f_5 + 27{,}5f_5$ $23 = 5f_5$ $f_5 = \mathbf{4{,}6} \approx \mathbf{5}$ $f_4 + 5 = 12$ $f_4 = \mathbf{7}$

Numero	Clase	Marca	f	F	f_r	F_r
1	[5,10)	7,5	2	**2**	0,05	**0,05**
2	[10,15)	12,5	8	10	0,2	**0,25**
3	**[15,20)**	17,5	**14**	24	0,35	**0,6**
4	[20,25)	22,5	7	31	0,175	0,775
5	[25,30)	27,5	5	**36**	0,125	0,9
6	[30,35)	32,5	3	39	0,075	**0,975**
7	[35,40)	35,5	1	40	0,025	1

EJERCICIOS PROPUESTOS

Tabla de distribuciones

1) **La tabla muestra la distribución del ingreso familiar mensual de 80 familias. Determine el número de familias que tiene un ingreso menor a 800 dólares mensuales.**

Clase	f	F	Fr
640-680			
680-720	**48**	**60**	
720-760			**0,125**
760-800			**0,075**
800-840			

Respuesta: 76

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios: Análisis Exploratorio de datos pag18.

2) Se dispone de los siguientes datos incompletos en una Tabla de Frecuencias, completar la Tabla:

Número	Clase	Marca	f	F	f/n	F/n
1	**[1,2)**		**1**			
2				**6**		
3					**0.25**	
4						**0.7**
5			**8**			**0.9**
6					**0.05**	
7						

3) En la en la tabla indicada los tiempos de espera en las ventanillas de un banco. Halle el tamaño de la muestra y complete la tabla de distribución de frecuencias.

Tiempo (min)	Frec. absoluta	Frec. Relativa
0-3	**32**	
3-6		**0.30**
6-9		
9-12	**8**	**0.05**
12-15		**0.10**

Respuesta n = 160

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios: Análisis Exploratorio de datos pag18.

4) Halle el tamaño de la muestra y reconstruya la siguiente tabla asimétrica de distribución de frecuencia.

Intervalo	Frec. Absoluta	Frec. Relativa	Frec. Relativa acumulada
10 – 12	**7**		
12 -			**0.24**
-		**0.52**	
-	**5**		
18-20			

Respuesta: n=50

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios: Análisis Exploratorio de datos pag18.

5) La tabla muestra la distribución del ingreso familiar mensual de 80 familias. Complete la tabla y determine el número de familias que tienen un ingreso menor a 800 dólares.

Intervalo	Frec. Absoluta	Frec. Absoluta acumulada	Frec. Relativa
640 – 680			
680 – 720	**48**	**60**	
720 – 760			**0.125**
760 – 800			**0.075**
800 – 840			

Respuesta: 76

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios: Análisis Exploratorio de datos

2.6 Medidas de forma

2.6.1 Asimetría

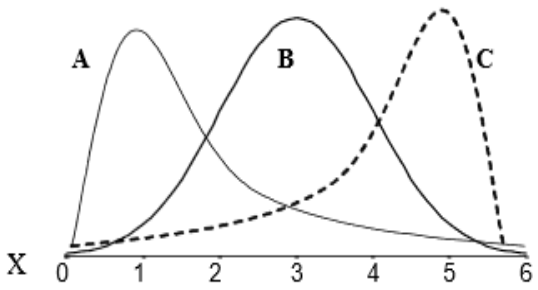
La asimetría de una distribución hace referencia al grado en que los datos se distribuyen en forma uniforme a partir de una medida de tendencia central a la derecha o a la izquierda, se define como:

$$As = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n}}{s^3}$$

Caso A: As > 0 Asimetría Positiva

Caso B: As = 0 Simetría

Caso C: As < 0 Asimetría Negativa



Gráfica 2 Asimetría.

Ejercicio 20:

Dado los siguientes datos calcular el coeficiente de asimetría

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
2	-2	4	-8
4	0	0	0
8	4	16	64
2	-2	4	-8
$\bar{x} = 4$	0	24	64

$$As = \frac{\frac{48}{4}}{\left(\left(\frac{24}{4}\right)^{1/2}\right)^3} = \frac{12}{14.69} = 0.82$$

Conclusión: El coeficiente de asimetría para los datos anteriores es igual a 0.82, lo que indica que hay una asimetría positiva es decir hay más datos a la izquierda de la media.

2.6.2 Curtosis

El coeficiente de apuntamiento o curtosis de una variable sirve para medir el grado concentración de los valores en torno a su media, se calcula de la siguiente manera:

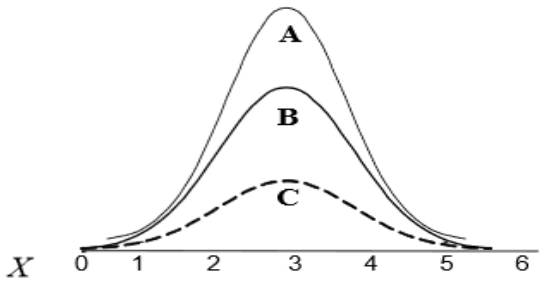
$$k = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}}{s^4} - 3$$

A la expresión se le resta 3 ya que valdría cero si la distribución fuera **Normal,** tomándose a ésta como referencia.

Caso A: k > 0, distribución Leptocúrtica

Caso B: k = 0, distribución Mesocúrtica (distribución Normal)

Caso C: K < 0, distribución Platicúrtica



Gráfica 3 Curtosis

Ejercicio 21:

Dados los siguientes datos 2, 4, 8, 2 calcular la curtosis.

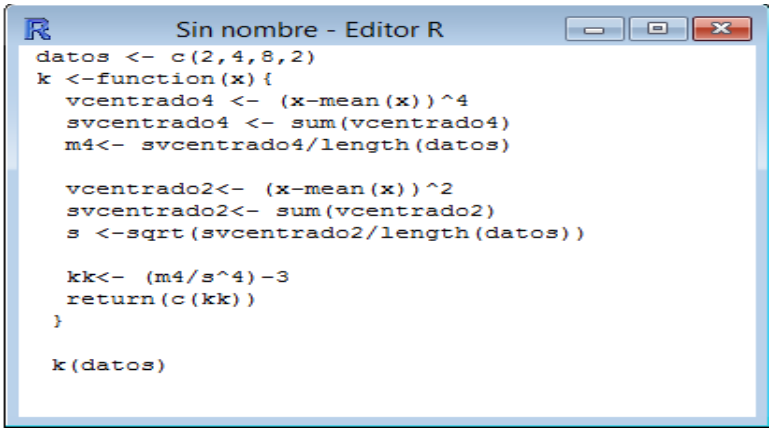
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$

2	-2	4	16
4	0	0	0
8	4	16	256
2	-2	4	16
TOTAL:	0	24	288

$$\bar{x} = 4$$

$$k = \frac{\frac{288}{4}}{\left(\left(\frac{24}{4}\right)^{1/2}\right)^4} - 3 = \frac{72}{36} - 3 = 2 - 3 = -1$$

La curtosis es igual a -1, su distribución es achatada alrededor de la media.



EJERCICIOS PROPUESTOS

Fuente: Libro Galindo página 43, Literales: (2,3) y libro Estadística Descriptiva de S. Fernández página 259)

1) Dadas η= 8 mediciones: 4, 2, 6, 5, 7, 5, 4, 6, encuentre:

a) $\bar{x}$;

b) La mediana;

c) s;

d) El rango;

e) La asimetría;

f) La curtosis ;

Respuestas: a) $\bar{x}$=4.875; b) Me=5; c) s=1.553; d) R=5; e) As=-0.644; f) Ap=0.592

2) Dadas η= 9 mediciones: 5, 8, 8, 4, 4, 9, 7, 5, 4. Determine:

- $\bar{x}$;
- La mediana;
- s;
- El rango;
- La asimetría;
- La curtosis ;

Respuestas: a) $\bar{x}$=6; (b) 5; (c) s=2; (d) R=5; e) As=0.362; (f) Ap=-1.826

3) La dirección general de tráfico está interesada en estudiar la educación vial en los jóvenes. Para ello selecciona una muestra aleatoria de sujetos que acaban de obtener el carnet de conducir (grupo 1) y otra con sujetos que lo tienen hace 5 años (grupo 2) y registra el nº de veces que han perdido puntos en el último año. Los resultados se muestran a continuación:

Grupo 1:	1	2	4	1
Grupo 2:	2	7	7	8

Respuesta: $\bar{x}$1=2; $\bar{x}$2=6; S1=1,5; S2=5,5

4) Calcular el valor Asimétrico y Curtosis de los siguientes valores: 2,4,8,2

Respuesta: $\bar{x}$=4; s=2,82, As=0,53; Ap=-1,86

5) Estudiar la simetría de los siguientes datos:

10-12-12-14-10-10-16-12-14-10

Respuestas: $\bar{x}$=12; s=2, As=0,53; As=0,6 Asimetría positiva.

2.8 Diagrama de caja

Son una presentación visual que describe varias características importantes al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente.

CONSTRUCCIÓN:

Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico $Q_3 - Q_1$. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero, recordemos que el segundo cuartil coincide con la mediana.
Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato que no se encuentre dentro de este rango es marcado e identificado individualmente.

Ejercicio 22:

Dadas las siguientes edades de 20 personas, realizar el diagrama de caja y explicar.

36,25,37,24,39,20,36,45,31,31,39,24,29,23,41,40,33,24,34,40

Desarrollo:

Para calcular los parámetros estadístico, lo primero es ordenar la distribución

20 23 24 24 24 25 29 31 31 33 34 36 36 37 39 39 40 40 41 45

Cálculos de cuartiles

Para calcular Q1, a los 20 datos lo dividimos para cuatro teniendo cada grupo 5 elementos, el primer cuartil es la media aritmética del valor que está en la quinta posición y el siguiente:

Q1 = (24 + 25) / 2 = 24,5

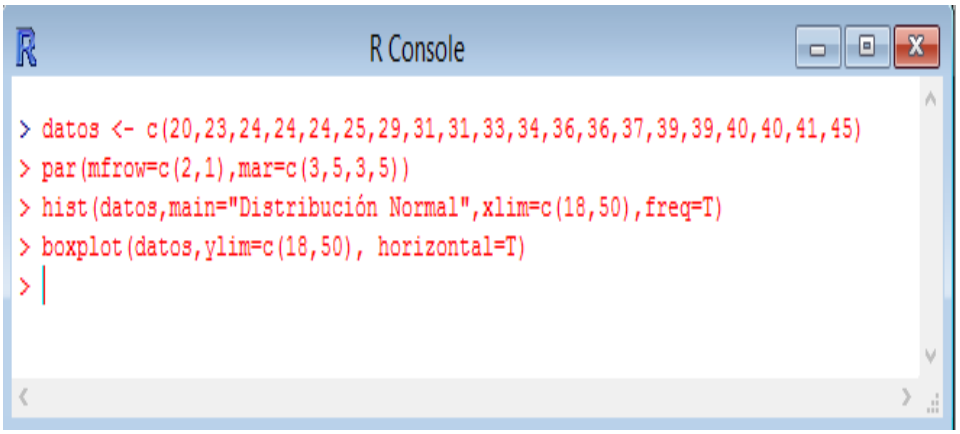
Para calcular Q2, a los 20 datos lo dividimos para dos grupos obteniendo para cada grupo 10 elementos, la mediana es la media aritmética del valor que está en la posición décima y el siguiente:

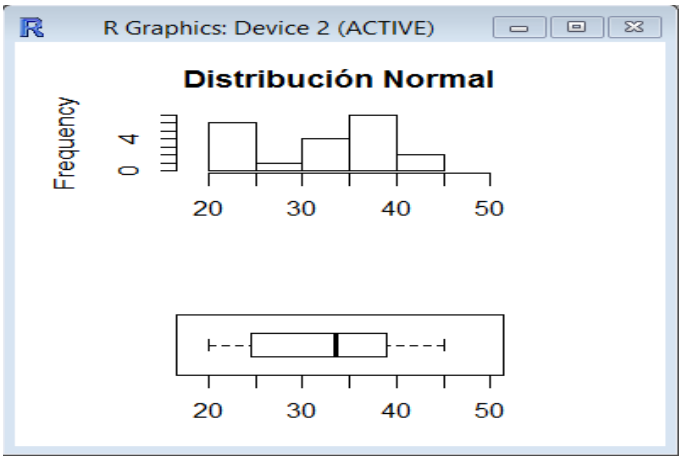
Md= Q2 = (33 + 34)/ 2 =33,5

Para calcular el Q3, a los 20 datos lo dividimos para el 75% de los valores, obteniéndose un grupo de 15 elementos, el tercer cuartil es la media aritmética del valor que está es la posición quince y el siguiente:

Q3= (39 + 39) / 2 = 39

DIBUJAR LA CAJA Y LOS BIGOTES





Gráfica 4 Diagrama de cuartiles

Interpretación del diagrama

La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que las edades comprendidas entre el 25% y el 50% de la población están más dispersa que entre el 50% y el 75%.

El bigote de la izquierda (Xmím, Q1) es más corto que el de la derecha; por ello el 25% de los más jóvenes, sus edades son más parecidas, que el 25% de los mayores.

El rango intercuartílico = Q3 - Q1 = 14,5; el 50% de la población centrada alrededor de la mediana y ordenada, la diferencia entre sus edades en promedio es de 14,5 años.

2.9 Coeficiente de variación

Es una de la variabilidad que indica la magnitud relativa de la desviación estándar en comparación con la media.

Es útil para contrastar la variación de dos o más variables que están medidas en diversas escalas, su cálculo puede realizarse mediante la siguiente expresión:

$$CV = \frac{S}{\bar{x}} * 100$$

Ejercicio 23:

Con un micrómetro, se realizan mediciones del diámetro de un disco, que tienen una media de 4.03 mm y una desviación estándar de 0.012 mm; con otro micrómetro se toman mediciones de la longitud de un tornillo que tiene una media de 1.76 pulgadas y una desviación estándar de 0.0075 pulgadas. ¿ Cuál de los dos micrómetros presenta una variabilidad relativamente menor?. Los coeficientes de variación son:

$$CV_{disco} = \frac{0.012}{4.03} * 100 = 0.29\ \%\ \ y\ \ CV_{tornillo} = \frac{0.0075}{1.76} * 100 = 0.42\ \%$$

Analizando, en la primera medición se exhibe una variabilidad menor con respecto a su media que las efectuadas por la otra medición.

Ejercicio 24:

Dos profesores que imparten diferentes materias a un mismo grupo deciden investigar el coeficiente de variación de una y otra materia, para lo cual se obtiene la media y la desviación estándar respectivamente:

Para la primera materia se tiene: $\bar{x} = 6.3$; $s_1 = 1.2$; $CV = \frac{1.2}{6.3} * 100 = 0.19\ \%$

Para la segunda materia se tiene: $\bar{x} = 8$; $s_2 = 3$; $CV = \frac{3}{8} * 100 = 0.38\ \%$

Se concluye que aunque las calificaciones en promedio son igual a 8 las calificaciones son más dispersas ya que el coeficiente de variación es mayor para la segunda muestra.

TERCER CAPÍTULO

Probabilidad

CONTENIDO

3.1 Fundamentos de la teoría de probabilidad

- 3.1.1 Permutaciones
- 3.1.2 Combinaciones

3.2 Espacio muestral (Ω)

- 3.2.1 Conjunto Partes de Ω (subconjuntos)

3.3 Eventos

3.4 Probabilidad de eventos

- 3.3.4 Propiedades de los eventos simples

3.5 Axiomas de la probabilidad

3.6 Propiedades de la probabilidad

- 3.6.1 Probabilidad de un evento nulo
- 3.6.2 Probabilidad del evento complemento
- 3.6.3 Probabilidad de eventos incluidos
- 3.6.4 La probabilidad de un evento está entre 0 y 1
- 3.6.5 Probabilidad de la diferencia de eventos
- 3.6.6 Regla aditiva de probabilidad de Eventos

3.7 Independencia de eventos

3.8 Condicionalidad de eventos

CAPÍTULO 3
PROBABILIDAD

INTRODUCCIÓN

El concepto de probabilidad es abstracto. Por ende tratamos de dar una explicación fácil y coherente del mismo, mediante la siguiente simulación.

Conocemos a priori que la probabilidad de obtener una cara en el lanzamiento de una moneda es 0.5, por lo tanto tenemos una inquietud ¿Cómo obtener 0.5?

Realizamos el siguiente experimento "Lanzamiento de n veces una moneda y observamos la f_r de obtener cara y sello", como se describe en los siguientes gráficos.

Experimento: "Lanzamiento de n veces de una moneda"			
Experimento 1 : n= 10			
Condiciones del experimento:		0 cara	
		1 sello	
1			
0		frecuencia relativa de cara=	0,5
0		frecuencia relativa sello=	0,5
0			
1			
0			
1			
1			
1			
0			

Gráfica 5 Probabilidad de n=1

Experimento: "Lanzamiento de n veces de una moneda"									
Experimento 1 : n= 100									
Condiciones del experimento:			0 cara						
			1 sello			frecuencia de cara reltiva=			0,48
						frecuencia de sello reltiva=			0,52
1	0	1	0	0	1	1	0	0	1
0	0	1	1	0	0	0	1	1	0
1	0	0	1	1	0	0	1	1	1
0	1	1	0	1	1	1	0	0	1
0	1	1	1	0	1	1	0	1	1
1	0	1	1	1	0	0	0	0	0
0	1	1	1	1	0	1	0	1	0
1	1	1	0	0	0	0	0	0	1
0	1	0	0	1	1	1	0	0	0
0	0	1	1	0	0	0	0	0	0

Gráfica 6 Probabilidad de n=100

Podemos concluir, que cuando el número de lanzamientos tiende al infinito:

"La frecuencia relativa de obtención de cara se estabiliza en un valor, llamado probabilidad de obtener cara en el lanzamiento de una moneda".

$$f_r(cara) \rightarrow lanzamientos\ tiende \infty \rightarrow P(cara)$$

Probabilidad de cara

3.1 Fundamentos de la teoría de probabilidad

3.1.1 Principio básico de conteo o principio de multiplicación

El principio indica, si un proceso se lo puede dividir en n subprocesos y cada subproceso se puede realizar de r_1,…, r_n maneras, entonces todo el proceso se lo puede realizar de r_1*r_2*…*r_n maneras.

Ejercicio 25:

Se lanza un dado y una moneda. ¿Cuántos resultados diferentes se obtiene de este experimento?

Desarrollo

Al lanzar el dado se puede tener m=6 resultados diferentes; mientras que al lanzar la moneda se obtiene n=2 resultados diferentes. Por lo tanto, el número total de resultados del experimento es **m x n = 6 x 2 = 12.**

Los doce resultados son:

Moneda ={ cara (c), sello(s)}

Dado = {1, 2, 3, 4, 5, 6}

{(1,c),(1,s), (2,c),(2,s), (3,c),(3,s), (4,c),(4,s), (5,c),(5,s), (6,c),(6,s)}

3.1.2 Permutaciones

Son los arreglos (grupos) diferentes que se pueden hacer con los elementos de un grupo.

En estos arreglos se debe **considerar el orden** de los elementos incluidos, su cálculo es:

$$nP_K = \frac{n!}{(n-k)!}$$

Ejercicio 26:

Dado el siguiente conjunto de letras= {a, b, c} encuentre las permutaciones de tres objetos tomados dos de ellos, es decir $3P_2$

Desarrollo

$3P_2 = \frac{3!}{(3-2)!} = 6$

Los grupos son {ab, ac, ba, bc, ca, cb}, cabe indicar que "ab" es distinto de "ba" ya que el orden sí importa.

Ejercicio 27:

Se tiene un grupo de personas entre ellas Pedro, Betty y Juan sobresalen se quiere formar una directiva de dos personas presidente y tesorero.

¿Se quiere conocer cuántas directivas se puede formar con las tres personas?.

$$3P_2 = \frac{3!}{(3-2)!} = 6$$

Las posibles directivas serían:

A = { {Pedro, Betty},{Pedro, Juan},{Juan, Betty},{Betty, Juan},{Juan, Pedro},{Betty, Pedro}}

TIPS DE R: PERMUTACIONES

Para este tema necesitaremos dos paquetes, que pueden ser instalados al digitar las siguientes expresiones indicadas abajo:

```
install.packages('gtools')

library(gtools)
```

La función "**permutations(n,r,v=x,repeats.allowed=F)"**, se utiliza para obtener los arreglos donde el orden si importa.

"n" es el número de elementos que tiene el conjunto.

"r" es el número de elementos que tiene los arreglos.

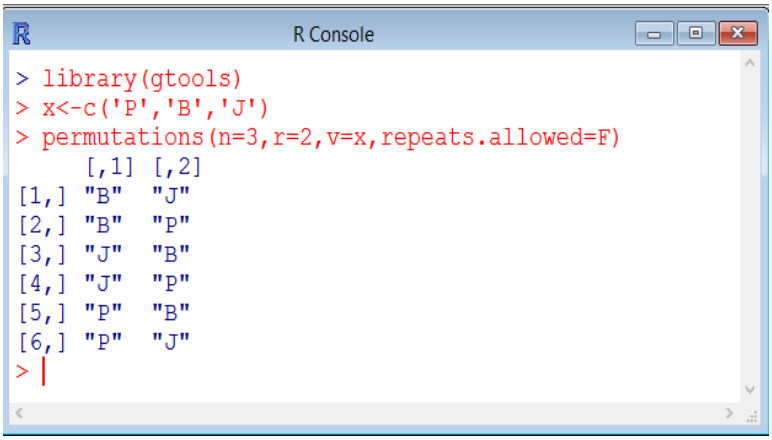


Ilustración 5 Tips R 17. Permutación

3.1.3 Combinaciones

Son los arreglos (subconjuntos) que se puede tener con los n elementos de un conjunto considerando que el orden de los elementos en cada arreglo no es de interés.

$$nC_K = \frac{n!}{(n-k)!\,k!}$$

Ejercicio 28:

Dado el conjunto de las vocales encontrar las combinaciones o subconjuntos de tamaño 2.

Desarrollo

$$5C_2 = \frac{5!}{(5-2)!\,2!}$$

$$= \frac{5!}{3!\,2!} = \frac{20}{2} = 10$$

Los elementos son {ae, ai, ao, au, ei, eo,eu, io, iu, o

Para calcular las combinaciones se utiliza la función "**combinations(n, r, v**)", los cuales son arreglos donde el orden no importa.

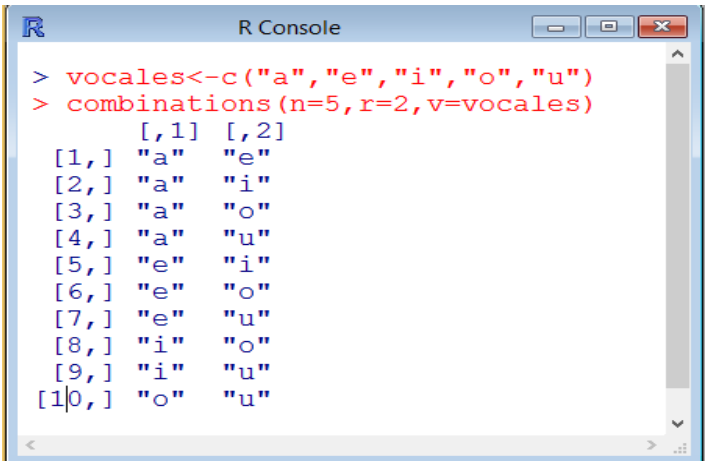


Ilustración 6 Tips R 18 Combinación.

n, es el número de elementos.

r, es el número de elementos que tiene el arreglo.

v, es el conjunto de datos.

3.2 Espacio muestral (Ω)

Es el conjunto de todos los posibles resultados de un experimento (eventos elementales).

Ejercicio 29:

a) Experimento: "Lanzamiento de una moneda"

Desarrollo

Ω = {cara, sello}

b) Experimento: "Lanzamiento dos monedas"

Desarrollo

El número de casos se calcula mediante la regla de la multiplicación, dos casos de la primera moneda (C_1, S_1), dos casos de la segunda moneda (C_2, S_2), por lo tanto hay cuatro casos posibles.

$\Omega = \{(C_1, C_2),(C_1, S_2), (S_1, C_2), (S_1, S_2)\}$

3.2.1 Conjunto Partes de Ω ($\wp(\Omega)$)

El conjunto partes de Ω, se refiere al número de subconjuntos que tiene dicho conjunto; su cardinalidad (número de elementos de un conjunto) es igual a 2^n, donde n es el número de elementos del conjunto.

Ejercicio 30:

Experimento: "Lanzamiento de una moneda", encuentre el conjunto partes de Ω.

Ω ={cara, sello}

$\text{Card}\ (\Omega) = 2^2 = 4$

= {{cara}, {sello}, {cara, sello}, {Ø}}

Ejercicio 31:

Experimento: "Lanzamiento de una moneda"

Desarrollo

Ω = {cara, sello}, este espacio muestral tiene la característica de ser discreto y numerable.

Ejercicio 32:

Experimento: "Sueldo de una empresa"

Desarrollo

$\Omega = \{ x/x \in [a,b], x\ es\ el\ sueldo \}$, tiene la característica de ser continuo y no numerable.

3.3 Eventos

Los eventos son subconjuntos de Ω, siendo estos de dos tipos, simples y compuestos.

3.3.1 Evento Simple*:*

Es aquel que tiene un solo elemento básico o fundamental.

Ejercicio 33:

Dado el siguiente experimento "Lanzamiento de un dado", encuentre un evento simple.

$$\Omega = \{1,2,3,4,5,6\}$$
$$A = \{"sale\ 1"\} = \{1\}$$

3.3.2 Evento Compuesto:

Es aquel que tiene más de un elemento elemental o básico.

Ejercicio 34:

Dado el siguiente experimento "Lanzamiento de un dado", encuentre un evento compuesto-

$$\Omega = \{1,2,3,4,5,6\}$$
$$A = \{"sale\ par"\} = \{2,4,6\}$$

3.4 Probabilidad de eventos

Matemáticamente la probabilidad es una función definida como conjunto de salida el conjunto partes de omega y conjunto de llegada el intervalo [0,1].

$$P: \wp(\Omega) \longrightarrow [0,1]$$
$$A \longrightarrow P(A)$$

La primera perspectiva de la probabilidad la desarrollaremos desde el punto de vista clásico, definida de la siguiente manera:

$$\textbf{Probabilidad de A} = \mathbf{P(A)} = \frac{\#\ de\ casos\ favorables\ al\ evento\ A}{\#\ casos\ totales} = \frac{card\ (A)}{card\ (\Omega)}$$

Ejercicio 35:

Dado el experimento "lanzamiento de un dado", calcular la probabilidad de que al lanzar un dado, salga un número mayor a 3.

A= "salga un número mayor a tres"

A= {4, 5, 6}

$$\boldsymbol{P(A)} = \frac{3}{6} = \frac{1}{2}$$

Conclusión: De cada 100 veces que se realiza el experimento en 50 de ellos se obtendrá un número mayor a tres.

Ejercicio 36:

Una caja con 12 baterías para carro contiene 5 defectuosas. Halle la probabilidad de que al elegir al azar tres baterías de la caja se obtengan:

a) Una defectuosa

Desarrollo

$$\text{P ("Escoger al azar una batería defectuosa")} = \frac{5C_1\ 7C_2}{12C_3} = \frac{105}{220} = \frac{21}{44} = 0{,}48$$

En el 48% de las veces se va obtener 1 batería defectuosa.

b) Ninguna defectuosa

Desarrollo

$$\text{P ("Escoger que no haya ninguna defectuosa")} = \frac{5C_0 * 7C_3}{12C_3} = \frac{35}{220} = \frac{7}{44}$$

$$= 0{,}16$$

En el 16% de las veces se va obtener 3 baterías buenas.

Ejercicio 37:

Un grupo de 15 personas, 7 leen la revista A, 5 leen la revista B y 6 ninguna revista.

	Leen B	No leen B	
Leen A	3	4	7
No leen A	2	6	8
Total	5	10	15

Del cuadro se obtiene directamente que:

4 leen A, únicamente (lea A pero no lea B)

2 leen B, únicamente (lea B pero no lea A)

3 leen A y B

a) Encuentre la probabilidad de que al escoger una persona al azar esta al menos haya leído una revista.

Desarrollo

1) Primera forma

P("al menos haya leído una revista")=P(haya leído la revista A o haya leído la revista B o ambos)= P("haya leído la revista A")+P("haya leído la revista B")+P("ambas") $= \frac{4}{15} + \frac{2}{15} + \frac{3}{15} = \frac{9}{15} = 0{,}6$

2) Segunda forma

$$P\ (\text{“ esta al menos haya leido una revista”}) = \frac{9}{15} = \frac{3}{5} = 0.6$$

Interpretación

Indica de cada 100 veces que se realiza el evento en 60 veces va a salir al menos una persona que lea una revista.

b) Encuentre la probabilidad de que al escoger tres personas al azar, dos lean ambas revistas y uno no lea.

Desarrollo

P ("elegir 3 personas, dos lean ambas revista y uno no lea") $= \frac{{}_3C_2 * {}_6C_1}{{}_{15}C_3} = \frac{3 * 6}{455}$

$= 0{,}039.$

3.4.1 Propiedades de los eventos simples

Sea $\Omega = \{E_1, E_2, \dots\dots, E_n\}$ eventos simples, entonces la probabilidad de la unión de esos eventos es igual a la suma de sus probabilidades, es decir:

$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i);$

P ($E_1 \cup E_2 \cup \dots\dots E_n$) = P ($E_1$) + P ($E_2$) +… … P ($E_n$)

se debe a que los eventos simples son disjuntos, es decir para todo i,j se tiene $E_i \cap E_j = \emptyset$.

Observación:

s $= \{E_1, E_2, \dots\dots, E_n\}$

$$P\left(\bigcup_{i=1}^{n} E_1\right) = \sum_{i=1}^{n} P(E_1) = 1$$

Ejercicio 38:

a) **Cuál es la posibilidad que al lanzar un dado se obtenga un número par**

Desarrollo

P ("al lanzar un dado, se obtenga un número par")= P ("sale 2 o sale 4 o sale 6")=

$= P(\frac{"Sale\ 2"}{E_1} o \frac{"Sale\ 4"}{E_2} o \frac{"Sale\ 6"}{E_3})$

=P ($E_1 \cup E_2 \cup E_3$) = P (E_1) + $P(E_2) + P(E_3)$ = $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \mathbf{\frac{1}{2}}$

b) **P ("Salga un número en el lanzamiento de un dado")**

Desarrollo

P ("sale 1" o "sale 2" o "sale 3" o… o "sale 6") = P ("sale1")+P ("sale 2")+…+P ("sale 6") = $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=1$

Nota: La suma de los eventos simples siempre es 1.

3.4.2 Axiomas de la probabilidad

$P(\mathrm{A}) \geq 0$

$P(\Omega) = 1$

$A_1, A_2 \in S\ (partes\ \Omega)\ ;\ A_1 \cap A_2 = \emptyset \implies P(A_1 \cup A_2) = P(A_1) + P(A_2)$

3.4.3 Propiedades de la probabilidad

a) Probabilidad de un evento vacío.

$$\boldsymbol{P(\emptyset) = 0}$$

Demostración:

$\Omega = \Omega \cup \emptyset$	Identidad
P(Ω) = P(Ω ∪ ∅); $\Omega \cap \emptyset = \emptyset$	Complemento
$P(\Omega) = P(\Omega) + P(\emptyset);\ \Omega = \{E_1, E_2, \dots, E_n\}$	Probabilidad

$1 = 1 + P(\emptyset)$

$$P(\emptyset) = 0.$$

b) Probabilidad del evento complemento

$$\boldsymbol{P(A^c) = 1 - P(A)}$$

Demostración:

$\Omega = A \cup A^c$	Complemento
$P(\Omega) = P(A \cup A^c);\ A \cup A^c = \emptyset$	Probabilidad
$P(\Omega) = P(A) + P(A^c)$	Conjuntos Disjuntos
$1 = P(A) + P(A^c)$	Axioma

$P(A^c) = 1 - P(A)$

c) Probabilidad de eventos incluidos

$$\boldsymbol{Si\ A\ \subset B, entonces\ P(A) \leq P(B)}$$

Demostración:

$B = A \cup (A^c \cap B)$ Definición de A subconjunto B

$P(B) = P(A) + P(A^c \cap B); A \cap (A^c \cap B) = 0$ Conjuntos Disjuntos

$P(B) \geq P(A)$ El todo es mayor igual que sus par

d) La probabilidad de un evento está entre 0 y 1

$$\boldsymbol{0 \leq P(A) \leq 1}$$

Demostración:

$\emptyset \subset A \subset \Omega$

$P(\emptyset) \leq P(A) \leq P(\Omega)$

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

e) Regla aditiva de probabilidad de Eventos

$$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Demostración:

$(A \cup B) = (A - B) \cup (A \cap B) \cup (B - A)$ Definición de conjuntos

$P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A)$ Conjuntos disjuntos

$P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$ Probabilidad de Diferencia

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

Ejercicio 39:

Dado un mazo de barajas, calcular:

Desarrollo

El número de cartas es 52, y los tipos de cartas son:

♦ Diamantes

♣ Trébol

♥ Corazón Rojo

♠ Espada

a) Encuentre la P (“Sacar un As o sacar un Rey”)

Sean los eventos A = “Sacar un As” y B = “Sacar un Rey”.

P (“Sacar un As o sacar un Rey”) =

P (A o B) = P (A ∪ B) = P (A) + P (B) – P (A∩B)

$$= \frac{4}{52} + \frac{4}{52} - 0$$

$$= \frac{8}{52} = \frac{2}{13} = 0{,}15$$

b) Encuentre la P (“Sacar un As” o “Sacar una espada”)

Sea el evento A = “sacar un As” y sea B = “sacar una espada”.

P (“Sacar un As” o “Sacar una espada”)

P (A ∪ B) = P (A) + P (B) – P (A ∩ B)

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$$

Ejercicio 40:

Entre 5 Matemáticos y 7 Físicos hay que hacer una comisión de 2 Matemáticos y 3 Físicos. ¿De cuantas formas podrá hacerse? si:

1.- Todos son elegibles

2.- Un físico en particular a de estar en esa comisión

3.- Dos matemáticos concretos tienen prohibido pertenecer a la comisión

Desarrollo

Número de personas = 12

Matemáticos = 5

Físicos = 7

Comisión = 2 Matemáticos y 3 Físicos

1.- Todos son elegibles = ${}_5C_2 \cdot {}_7C_3$

$$= \frac{5!}{3!\,2!} \cdot \frac{7!}{4!\,3!} = 350 \text{ maneras de combinar}$$

2.- Un físico en particular a de estar en esa comisión = ${}_5C_2 \cdot {}_6C_2$

$$= \frac{5!}{3!\,2!} \cdot \frac{6!}{4!\,3!} = 150 \text{ maneras de combinar}$$

3.- Dos matemáticos tienen prohibido pertenecer a la comisión = ${}_3C_2 \ {}_7C_3$

$$= \frac{3!}{1!\,2!} \cdot \frac{7!}{4!\,3!} = 105 \text{ maneras de combinar}$$

Ejercicio 41:

Se tiene 3 rosas rojas, 5 claveles y 2 tulipanes ¿Cuál es la probabilidad de escoger 3 veces consecutivamente a los claveles, con reposición y sin reposición?

a) P ("Escoger 3 veces consecutivamente claveles con reposición")

$$\frac{5}{10} * \frac{5}{10} * \frac{5}{10} = \frac{125}{1000}$$

b) P ("Escoger 3 veces consecutivamente claveles sin reposición")

$$\frac{5}{10} * \frac{4}{9} * \frac{3}{8} = \frac{60}{720}$$

3.5 Independencia de eventos

Se dice que dos eventos A y B son independientes si el uno no influye en la consecución del otro y viceversa, es decir:

$$\boldsymbol{P(A\ y\ B) = P(A \cap B) = P(AB) = P(A)P(B)}$$

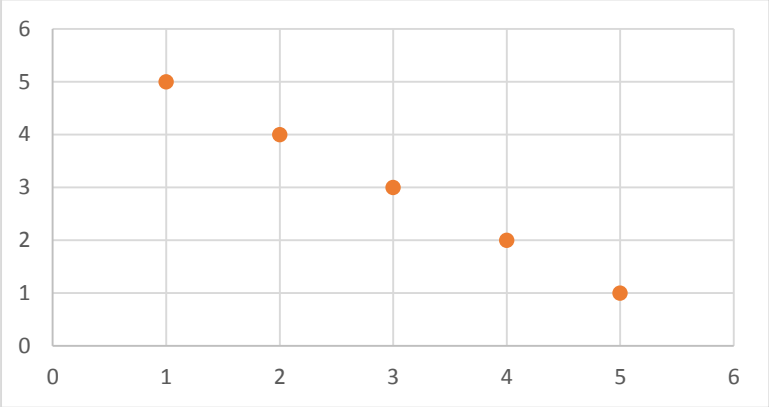
3.6 Condicionalidad de eventos

Sean dos eventos A y B y $P(B) \neq 0$ se dice que el evento A está condicionada a la consecución del evento B si:

$$\boldsymbol{P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}}$$

Ejercicio 42:

Se lanzan 2 dados, ¿cuál es la probabilidad de que en los dos dados salga el tres, si se sabe que la suma es 6?.



Los siguientes eventos cuya suma es seis son: A = { {1,5},{2,4},{3,3},{4,2},{5,1}} y de los eventos solo uno hay uno el {3,3} que sale en los dos dados el 3, por lo tanto si B = "Salga tres en ambos dados, si se sabe que su suma es 6"

$$P(B) = \frac{1}{5}$$

Ejercicio 43:

La probabilidad del que el evento M ocurra es igual a $\frac{1}{4}$, la probabilidad de que el evento N ocurra es $\frac{3}{5}$, y la probabilidad de que los complementos de M y N ocurran es de $\frac{3}{10}$ ¿Son independientes M y N?

Desarrollo

$$P(M) = \frac{1}{4}$$

$$P(N) = \frac{3}{5}$$

$$P(M^c \cap N^c) = \frac{3}{10}$$

$$P(M^c \cap N^c) = P((M \cup N)^c) = 1 - P(M \cup N)$$

$$\frac{3}{10} = 1 - P(M) - P(N) + P(M \cap N)$$

$$\frac{3}{10} = 1 - \frac{1}{4} - \frac{3}{5} + P(M \cap N)$$

$$\frac{3}{10} = \frac{3}{20} + P(M \cap N)$$

$$\frac{3}{20} = P(M \cap N)$$

Para demostrar que M y N son independientes se deberá demostrar que:

$$P(M \cap N) = P(M)P(N)$$

$$P(M \cap N) = \frac{3}{20} \quad (1)$$

$$P(M).P(N) = \frac{1}{4} * \frac{3}{5} = \frac{3}{20} \quad (2)$$

(1) = (2)

Se concluye que M y N son independientes.

Ejemplo 44:

Se conoce que el 60% de las mujeres y el 40% de los hombres reaccionan positivamente ante cierto medicamento. Se les proporcionó el medicamento a 20 personas de las cuales 12 eran mujeres, se tomó al azar uno de los resultados y se observó que era de reacción negativa. Halle la probabilidad de que el resultado corresponda a un hombre.

	R^-	R^+	
H	4,8	**3,2**	8
M	4,8	**7,2**	**12**
Total	9,6	10,4	**20**

Se conoce que el 60% de las mujeres, reaccionan positivamente es decir:

0,6*12 = 7,2.

Se conoce que el 40% de los hombres, reaccionan positivamente es decir:

0,4*8 = 3,2.

Antes de dar solución al ejercicio, se realizará un análisis de las probabilidades de la tabla.

Se requiere conocer los eventos, dadas las probabilidades:

$$P(A_1) = \frac{8}{20}$$

$\boldsymbol{A_1}$ = "al escoger una persona al azar esta sea hombre"

$$P(A_2) = \frac{12}{20}$$

$\boldsymbol{A_2}$ = "al escoger una persona al azar esta sea mujer"

$$P(A_3) = \frac{9{,}6}{20}$$

$\boldsymbol{A_3}$ = "al elegir una persona al azar esta reaccione negativamente"

$$P(A_4) = \frac{10{,}4}{20}$$

$\boldsymbol{A_4}$ = "al elegir una persona al zar esta reaccione positivamente"

$$P(A_5) = \frac{4{,}8}{20}$$

$\boldsymbol{A_5}$ = "al elegir una persona al azar este sea hombre y reaccione negativamente"

$$P(A_6) = \frac{7{,}2}{20}$$

$\boldsymbol{A_6}$ = "al elegir una persona al azar sea mujer y reaccione positivamente"

$$P(A_7) = \frac{4{,}8}{9{,}6}$$

$\boldsymbol{A_7}$ = "al elegir una persona al azar sea hombre y reaccione negativamente"

$$P(A_8) = \frac{3{,}2}{10{,}4}$$

$\boldsymbol{A_8} = \boldsymbol{H} \,/\, \boldsymbol{R^+}$ = "conociendo que es de reacción positiva, este sea hombre".

$$P(A_9) = \frac{4{,}8}{12}$$

$\boldsymbol{A_9} = \boldsymbol{R^-} \,/\, \boldsymbol{M}$ = "conociendo que es mujer, esta sea de reacción negativa".

La probabilidad de que al elegir una persona al azar sea hombre conociendo que reaccionó negativamente se puede calcular de dos formas:

Forma 1:

$$P(A) = P(H/R^{-}) = \frac{4{,}8}{9{,}6} = 0{,}5$$

Forma 2:

$$P(A) = P(H/R^{-}) = \frac{P(H \cap R^{-})}{P(R^{-})} = \frac{\frac{4{,}8}{20}}{\frac{9{,}6}{20}} = \frac{4{,}8}{9{,}6} = 0{,}5$$

Ejercicio 45:

En un taller trabajan 7 hombres y 3 mujeres, se escogen al azar 3 personas. Halla la probabilidad de que todas las personas seleccionadas sean hombres.

Desarrollo

P ("3 personas seleccionadas sean hombre") = P("la primera seleccionada sea hombre y la segunda sea hombre y la tercera sea hombre") = P ("1ra persona seleccionada sea hombre") P ("2da persona seleccionada sea hombre", dado que la "1ra persona seleccionada fue hombre") P ("3ra persona seleccionada sea hombre, dado que las anteriores fueron hombres")

$$= \frac{7}{10}\,\frac{6}{9}\,\frac{5}{8} = \frac{7}{24} \qquad \text{sin } reposición$$

$$= \frac{7}{10}\,\frac{7}{10}\,\frac{7}{10} = \frac{343}{1000} \qquad \text{con } reposición$$

Ejercicio 46:

Una caja contiene 8 bolas rojas 3 blancas y 9 azules, si se sacan 3 al azar. Determine la probabilidad de que

a) Las 3 sean rojas

Tenemos en la caja un total de:

8 bolas r. + 3 bolas b. + 9 bolas a. = 20 bolas

P("las tres salgan rojas")= $\frac{8}{20}\ \frac{7}{19}\ \frac{6}{18} = \frac{14}{285} = 0{,}049$ (sin reposición)

b) Las 3 sean blancas

P("Las tres salgan blancas")= $\frac{3}{20} \times \frac{2}{19} \times \frac{1}{18} = \frac{1}{18} = 0{,}0008771$(sin reposición)

EJERCICIOS PROPUESTOS

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios 2.9, literales (8, 9, 10, 11, 12) pág. 75.

1) Se lanzan dos dados ¿Cuál es la probabilidad de que en los dos dados salga el 3, si se sabe que la suma es 6?

R: 1/5

2) En una biblioteca hay 8 libros de literatura de ciencia ficción, 3 de los cuales son de Isaac Asimov, la biblioteca toma al azar 2 libros. Determine la probabilidad de que ambos libros puedan ser de Isaac Asimov.

R: 3/28

3) La empresa de correos ha determinado que el 10% de los paquetes enviados al exterior no llegan a su destino. Dos libros se pueden enviar separadamente o en un solo paquete. Para cada una de las formas , encuentre:

a) La probabilidad de que ambos libros lleguen a su destino
b) La probabilidad de que al menos un libro llegue a su destino

R: Envío conjunto a) 0,9 b) 0,9; Envío por separado a) 0,81 b) 0,99

4) Suponga que el 5% de todos los hombres y el 0.25% de todas las mujeres sufren daltonismo. Una persona escogida al azar resulta ser daltónica ¿Cuál es la probabilidad de que esta persona sea un hombre? (Considere que la cantidad de hombres y mujeres es igual)

R: 20/21

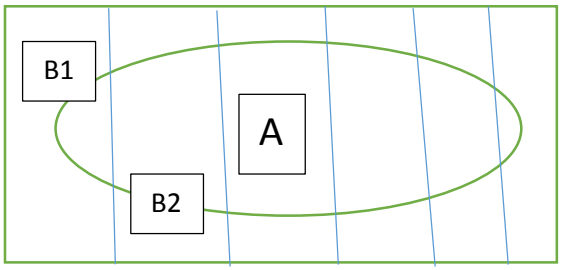
5) El 35% de los créditos de un banco son para vivienda, el 50% son para industria y el 15% para consumo diverso. Resultan fallidos el 20% de los créditos para vivienda, el 15% de los créditos para industrias y el 70% de los créditos para consumo. Calcula la probabilidad de que se pague un crédito elegido al azar.

3.7 Probabilidad total

Sea A un suceso cualquiera y $B_1, B_2 \dots, B_n$ eventos disjuntos, $B_i \cap B_j = \emptyset$, entonces

$$P(A) = \sum_{i=1}^{n} P(B_i)\, P(A \backslash B_i)$$

Demostración:



Gráfica 7 Demostración Probabilidad total

.

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)$$

$$P(A) = P[(A \cap B_i) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)]$$

Conocemos que los conjunto son disjuntos ya que $(A \cap B_i) \cap (A \cap B_j) = \emptyset$

$$P(A) = P(A \cap B_i) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

Aplicando la definición de probabilidad condicional $P(A \backslash B_i) = \frac{P(A \cap B_i)}{P(B_i)}$

$$P(A) = P(B_1)P(A \backslash B_1) + P(B_2)P(A \backslash B_2) + \cdots + P(B_n)P(A \backslash B_n)$$

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A \backslash B_i) \qquad (1)$$

La expresión (1) se la llama probabilidad total.

3.8 Teorema de Bayes

$$P(B_k \backslash A) = \frac{P(A \cap B_k)}{P(A)}$$

$$P(B_k \backslash A) = \frac{P(B_k)P(A \backslash B_k)}{P(A)}$$

$$P(B_k \backslash A) = \frac{P(B_k)P(A \backslash B_k)}{\sum_{k=1}^{n} P(B_k)P(A \backslash B_k)}$$

Ejercicio 47:

Demuestre $P\big((B_1 \cup B_2) \backslash A\big) = P(B_1 \backslash A) + P(B_2 \backslash A)$

Por definición $P\big((B_1 \cup B_2) \backslash A\big) = \frac{P\big((B_1 \cup B_2) \cap A\big)}{P(A)} = \frac{P((B_1 \cap A) \cup (B_2 \cap A))}{P(A)}$ (1)

Pero $(B_1 \cap A) \cap (B_2 \cap A) = B_1 \cap A \cap B_2 \cap A = (B_1 \cap B_2) \cap A = \emptyset$

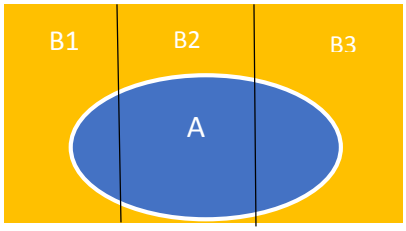
Ya que $B_1 \cap B_2 = \emptyset$, son disjuntos

En (1)

$$P\big((B_1 \cup B_2) \backslash A\big) = \frac{P(B_1 \cap A)}{P(A)} + \frac{P(B_2 \cap A)}{P(A)} = P(B_1 \backslash A) + P(B_2 \backslash A)$$

Ejercicio 48:

En una ciudad, el 25 % de los habitantes son ancianos el 35 % adulto y el 40% son niños. Se sabe que la gripe afecta al 5 % de los ancianos al 4% de los adultos y al 2% de los niños.



A= "se escoja al azar un habitante este tenga gripe".

$\boldsymbol{B}_1$ = "se escoja al azar un habitante este sea anciano".

$\boldsymbol{B}_2$ ="se escoja al azar un habitante este sea adulto".

$\boldsymbol{B}_3$ ="se escoja al azar un habitante este sea niño".

a) Calcular la probabilidad de que un habitante seleccionado aleatoriamente tenga gripe.

$$P(A) = P(B_1)P(A\setminus B_1) + P(B_2)P(A\setminus B_2) + P(B_3)P(A\setminus B_3)$$

$$P(A) = 0{,}25 \times 0{,}05 + 0{,}35 \times 0{,}04 + 0{,}40 \times 0{,}02$$

$$P(A) = 0{,}0345$$

b) Si un habitante tiene gripe cual es la probabilidad de que sea anciano o niño.

$$P((B_1 \cup B_3)\setminus A) = P(B_1\setminus A) + P(B_3\setminus A)$$

$$P(B_1\setminus A) = \frac{P(B_1)P(A\setminus B_1)}{P(A)}$$

$$P(B_1\setminus A) = \frac{0{,}25 \times 0{,}05}{0.0345} = 0{,}36$$

$$P(B_3\setminus A) = \frac{P(B_3)P(A\setminus B_3)}{P(A)}$$

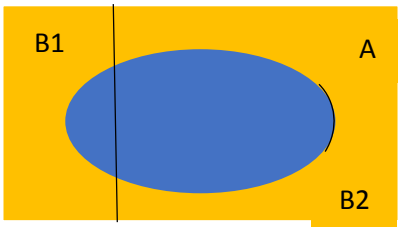
$$P(B_3\setminus A) = \frac{0{,}40 \times 0{,}02}{0.0345} = 0{,}23$$

$$P((B_1 \cup B_3)\setminus A) = 0{,}36 + 0{,}23 = 0{,}59$$

Ejercicio 49:

En una oficina hay seis computadoras de marca y cuatro clones, la probabilidad que al utilizar una máquina, esta encienda correctamente es 0,95 para la de marca y 0,8 para los clones, un empleado utiliza al azar una computadora, hallar la probabilidad:

a.- Se encienda correctamente.



A= "al escoger al azar una máquina, esta encienda correctamente".

$\boldsymbol{B}_{1=}$"al escoger al azar una máquina, esta sea de marca".

$\boldsymbol{B}_{2=}$"al escoger al azar una máquina, esta sea clon".

$$P(A) = P(B_1)P(A\backslash B_1) + P(B_2)P(A\backslash B_2)$$

$$P(A) = \frac{6}{10} \times 0{,}95 + \frac{4}{10} \times 0{,}8$$

$$P(A) = 0{,}89$$

De cada 100 veces al encender una maquina el 89% se enciente correctamente.

b.- La probabilidad que al escoger al azar una máquina, esta sea de marca dado que encienda correctamente.

$$P(B_1\backslash A) = \frac{P(B_1)P(A\backslash B_1)}{P(A)}$$

$$P(B_1\backslash A) = \frac{\frac{6}{10} \times 0{,}95}{0{,}89} = 0{,}64$$

De cada 100 máquinas que han funcionado correctamente en 64 veces corresponden a las de marca.

c.- La probabilidad que al escoger al azar una máquina , esta sea clon dado que encienda correctamente.

$$P(B_2\backslash A) = \frac{P(B_2)P(A\backslash B_2)}{P(A)}$$

$$P(B_2\backslash A) = \frac{\frac{4}{10} \times 0{,}8}{0{,}89} = 0{,}36$$

De cada 100 veces que se han encendido una maquina correctamente en 36 veces corresponden a clones.

EJERCICIOS PROPUESTOS

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios 2.9, literales (22, 24, 25, 26, 27) pág. 78

1) En una fábrica, el 70% de los empleados son lojanos. De entre los lojanos, el 50% son hombres, mientras que los no lojanos, solo son hombres el 20%.Encuentre:

a) ¿Qué porcentaje de empleados no lojanos son mujeres?

b) Probabilidad de que un empleado de la oficina sea mujer

c) Fernando trabaja en dicha oficina ¿Cuál es la probabilidad de que sea lojano?

R: a) 80% **b)** 0.59 **c) 0.8537**

2) Se estima que solo un 20% de los que compran acciones en Bolsa tienen conocimientos bursátiles de ellos el 80% obtienen beneficios. De los que compran acciones sin conocimiento bursátiles , solo un 10% obtienen beneficios .Se desea saber:

a) El tanto por ciento de los que compran acciones en bolsa que obtienen beneficios.

b) Si se elige al azar una persona que ha comprado acciones en bolsa y resulta que ha obtenido beneficios, ¿Cuál es la probabilidad de que tenga conocimientos bursátiles?

R: a) 24% **b)** 0.667

3) En un supermercado el 70% de la compras la realizan las mujeres, de la compras realizadas por estas, el 80% supera los 20$, mientras que de las compras realizadas por hombres solo el 30% supera esta cantidad.

a) Elegido un comprobante al azar, ¿Cuál es la probabilidad de que supere los 20$?

b) Si se sabe que el comprobante de compra no supera los 20$, ¿Cuál es la probabilidad de que la compra haya sido hecha por una mujer?

R: a) 0.65 b) 0.4

4) En una universidad existen tres facultades: A, B y C. En A hay matriculadas 150 chicas y 50 chicos; en B, 300 chicas y 200 chicos, y en C, 150 chicas y 150 chicos.

a) Calcule la probabilidad de que un estudiante, elegido al azar, sea chico

b) Si un estudiante elegido al azar resultara ser chico ¿Cuál es su faculta más probable?

R: a) 0.4 **b)** La B: Pr (B/H) = 0.5

5) Entre los cinco aspirantes a un cargo de gerente, a dos se les considera excelentes y a los demás se les considera buenos. Para una entrevista se escoge al azar a dos de los cinco. Calcule la probabilidad de que se escoja:

a) A los dos excelentes

b) Por lo menos a uno de los excelentes

c) A los dos excelentes dado que se sabe que por lo menos uno de los seleccionados es excelente

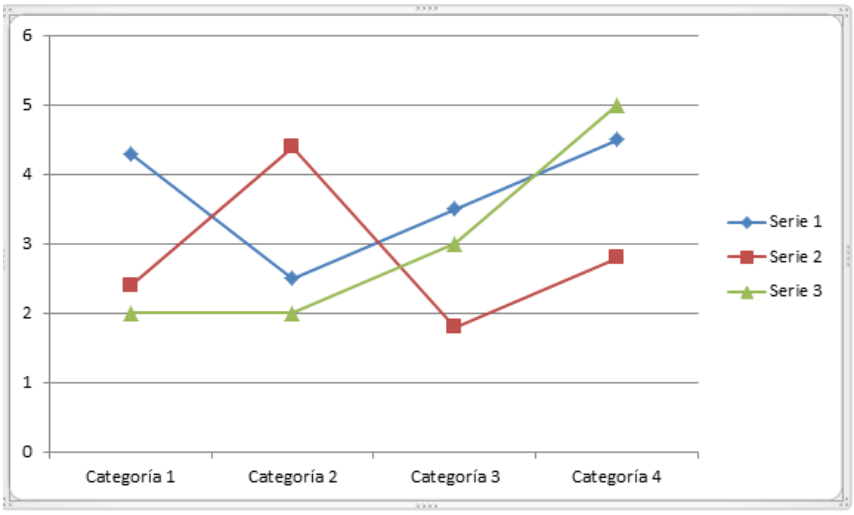
R: a) 1/10 **b)** 7/10 **c)** 1/7

Estadística y probabilidad con R

Ing. Mat. Willam Caiza, docente UPS.

CAPÍTULO 4

Variables Aleatorias



CONTENIDO

4.1 Definición de Variable aleatoria

- **4.1.1 Función de distribución v.a.d (variable aleatoria discreta**

4.2 Esperanzas, varianzas y correlaciones de una y dos variables aleatorias

- **4.2.1 Correlación**
- **4.2.2 Esperanza matemática**
- **4.2.3 Esperanza matemática de una variable aleatoria continua**
- **4.2.4. Propiedades de la esperanza matemática**
- **4.2.5 La varianza**

4.3 Función Generadora de Momentos

- **4.3.1. Distribución de poisson**
- **4.3.2 Distribución uniforme discreta**
- **4.3.3. Distribución hipergeométrica**
- **4.3.4. Distribución geométrica**
- **4.3.5. Distribución binomial negativa**
- **4.3.6. Distribución de bernoulli**
- **4.3.7. Distribución binomial**

CAPÍTULO 4

VARIABLES ALEATORIA

4.1 Definición de Variable aleatoria

Se llama variable aleatoria a cualquier función definida en un espacio muestral con recorrido en un subconjunto finito o infinito de los reales.

$$X : \Omega \rightarrow R$$
$$A \rightarrow X(A) = y$$

Ejercicio 50:

Experimento: "Lanzamiento de una moneda"

A = "sale cara"= 1

B = "sale sello"= 0

X= "valor que sale en el lanzamiento de una moneda"

X= {0,1}

P("Sale cara") = P(X = 1) = $\frac{1}{2}$

Ejercicio 51:

Experimento: "Lanzamiento de un dado"

X= " valor que sale en el lanzamiento de un dado "

X= "Sale 1" = 1

x = "Sale 2" = 2

x = "Sale 3" = 3

x = "Sale 4" = 4

x = "Sale 5" = 5

x = "Sale 6" = 6

X= {1, 2, 3, 4, 5,6}

P(X=2) = P("sale 2") = $\frac{1}{6}$

Ejercicio 52:

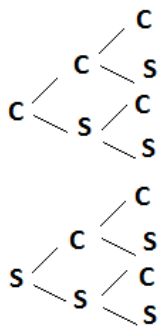
Experimento: "Tomar los pesos del curso de probabilidad y estadística"

X= "Peso de los estudiantes"

X ∈ [a, b], la cual es una variable aleatoria continua

Ejercicio 53:

Consideremos la prueba consistente en arrojar 3 monedas si X es la variable aleatoria "cuenta el número de caras", entonces X puede tomar los valores.



$\Omega = \{ CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}$

X= "Número de caras resultantes"

x = {0, 1, 2, 3}

x = {SSS}= 0

x = {CSS}= 1

x = {CSS}= 2

x = {CCC}= 3

4.1.1 Función de distribución variable aleatoria discreta

Sea X una variable aleatoria discreta, la función real F se define: $\forall\, t \in \mathbb{R},\ F(t) = P(X \leq t)$, se denomina función de distribución de la variable aleatoria discreta.

PROPIEDADES

- La función es creciente y sus gráfico tiene las siguientes características

$$\lim_{t \to +\infty} F(t) = 1 \qquad y \qquad \lim_{t \to -\infty} F(t) = 0$$

- $\boldsymbol{F(t) = P(X \leq t) = \sum_{j \leq t} p_j}$

En la propiedad se puede notar p_j , lo cual se podría ilustrar con el siguiente ejemplo.

Experimento: "Lanzamiento tres monedas"

X = "Número de caras"

P(X=3) = 1/8 = p_3

P(X=0) = 1/8 = p_0

P(X=1) = 3/8 = p_1

- $P(a < x \leq b) = F(b) - F(a)$
- $P(a \leq x \leq b) = F(b) - F(a) + P(X = a)$
- $P(x = a) = F(a) - F(a_)$
- $P(x < a) = F(a_)$
- $P(a < x < b) = F(b) - F(a) - P(X = b)$

Ejercicio 54:

Sea X una variable aleatoria discreta cuya función de probabilidad es $\boldsymbol{p(x) = \frac{k}{x}}$, x= 1,2,3,4,5.

a) Encuentre el valor de k para que la función p(x) sea la función de probabilidad de x.

Para que sea función de probabilidad se debe tener $\sum_{i=1}^{n} p_i = 1$

$$p_1 + p_2 + p_3 + p_4 + p_5 = 1$$

$$\frac{k}{1} + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k}{5} = 1$$

$$274\, k = 120$$

$$\boldsymbol{k = \frac{120}{274} = 0,437}$$

b) Calcule la probabilidad P(1< X ≤ 4).

Primera forma

$$\mathrm{P}(1 < \mathrm{X} \leq 4) = \mathrm{P}(\mathrm{X} = 2) + \mathrm{P}(\mathrm{X} = 3) + \mathrm{P}(\mathrm{X} = 4)$$
$$= \frac{0{,}437}{2} + \frac{0{,}437}{3} + \frac{0{,}437}{4} = 0{,}4734$$

Ejercicio 55:

Dada la siguiente ley de probabilidad

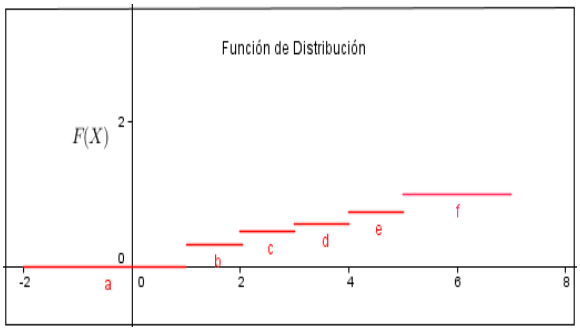
x_j	1	2	3	4	5
$\mathrm{P}(\mathrm{X} = \mathrm{x}_j) = p_j$	0.3	0.2	0.1	0.15	0.25

a.- Calcular F(x)

Se tiene que el dominio de la función de distribución es {1,2,3,4,5}, por lo tanto F(x) = 0, si x < 1 y F(x) = 1 para x ≥ 5.

- Si 1 ≤ x < 2, $\boldsymbol{F(1) = P(X \leq 1) = \sum_{j \leq 1} p_j = p_1 = 0,3}$
- Si 2 ≤ x < 3, $\boldsymbol{F(2) = P(X \leq 2) = \sum_{j \leq 2} p_j = p_2 + p_1 = 0,2 + 0,3 = 0,5}$
- Si 3 ≤ x < 4, $\boldsymbol{F(3) = P(X \leq 3) = \sum_{j \leq 3} p_j = p_3 + p_2 + p_1 = 0,1 + 0,2 + 0,3 = 0,6}$
- Si 4 ≤ x < 5, $\boldsymbol{F(4) = P(X \leq 4) = \sum_{j \leq 4} p_j = p_4 + p_3 + p_2 + p_1 = 0,15 + 0,1 + 0,2 + 0,3 = 0,75}$
- Si x ≥ 5, F(5) =1.

$$F(X) = \begin{cases} 0 & si\ x < 1 \\ 0{,}3 & si\ 1 \leq x < 2 \\ 0{,}5 & si\ 2 \leq x < 3 \\ 0{,}6 & si\ 3 \leq x < 4 \\ 0{,}75 & si\ 4 \leq x < 5 \\ 1 & si\ x \geq 5 \end{cases}$$



b.- Calcular la P(X=1)

$$P(X = 1) = F(1) - F(1_) = 0.3 - 0 = 0.3$$

c.- Calcular la P(X ≤ 1)

$$P(X \leq 1) = P(-a < X \leq 1) = F(1) - F(-a) = 0.3 - 0 = 0.3$$

d.- Calcular P(X < 1)

$$P(X < 1) = F(1_) = 0$$

e.- Calcular P(1 < X ≤ 2)

P(1 < X ≤ 2)= F(2) – F(1) = 0.5 -0.3 = 0.2

f.- Calcular F(2,5)

$$F(2.5) = P(X \leq 2.5) = \sum_{j \leq 2.5} p_j = p_{2.5} + p_2 + p_1 = 0 + 0.2 + 0.3 = 0.5$$

4.1.2 Función de distribución con variable aleatoria continua

Sea X una variable aleatoria continua, la función real F se define: $\forall\, t \in \mathbb{R},\ F(t) = P(X \leq t)$, se denomina función de distribución de la variable aleatoria continua.

PROPIEDADES

- La función es creciente y sus gráfico tiene las siguientes características

$$\lim_{t \to +\infty} F(t) = 1 \qquad y \qquad \lim_{t \to -\infty} F(t) = 0$$

- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$

- $P(X = a) = 0$

4.1.3 Función de densidad

La función de densidad de una variable X continua es una función $f: R \rightarrow R$.

Propiedades

- La función es estrictamente positiva, es decir: f(x) ≥ 0.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

Matemáticamente la función de distribución F(X) y la función de densidad f(x) se relacionan mediante el Teorema Fundamental del Cálculo:

$$F(x) = \int_{-\infty}^{x} f(t)dt \; ; \quad F'(x) = f(x).$$

Si f(x) es una función de densidad entonces para cualquier intervalo en [a, b], se define $P(a \leq X \leq b) = \int_a^b f(x)dx$= F(b) –F(a).

Ejercicio :

Dada la función de densidad f, encuentre el valor de la constante c de tal manera que esté bien definida.

$$f(x) = \begin{cases} cx(x-1), si \;\; 2 \leq x < 4 \\ 0\,, \qquad caso\; contrario \end{cases}$$

Desarrollo:

$$\int_2^4 (\boldsymbol{cx^2 - cx})\boldsymbol{dx} = \boldsymbol{c}\int_2^4 \boldsymbol{x^2 dx} - \boldsymbol{c}\int_2^4 \boldsymbol{x dx} = \boldsymbol{c}\left(\frac{\boldsymbol{4^3}}{\boldsymbol{3}} - \frac{\boldsymbol{2^3}}{\boldsymbol{3}}\right) - \boldsymbol{c}\left(\frac{\boldsymbol{4^2}}{\boldsymbol{2}} - \frac{\boldsymbol{2^2}}{\boldsymbol{2}}\right) = \boldsymbol{1}$$

$$\frac{\boldsymbol{56}}{\boldsymbol{3}}\boldsymbol{c} - \frac{\boldsymbol{12}}{\boldsymbol{2}}\boldsymbol{c} = \boldsymbol{1}$$

$$\frac{\boldsymbol{76}}{\boldsymbol{6}}\boldsymbol{c} = \frac{\boldsymbol{38}}{\boldsymbol{3}}\boldsymbol{c} = \boldsymbol{1}$$

Resolviendo $c = \frac{3}{38}$

Ejercicio

Considere la variable aleatoria continua z con función de densidad

$$f(z) = \begin{cases} (1+b)z^b & si \;\; 0 \leq z \leq a \\ 0 & en\ caso\ contrario \end{cases}$$

a.- Calcule los valores de a y b conociendo $P\left(z \leq \frac{1}{2}\right) = \frac{1}{8}$

$$\int_0^{1/2} (1+b)z^b dz = \frac{1}{8}$$

$$(1+b)\int_0^{1/2} z^b dz = \frac{1}{8}$$

$$(1+b)\frac{z^{b+1}}{b+1}|_0^{\frac{1}{2}} = \frac{1}{8}$$

$$1/2^{b+1} - 0^{b+1} = \frac{1}{8}$$

$$\left(\frac{1}{2}\right)^{b+1} = \left(\frac{1}{2}\right)^3$$

resolviendo la ecuación exponencial b+1 =3 por lo tanto b=2

$$3\int_0^a z^2 = 3\frac{z^3}{3}|_0^a = 1$$

$$a^3 - 0^3 = 1$$

Resolviendo la ecuación tenemos que a =1.

b.- Encuentre la función de distribución F(Z)

Con los datos calculados

$$f(z) = \begin{cases} 3z^2 & si \;\; 0 \leq z \leq 1 \\ 0 & en\ caso\ contrario \end{cases}$$

Por definición de función de distribución conocemos:

- F(z) = 0 si x <0
- F(z) =1 si x≥1

Si 0 < z <1

$$F(z) = \int_{-\infty}^{z} 3t^2\, dt = t^3|_0^z = z^3 - 0^3 = z^3$$

Finalmente la F(Z)

$$\begin{cases} 0 & si\ z < 0 \\ z^3 & si\ 0 < z < 1 \\ 1 & si\ z \geq 1 \end{cases}$$

Ejercicio

Una empresa alquila el tiempo de cómputo de un tipo de computadora. La empresa debe planear su presupuesto, por lo que ha estudiado el tiempo de empleo de la computadora.

El tiempo semanal de alquiler en horas sigue una función de densidad

$$f(t) = \begin{cases} \frac{3}{64}t^2(4-t) & si \ \ 0 \le t \le 4 \\ 0 & en\ caso\ contrario \end{cases}$$

a.- Determinar la función de distribución del tiempo de empleo de la computadora.

Por definición de función de distribución se tiene

- Si x < 0, F(X) = 0
- Si x≥4, F(X) = 1

Si 0 ≤ X < 4

$$F(X) = \int_{-\infty}^{x} \frac{3}{16}t^2 dt - \int_{-\infty}^{x} \frac{3}{64}t^3\, dt$$

$$= \frac{1}{16}t^3|_0^x - \frac{3}{64}\frac{t^4}{4}\ |_0^x = \frac{1}{16}x^3 - \frac{3}{256}x^4$$

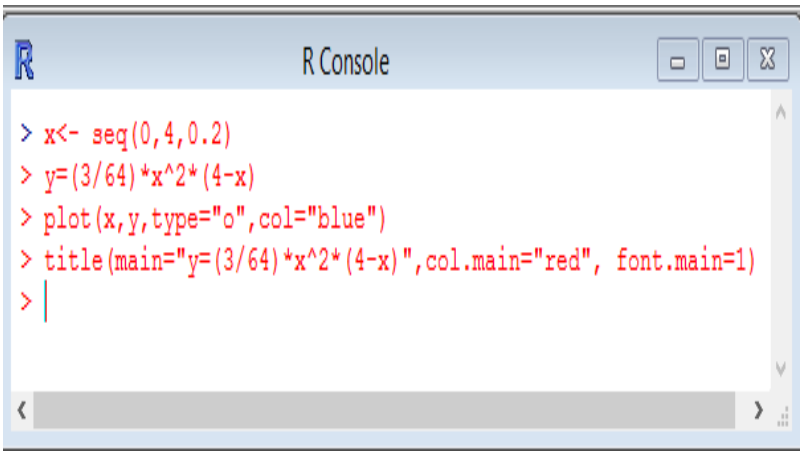
$$F(X) = \begin{cases} 0\,, & si \ \ x < 0 \\ \frac{1}{16}x^3 - \frac{3}{256}x^4\,, & si\ 0 \le \mathrm{X} < 4 \\ 1\,, & si \ \ x \ge 4 \end{cases}$$

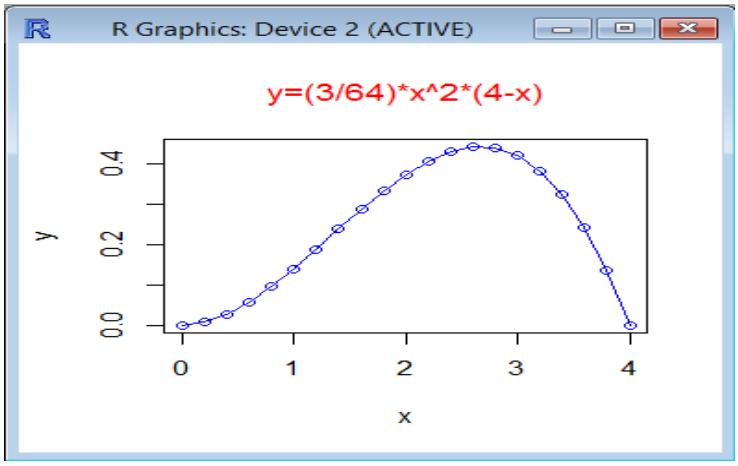
b.- Calcule la probabilidad de que el tiempo de uso de la computadora, en una semana sea mayor que 2 horas.

X= “tiempo de uso de la computadora”

P (X > 2) = $\int_2^4 \frac{3}{64}t^2(4-t)\,dt = \frac{1}{16}t^3|_2^4 - \frac{3}{64}\frac{t^4}{4}\ |_2^4 = \frac{56}{16} - \frac{3(240)}{256} = \frac{176}{256} = \frac{11}{16}$

c.- ¿Cuánto tiempo de alquiler se debe presuponer por semana si esta cifra solo se puede rebasar con una probabilidad de 0.1?





Si el tiempo de uso de la computadora tiene a f(x) como función de densidad se podría decir que en f(x) está el máximo valor de uso, tomando el 10% de la derecha de la gráfica se tendría

$$P(X \leq t) = 0.9$$

$$\frac{3}{64}\int_0^t t^2(4-t)dt = 0.9$$

$$\int_0^t t^2(4-t)dt = 19{,}2$$

$$\frac{4}{3}t^3|_0^x - \frac{t^4}{4}|_0^x = 19{,}2$$

$$16x^3 - 3x^4 - 230{,}4 = 0$$

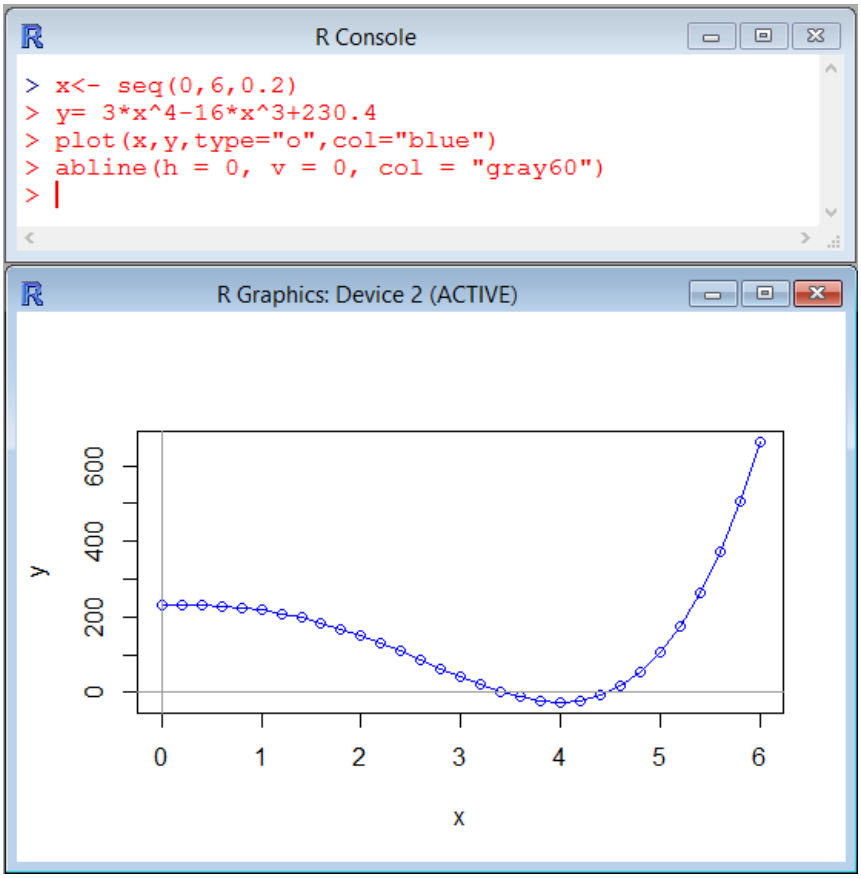
$$3x^4 - 16x^3 + 230{,}4 = 0$$

Resolviendo mediante el método de Newton- Raphson

$$X_{i+1} = X_i - \frac{f(X_i)}{f'(X_i)}$$

Con $f(x) = \mathbf{3x^4 - 16x^3 + 230,4}$ y $f'(x) = \mathbf{12x^3 - 48x^2}$

Para iniciar la resolución de la ecuación realicemos el gráfico correspondiente para tener el valor inicial de la raíz para el cálculo numérico.



En el gráfico anterior se puede observar que existen dos raíces entre [3.5], pero necesitamos conocer la raíz que se encuentra entre [3, 4] ya que la otra raíz se encuentra fuera del dominio.

Sin nombre - Editor R

```
NewtonRaphson <- function(p){

 x<-numeric()
 x[1]<- p
 print(c('columna 1-> iteración i',' columna 2 -> raíz'))
 for (i in 1:10){
    r<-x[i]
    f<-3*r^4-16*r^3+230.4
    f1<- 12*r^3-48*r^2

    x[i+1]= x[i]- (f/f1)

    print(c(i,x[i]))
  }

 }

 NewtonRaphson(2)
```

Realizando el programa en R del método de Newton – Raphson, se tiene que su valor se aproxima a 3,4297, como se puede observar claramente en el gráfico siguiente.

R Console

```
[1] "columna 1-> iteración i" " columna 2 -> raíz"
[1] 1 2
[1] 2.000000 3.566667
[1] 3.000000 3.414415
[1] 4.000000 3.429629
[1] 5.000000 3.429763
[1] 6.000000 3.429763
[1] 7.000000 3.429763
[1] 8.000000 3.429763
[1] 9.000000 3.429763
[1] 10.000000  3.429763
> 
```

4.2 ESPERANZAS, VARIANZAS, COVARIANZAS Y CORRELACIONES.

4.2.1 Covarianza muestral

La covarianza muestral se define para las variables X e Y, la misma que mide la variabilidad conjunta, definida por:.

$$S_{xy} = \frac{1}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y})$$

Si x=y, la covarianza se transforma en varianza y mide la variabilidad entre la misma variable.

$$S_{xy} = S_{xx} = \frac{1}{n-1}\sum (x_i - \bar{x})(x_i - \bar{x})$$

$$S_{xx} = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = Var\,(X)$$

4.2.2 Coeficiente de correlación

Al igual que la covarianza el coeficiente de correlación mide la relación lineal existente entre las dos variable, y es la más utilizada ya que no depende de la escala en la que se mida.

$$r = \frac{S_{xy}}{S_x S_y} \qquad ; \qquad -1 \leq r \leq 1$$

Si el valor de r es cercano a 1, se podría decir que existe una correlación lineal positiva fuerte.

Si el valor de r es cercano a -1, se podría decir que existe una correlación lineal negativa fuerte.

Si el valor de r es cercano a 0, se podría decir que no existe correlación lineal

4.2.3 Esperanza matemática discreta.

La esperanza matemática de una variable aleatoria discreta X es:

$$E(X) = \sum_{i=1}^{n} x_i\, pi = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$$

Se podría considerar a la esperanza matemática como un promedio, o lo que se esperaría de la variable aleatoria X.

Como ejemplo se podría tomar $E(\bar{x}) = \mu$, lo que significa que la esperanza o promedio de la media aritmética es igual a la media poblacional, base fundamental donde se une la estadística descriptiva e inferencial.

Ejercicio 54:

Halle la esperanza de las variables aleatorias discretas definidas por:

X	-0,71	0,24	0,61
P	0,2	0,5	0,3

$$E(X) = \sum_{i=1}^{n} x_i\, pi$$

$$E(X) = \sum_{i=1}^{n} (-0{,}71)(0{,}2) + (0{,}24)(0{,}5) + (0{,}61)(0{,}3)$$

$$E(x) = 0{,}158$$

Lo que indica que en promedio X tendría el valor de 0,158.

4.2.4 ESPERANZA MATEMÁTICA DE UNA VARIABLE ALEATORIA CONTINUA

La esperanza matemática de una variable aleatoria continua es:

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx\,, \forall\, x\ \in I$$

Si X se encuentra definido en el intervalo [a, b], entonces la esperanza matemática se define:

$$E(X) = \int_{a}^{b} x\, f(x)\, dx\,, \forall\, x\ \in [a, b]$$

f(x) es función de densidad.

Ejercicio 55:

Hallar la esperanza matemática de la siguiente variable aleatoria continua:

$$f(x) = \begin{cases} \frac{1}{3} & \text{si } 1 < x \leq 4 \\ 0 & \text{caso contrario} \end{cases}$$

$$E(X) = \int_{a}^{b} x\, f(x)\, dx$$

$$E(X) = \int_{1}^{4} x\ \frac{1}{3} dx$$

$$E(X) = \frac{1}{3}\int_{1}^{4} x\ dx$$

$$E(X) = \frac{1}{3}\frac{x^2}{2}$$

$$E(X) = \frac{1}{3}\left[\frac{16}{2} - \frac{1}{2}\right]$$

$$E(X) = \frac{5}{2}$$

4.2.5. PROPIEDADES DE LA ESPERANZA MATEMÁTICA

- **E(C) = C, C es una constante**

Conocemos

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

Por lo tanto

$$E(c) = \int_{-\infty}^{+\infty} c\, f(x)\, dx$$

$$E(c) = c\int_{-\infty}^{+\infty} f(x)\, dx$$

$$\boldsymbol{E(c) = c}$$

- $\boldsymbol{E(X+Y) = E(X) + E(Y)}$

Conocemos

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

además

$$E(g(z)) = \int_{-\infty}^{+\infty} g(z)\, f(x)\, dx$$

Sea g (z)= x + y

$$E(x+y) = \int_{-\infty}^{+\infty} (x+y)\, f(x)\, dx$$

$$E(x+y) = \int_{-\infty}^{+\infty} x\, f(x)\, dx + \int_{-\infty}^{+\infty} y\, f(x)\, dx$$

$$\boldsymbol{E(x+y) = E(x) + E(y)}$$

- $\boldsymbol{E(cx) = cE(x)}$

$$E(cx) = \int_{-\infty}^{+\infty} cx\, f(x)\, dx$$

$$E(cx) = c\int_{-\infty}^{+\infty} x\, f(x)\, dx$$

$$\boldsymbol{E(cx) = c\, E(x)}$$

4.2.6 LA VARIANZA

La varianza de una variable aleatoria x (discreta o continua) se define:

$$\boldsymbol{Var(X) = E(X - E(X))^2}$$

Demuestre: $\boldsymbol{Var(X) = EX^2 - (E(X))^2}$

Demostración:

$$Var(X) = EX^2 - (E(X))^2$$

por definición

$$Var(x) = E\big(x - E(x)\big)^2$$

factorando

$$Var(X) = E(X^2 - 2XE(X) + (E(X))^2)$$

Aplicando la esperanza

$$Var(X) = E(X^2) - 2E(XE(X)) + E((E(X))^2)$$

La primera observación a tomar en cuenta es E(X) es constante

$$Var(X) = E(X^2) - 2E(X)E(X) + (E(X))^2$$

$$Var(X) = E(X^2) - 2(E(X))^2 + (E(X))^2$$

$$\boldsymbol{Var(X) = EX^2 - (E(X))^2}$$

4.2.6.1 Varianza de una variable aleatoria discreta

Sea X una la variable aleatoria discreta varianza se define:

$$Var(X) = \sum_{i=1}^{n} (x_i - EX)^2 p_i$$

$$Var(X) = \sum_{i=1}^{n} x_i^2\, p_i - (EX)^2$$

4.2.6.2 Definición de varianza variable aleatoria continúa:

Sea x una variable aleatoria continua la varianza se define:

$$Var(X) = \int_{-\infty}^{+\infty} (x - EX)^2 f(x)dx$$

$$Var(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2$$

si f(x) se define en $I \in [a, b]$

$$Var(X) = \int_{a}^{b} x^2 f(x)dx - (EX)^2$$

4.2.6.3 PROPIEDADES DE LA VARIANZA

- $\boldsymbol{Var(c) = 0\ ; c\ es\ cte}$

$$Var(x) = \int_{-\infty}^{+\infty} (x - Ex)^2 f(x)dx$$

$$Var(c) = \int_{-\infty}^{+\infty} (c - E(c))^2 f(x)dx$$

$$Var(c) = \int_{-\infty}^{+\infty} (c - c)^2 f(x)dx$$

$$\boldsymbol{Var(c) = 0}$$

- $\boldsymbol{Var(cX) = c^2 Var(X)}$

$$Var(cX) = \int_{-\infty}^{+\infty} (cx - Ecx)^2 f(x)dx$$

$$Var(cX) = \int_{-\infty}^{+\infty} (c(x - E(x)))^2 f(x)dx$$

$$Var(cX) = c^2 \int_{-\infty}^{+\infty} (x - E(x))^2 f(x)dx$$

$$\boldsymbol{Var(X) = c^2 Var(X)}$$

- $\boldsymbol{Var(X+Y) = Var(X) + Var(Y)}$

Conocemos: $Var(X) = E(X - EX)^2$

$$Var(X+Y) = E[(X+Y) - E(X+Y)]^2$$

$$Var(X+Y) = E(x+y)^2 - 2E((x+y)E(x+y)) + (E(x+y))^2$$

$$Var(X+Y) = E(X^2 + 2XY + Y^2) - 2E(X+Y)E(X+Y) + (E(X+Y))^2$$

$$Var(X+Y) = E(X^2 + 2XY + Y^2) - 2((E(X) + E(Y))(E(X) + E(Y))) + (E(X) + E(Y))^2$$

$Var(X+Y) = E(X^2) + 2E(XY) + E(X^2) - 2(EXEX + EXEY + EXEY + EYEY) + (EX)^2 + 2EXEY + (EY)^2$ (1)

pero $Var(X) = E(X - EX)^2 = EX^2 - 2EXEX + (EX)^2$

Y Covarianza entre X e Y es: Cov (X,Y) = E((X-EX)(Y-EY)) = EXY-EYEX-EXEY+EXEY

En (1)

$$Var(X+Y) = [E(x^2) - 2EXEX + (EX)^2] + [E(y^2) - 2EYEY + (EY)^2] + 2[EXY - EXEY - EYEX + EXEY]$$

$$Var(X+Y) = Var(X) + Var(Y) + 2\,cov(X,Y)$$

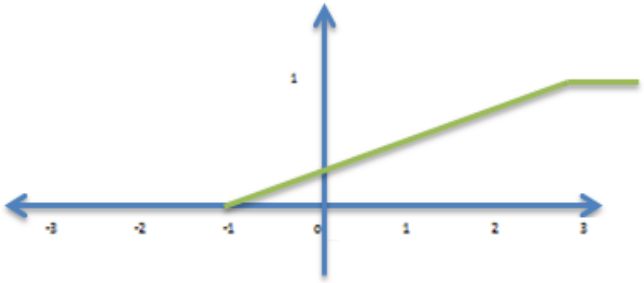
Pero si X e Y son independientes la Cov(X,Y) = 0

$$\boldsymbol{Var(X+Y) = Var(X) + Var(Y)}$$

Ejercicio 34

Halle la esperanza y la varianza de una variable aleatoria X que tiene como función de distribución:

$$F(x) = \begin{cases} 0; & x \leq -1 \\ \dfrac{x+1}{4}; & -1 < x \leq 3 \\ 1; & x > 3 \end{cases}$$



Desarrollo:

Por definición $F'(x) = f(x)$, derivando tenemos

$$f(x) = \begin{cases} \frac{1}{4} & ; -1 < x \leq 3 \\ 0 & ; en\ otro\ caso \end{cases}$$

$$E(X) = \int_{-1}^{3} xf(x)dx = \int_{-1}^{3} \frac{x}{4}dx = \frac{1}{4}\int_{-1}^{3} xdx = \frac{1}{4}\frac{\boldsymbol{x^2}}{\boldsymbol{2}}\ |_{\boldsymbol{-1}}^{\boldsymbol{3}} = \frac{\boldsymbol{1}}{\boldsymbol{8}}\left(\boldsymbol{3^2} - (\boldsymbol{-1})^{\boldsymbol{2}}\right) = \boldsymbol{1}$$

$$Var(X) = EX^2 - (EX)^2$$

$$EX^2 = \int_{-1}^{3} x^2 f(x)\,dx = \frac{1}{4}\int_{-1}^{3} x^2\,dx = \frac{1}{4}\frac{x^3}{3}|_{-1}^{3} = \frac{1}{12}(3^3 - (-1)^3) = \frac{7}{3}$$

$$Var(X) = \frac{7}{3} - 1 = \frac{4}{3}$$

Ejercicio 35

Halle la esperanza y la varianza de la variable aleatoria discreta definida por:

Y	2	4	5	6
P	0,3	0,1	0,2	0,4

Desarrollo

$$E(Y) = \sum_{i=1}^{4} y_i p_i = (2(0{,}3) + 4(0{,}1) + 5(0{,}2) + 6(0{,}4)) = 4{,}4$$

$$Var(Y) = \sum_{i=1}^{4} {y_i}^2\, p_i - (EY)^2$$

$$\sum_{i=1}^{4} {y_i}^2\, pi = [4(0{,}3) + 16(0{,}1) + 25(0{,}2) + 36(0{,}4)] = \frac{111}{5} = 22{,}2$$

$$Var(Y) = 22{,}2 - (4{,}4)^2 = 2{,}84$$

Ejercicio 36

Sean X, Y, Z variables aleatorias independientes cada una con media μ y varianza σ^2.

a.- Calcule la esperanza y varianza de S=X+Y+Z.

- $E(S) = E(X + Y + Z)$

$E(S) = E(X) + E(Y) + E(Z)$

$E(S) = \mu + \mu + \mu$

$E(S) = 3\mu$

- $Var(S) = Var(X + Y + Z)$

Por ser X,Y,Z independientes

$Var(S) = Var(X) + Var(Y) + Var(Z)$

$Var(S) = \sigma^2 + \sigma^2 + \sigma^2$

$Var(S) = 3\sigma^2$

b.- Encuentre la esperanza y varianza T=3X

- $E\ (T) = E(3X)$

$E(T) = 3E(X)$

$E(T) = 3\mu$

- $Var(T) = Var(3X)$

$Var(T) = 9\,Var(X)$

$Var(x) = 9\sigma^2$

Ejercicio 37

Sea X una variable aleatoria discreta en el intervalo [-1; 1] y f(x) su función de densidad. Encuentre E(x) si f(x)= $|x|$.

Desarrollo

$$f(x) = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

$$E(x) = \int_{-1}^{1} xf(x)dx = \int_{-1}^{0} \mathrm{x}(-\mathrm{x})dx + \int_{0}^{1} x(x)dx$$

$$E(x) = \int_{-1}^{0} -x^2dx + \int_{0}^{1} x^2dx = -\left(\frac{x^3}{3}\right)|_{-1}^{0} + \frac{x^3}{3}|_{0}^{1} = -\frac{1}{3}(0+1) + \frac{1}{3}(1) = -\frac{1}{3} + \frac{1}{3}$$

$$= \mathbf{0}$$

4.3 FUNCIÓN GENERADORA DE MOMENTOS

Definición:

El r-ésimo momento alrededor del origen de la variable aleatorio X está dado por:

$$\mu'_r = E(x^r) = \sum_i x^r p_i$$

Como el primer y segundo momento alrededor del origen son dados por $\mu'_1 = E(x)$ y $\mu'_2 = E(x^2)$, podemos escribir la media y la varianza de una variable aleatoria como

$$\mu = E(x) = \mu'_1$$

$$\sigma^2 = Var(x) = \mu'_2 - (\mu'_1)^2$$

Definición:

La función generadora de momentos de la variable aleatoria discreta X, se denota como Mx(t).

$$Mx(t) = E(e^{tx}) = \sum_x e^{tx} p_i$$

La función generadora de momentos existirá si la sumatoria converge, si existe una función generadora de momentos de una variable aleatoria discreta, se puede utilizar para generar todos los momentos de dicha variable.

Teorema:

Sea X una variable aleatoria con función generadora de momentos Mx(t), se define:

$$\left.\frac{d^r Mx(t)}{dt^r}\right|_{t=0} = E[x^r] = \mu'_r$$

Desarrollo

$$Mx(t) = E(e^{tx})$$

$$\frac{dMx(t)}{dt} = \frac{d}{dt}[E(e^{tx})]$$

La derivada de la E(X) es igual a la esperanza de la derivada de X, $\frac{d}{dt}E(X) = E(\frac{d}{dt}X)$

$$\frac{dMx(t)}{dt} = E\left[\frac{d}{dt}(e^{tx})\right]$$

$$\frac{dMx(t)}{dt} = E[xe^{tx}]$$

$$\left.\frac{dMx(t)}{dt}\right|_{t=0} = E[xe^{0x}]$$

$$\left.\frac{dMx(t)}{dt}\right|_{t=0} = E(x) = \mu'_1$$

Calculando la segunda derivada

$$Mx(t) = E(e^{tx})$$

$$\frac{dMx(t)}{dt} = \frac{d}{dt}[E(e^{tx})]$$

$$\frac{dMx(t)}{dt} = E[xe^{tx}]$$

$$\frac{d^2Mx(t)}{dt^2} = E\left(x\,\frac{d}{dt}e^{tx} + e^{tx}\frac{d}{dt}x\right) = E[x^2e^{tx} + 0]$$

$$\left.\frac{d^2Mx(t)}{dt^2}\right|_{t=0} = E[x^2e^{0x}]$$

$$\left.\frac{d^2Mx(t)}{dt^2}\right|_{t=0} = E(x^2) = \mu'_2$$

En general siguiendo con el proceso de diferenciación se obtiene:

$$\left.\frac{d^rMx(t)}{dt^r}\right|_{t=0} = E[x^r] = \mu' r$$

Ejercicio

Encontrar la esperanza y la varianza de la Distribución Binomial, mediante el método de los momentos.

$P(X = x_k) = p_k = {}_nC_k p^k q^{n-k}$; k=0,1,2,3, …,n

Demostrar:

$$Mx(t) = (pe^t + q)^n$$

Desarrollo

$Mx(t) = E(e^{tx}) = \sum_{k=0}^{n} p_k\, e^{tx_k}$; $\mathrm{E}(x) = \sum p_k\, x_k$

$= \sum_{k=0}^{n} {}_nC_k\, p^k q^{n-k} e^{tk}$; $p_k = {}_nC_k p^k\, q^{n-k}$; $X_k = k$, y k es constante por ser X discreta e inicia en 0 hasta n.

$= \sum_{k=0}^{n} nC_k (pe^t)^k q^{n-k}$; $(x+y)^n = \sum_{i=0}^{n} {}_nC_i y^i\, x^{n-i}$

$= (q + pe^t)^n$.

Demostrar que E(X) = np

$Mx(t) = (pe^t + q)^n$; derivando la función generadora de momentos

$$\frac{dMx(t)}{dt} = n(pe^t + q)^{n-1}(pe^t)$$

$\frac{dMx(t)}{dt}\Big| t = 0 = n(pe^0 + q)^{n-1}(pe^0)$; evaluando en t=0.

$$\frac{dMx(t)}{dt}\Big| t = 0 = n(p+q)^{n-1}p$$

$\frac{dMx(t)}{dt}\Big| t = 0 = n(1)^{n-1}p$; p + q =1

$$\frac{dMx(t)}{dt}\Big| t = 0 = np = E(X)$$

Encontrar $E(X^2)$

Si tenemos

$\frac{dMx(t)}{dt} = npe^t(pe^t + q)^{n-1}$

$\frac{d^2Mx(t)}{dt^2} = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + npe^t(pe^t + q)^{n-1}$

$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = n(n-1)(pe^0 + q)^{n-2}(pe^0)^2 + n(pe^0 + q)^{n-1}(pe^0)$

$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = n(n-1)(p+q)^{n-2}p^2 + np(p+q)^{n-1}$; p+q=1

$$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = n(n-1)(1)^{n-2}(p)^2 + np(1)^{n-1}$$

$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = np^2(n-1) + np = np(p(n-1)+1) = np(np-p+1)$; p=1-q

$$= np(np - 1 + q + 1) = np(np+q) = (np)^2 + npq = EX^2$$

$$\boldsymbol{Var(X) = E(x^2) - (EX)^2 =} \ (np)^2 + npq - (np)^2 = npq$$

Ejercicio

Encontrar la esperanza y varianza de la distribución Geométrica, utilizando el método de los momentos.

$P(X = k) = P_k = p(1-p)^{k-1}$; $x_k = \boldsymbol{k}$ y k= 1,2,3,…,∞

Demostrar $M_x(t) = E(e^{tx}) = \frac{pe^t}{1-pe^t}$

Desarrollo

$$E(e^{tx}) = \sum_{k=1}^{\infty} p_k e^{tx_k} = \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{tx_k}$$

$$= p\sum_{k=1}^{\infty} (1-p)^{k-1} \ e^{tk}$$

$$= \ p\sum_{k=1}^{\infty} (1-p)^{k}(1-p)^{-1} \ e^{tk}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} [(1-p)e^t]^k$$

$$= \frac{p}{q}\sum_{k=1}^{\infty} [qe^t]^k$$

Sea la serie geométrica

$$\frac{p}{q}(1 + (qe^t) + (qe^t)^2 + (qe^t)^3 + \cdots + (qe^t)^n + \cdots - 1)$$

A la serie anterior se le suma 1 y -1, para que la serie comience desde k=0.

$$= \frac{p}{q}\left(\frac{1}{1-qe^t} - 1\right)$$

$$Mx(t) = \frac{p}{q}\left(\frac{\cancel{1} - \cancel{1} + qe^t}{1-qe^t}\right) = \frac{pe^t}{1-qe^t}$$

Demostrar $E(x) = \frac{1}{p}$

$$\frac{dMx(t)}{dt} = \frac{pe^t(1-qe^t) - pe^t(-qe^t)}{(1-qe^t)^2}$$

$$\frac{dMx(t)}{dt} = \frac{pe^t - \cancel{pqe^{2t}} + \cancel{pqe^{2t}}}{1-2qe^t+q^2e^{2t}}$$

$$\frac{dMx(t)}{dt} = \frac{pe^t}{(1-qe^t)^2}$$

$$\frac{dMx(t)}{dt}\Big|_{t=0} = \frac{pe^0}{1-2qe^0+q^2e^o}$$

$$\frac{dMx(t)}{dt}\Big|_{t=0} = \frac{p}{1-2q+q^2}$$

$$\frac{dMx(t)}{dt}\Big|_{t=0} = \frac{p}{(1-q)^2}$$

$$\frac{dMx(t)}{dt}\Big|_{t=0} = \frac{1}{p} = \boldsymbol{EX}$$

Calcular la EX^2

$$\frac{dMx(t)}{dt} = \frac{pe^t}{(1-qe^t)^2}$$

$$\frac{d^2Mx(t)}{dt^2} = \frac{pe^t(1-qe^t)^2 - pe^t[2(1-qe^t)(-qe^t)]}{(1-qe^t)^4}$$

$$= \frac{pe^t - 2pqe^{2t} + pq^2e^{3t} - pe^t(-\mathbf{2}\boldsymbol{q}e^t + \mathbf{2}\boldsymbol{q}^2\boldsymbol{e}^{\mathbf{2}\boldsymbol{t}})}{(1-qe^t)^4}$$

$$\frac{d^2Mx(t)}{dt^2} = \frac{pe^t - 2pqe^{2t} + pq^2e^{3t} + 2pqe^{2t} - 2pq^2e^{3t}}{(1-qe^t)^4}$$

$$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = \frac{pe^0 - pq^2e^0}{(1-qe^0)^4}$$

$$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = \frac{p-pq^2}{(1-q)^4}$$

$$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = \frac{p(1-q^2)}{p^4}$$

$$\frac{d^2Mx(t)}{dt^2}\Big|_{t=0} = \frac{1-q^2}{p^3} = E(x^2)$$

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

$$Var(x) = \frac{1-q^2}{p^3} - \frac{1}{p^2}$$

$$Var(x) = \frac{1-q^2-p}{p^3}$$

$$Var(x) = \frac{1-p-(1-p)^2}{p^3}$$

$$Var(x) = \frac{1-p-1+2p-p^2}{p^3} = \frac{\boldsymbol{p}-p^2}{p^3} = \frac{p(1-p)}{p^3} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Ejercicio

Encontrar la esperanza y varianza de la distribución Poisson, mediante el método de los momentos.

$P(X=k) = p_k = \frac{e^{-\lambda}\lambda^k}{k!}$; $\lambda > 0$; $x = 0,1,2,3, \ldots$

Demostrar

$$Mx(t) = e^{\lambda e^t - \lambda} = e^{\lambda(e^t-1)}$$

Desarrollo

$$Mx(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx_k}\ p_k = \sum_{x=0}^{\infty} e^{tk}\ p_k$$

$$Mx(t) = \sum_{x=0}^{\infty} e^{tk}\ (\frac{e^{-\lambda}\lambda^k}{k!})$$

$$Mx(t) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{e^{tk}\lambda^k}{k!}$$

$$Mx(t) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

$$Mx(t) = e^{-\lambda}\left[1 + \frac{\lambda e^t}{1!} + \frac{(\lambda e^t)^2}{2!} + \frac{(\lambda e^t)^3}{3!} + \cdots + \frac{(\lambda e^t)^n}{n!} + \cdots..\right]$$

Se puede deducir que

$$\frac{z^0}{0!}+\frac{z'}{1!}+\frac{z^2}{2!}+\cdots+\frac{z^n}{n!}+\cdots=e^z=e^{\lambda e^t}$$

$$Mx(t)=e^{-\lambda}e^{\lambda e^t}=e^{\lambda e^t-\lambda}=e^{\lambda(e^t-1)}$$

Encontrar E (X) = λ

$$\frac{dMx(t)}{dt}=e^{\lambda(e^t-1)}\,\lambda e^t$$

$$\left.\frac{dMx(t)}{dt}\right|t=0=e^{\lambda(e^0-1)}\,\lambda e^0=e^{\lambda(1-1)}\,\lambda(1)=e^{\lambda(0)}\,\lambda(1)=\lambda$$

$$\left.\frac{dMx(t)}{dt}\right|t=0=\lambda=E(x)$$

Calcular EX^2

$$\frac{dMx(t)}{dt}=e^{\lambda e^t-\lambda}\,\lambda e^t$$

$$\frac{d^2Mx(t)}{dt^2}=e^{\lambda e^t-\lambda}\,\lambda e^t\;+e^{\lambda e^t-\lambda}\,\lambda e^t\,\lambda e^t$$

$$\left.\frac{d^2Mx(t)}{dt^2}\right|t=0=\;e^{\lambda e^0-\lambda}\,\lambda e^0\;+e^{\lambda e^0-\lambda}\,\lambda e^0\,\lambda e^0=\lambda+\lambda^2$$

$$\left.\frac{d^2Mx(t)}{dt^2}\right|t=0=\lambda^2+\lambda=E(x^2)$$

$$Var(X)=\lambda^2+\lambda-\lambda^2=\;\lambda$$

EJERCICIOS PROPUESTOS

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios 3.6, literales (23, 24, 25, 26) pag 112,113

1. Un supermercado tiene una demanda diaria variable x de la cantidad de carne que vende, de tal manera que X (medida en cientos de kilogramos) tiene una función de densidad de:

$$f(x)=\begin{cases}\frac{3}{64}x^2, & \text{si } 0 \le x < 1\\ 0, & \text{Caso contrario}\end{cases}$$

a) Calcule el valor esperado de la cantidad de carne demandada y su desviación estándar.

b) Si la ganancia está dado por Y= 2x – 0.5. Calcule la ganancia esperada y su varianza.

R: a) E(x) = 3, σ(x) = 0.7746; b) E (y)=5.5, Var (y)=12/5

2. La longitud de ciertos gusanos se distribuye según la función de densidad

$$f(x)=\begin{cases}\frac{3}{4}(x-1)(3-x), & \text{si } 1 \le x < 3\\ 0, & \text{Caso contrario}\end{cases}$$

a) Calcule la esperanza y la varianza de la longitud de los gusanos

b) Si para un estudio se consideran aquellos ejemplares que tenga una longitud entre 1.7 cm y 2.4 cm, calcule el porcentaje de gusanos que tienen esta característica.

R: a) E(x) = 2, Var(x)= 0.2; b) 50.225%

3. El tiempo de uso diario de la red internet en una oficina tiene por función de densidad (medida en horas)

$$f(x)=\begin{cases}\frac{3x^2(8-x)}{1024}, & \text{si } 0\leq x<8\\ 0, & \text{caso contrario}\end{cases}$$

- Calcule el valor esperado y la varianza del tiempo diario de uso de internet
- El tiempo de uso de internet cuesta 2 dólares por hora. Calcule el valor esperado y la desviación estándar del costo semanal (en 5 días laborables) por el citado uso.

R: a) E (T) = 4.8 h, Var (T)=2.56 h^2 ; b) E(c)=48, σ(c) = 16

4. La ley de probabilidad que describe la distancia (en metros) a la que un atleta lanza la jabalina

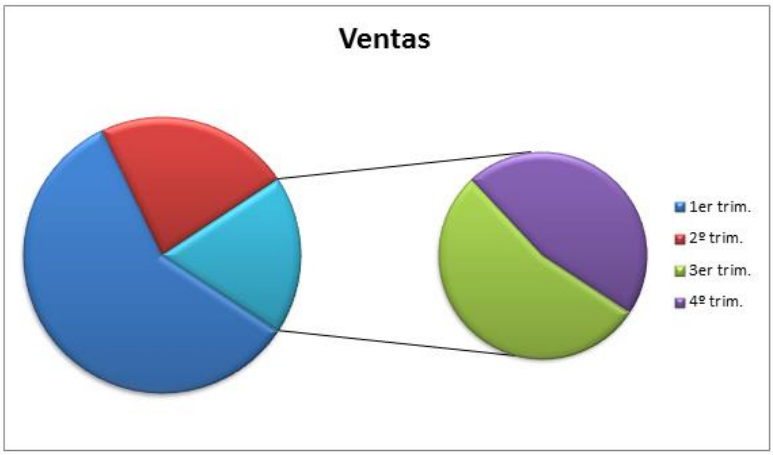
$$f(x)=\begin{cases}\frac{1}{20\pi}\ \text{sen}\left(\frac{x}{10\pi}\right), & \text{si } x\in[0, 10\pi^2]\\ 0, & \text{caso contrario}\end{cases}$$

- Halle la probabilidad de que una jabalina lanzada llegue a una distancia mayor que 60m
- Determine el valor esperando de la distancia a la que llega la jabalina
- Halle la varianza y la desviación estándar de la distancia cubierta por la jabalina.

R: a) Pr(x>60)=0.334; b) E(x) = $5\pi^2$ = 49.348; c) Var(x)=461.3m^2, σ(x) = 21.478 m

CAPÍTULO 5

Distribuciones de variables aleatorias



Contenido

5.1 Distribución variable aleatoria discretas

- **5.1.1 Distribución Binomial y Bernoulli**
- **5.1.3 Distribución hipergeométrica**
- **5.1.4 Distribución geométrica**
- **5.1.5 Distribución de poisson**
- **5.1.6 Distribución binomial negativa**

5.2 Distribución Variable Aleatoria Continuas

- **5.2.1 Distribución Uniforme**
- **5.2.2 Distribución exponencial**
- **5.2.3 Distribución normal**
- **5.2.4. Teorema del Límite Central**
- **5.2.5 Distribución gamma**

CAPÍTULO 5: DISTRIBUCIONES DE VARIABLES ALEATORIAS

5.1 Distribución variable aleatoria discretas

5.1.1 Distribución Bernoulli

Una prueba de Bernoulli es un experimento aleatorio cuyos posibles resultados son éxito y fracaso, la P(éxito) = p y P(fracaso)= q y p + q = 1.

Si X es una variable aleatoria discreta de Bernoulli, se puede definir

$$X = \begin{cases} 1\,; para\ el\ éxito \\ 0\ \ ; para\ el\ fracaso \end{cases}$$

Propiedades:

- E(X) =p
- Var(X) = pq

Dado el evento "Pasar el semestre", los posibles resultados de la consulta son "pasó el semestre" ó "no pasó el semestre", lo que nos daría la idea de una prueba de Bernoulli, donde el éxito = "pasó el semestre" y fracaso = "no pasó el semestre".

5.1.2 Distribución Binomial

Dado n pruebas de Bernoulli, si se requiere conocer, la probabilidad de obtener k éxitos en las n pruebas de Bernoulli, se hace uso de la probabilidad Binomial.

$$P(X = k) = \quad {}_n C_k\, p^k\, q^{n-k}$$

n = número de pruebas de Bernoulli

p = P (éxito)

q = P (fracaso)

k = k-éxitos

n - k = número de fracasos

Propiedades:

- X ~ Bin(n,p) ; X sigue una distribución Binomial de parámetros n, p.
- $E(X) = np$

- $Var(X) = npq$

Ejercicio 58

Un dispositivo está compuesto por tres elementos que trabajan independientemente. La probabilidad de falla de cada elemento en un día es 0,1. Formar la ley de distribución de probabilidad del número de elementos de falla en un día.

Primera Forma

Si X = 0

P(" cero fallen")= P(el primero funcione y el segundo funcione y el tercero funcione)

Los eventos son independientes.

P("cero fallen")=P(el primero funcione) P(el segundo funcione) P(el tercero funcione)

P("cero fallen")=(0.9) (0.9) (0.9)= 0,729

Si X=1 entonces:

P(uno falle) = P(primero falle y segundo funcione y tercero funcione o primero funcione y el segundo falle y tercero funcione o primero funcione y segundo funcione y tercero falle)

= P(primero falle y segundo funcione y tercero funcione)+P(primero funcione y el segundo falle y tercero funcione) + P(primero funcione y segundo funcione y tercero falle) = P(primero falle)P(segundo funcione)P(tercero funcione)+ P(primero funcione)P(segundo falle)P(tercero funcione)+ P(primero funcione)P(segundo funcione)P(tercero falle)=(0,1) (0,9)(0,9)+ (0,9)(0,1)(0,9)+(0,9)(0,9)(0,1)= 3 (0,1)(0,9)(0,9)= 0,243.

Si X=2

P(dos falle)= P(primero falle y segundo falle y tercero funcione o primero funcione y segundo falle y tercero falle o primero falle y segundo funcione y tercero falle) =

= P(primero falle)P(segundo falle)P(tercero funcione)+ P(primero funcione)P(segundo falla)P(tercero falle)+ P(primero falle)P(segundo funcione)P(tercero falle)= (0,1)(0,1)(0,9)+(0,9)(0,1)(0,1)+(0,1)(0,9)(0,1)= 3 (0,1)(0,1)(0,9)= 0,027

Si X = 3

P(tres fallen) = P(primero falle y segundo falle y tercero falle) = P(primero falle)P(segundo falle)P(tercero falle) = (0,1)(0,1)(0,1)

Segunda Forma

X= "Número de elementos que fallan"

X= {0, 1, 2, 3}

Éxito= "elemento falle"

P(éxito) = p = 0,1

$$P(x = 0) = \quad {}_3C_0 0{,}1^0\ 0{,}9^3 = \ 0{,}9^3$$

$$P(x = 1) = \quad {}_3C_1\ 0{,}1^1\ 0{,}9^2$$
$$= 3(0.1^1)(0.9^2)$$

$$P(x = 2) = \quad {}_3C_2\ 0{,}1^2\ 0{,}9^1 = 3\ (0{,}1^2)\ (0{,}9^1)$$

$$P(x = 3) = \quad {}_3C_3\ 0{,}1^3\ 0{,}9^0 = \ 0{,}1^3$$

Se puede observar que en la primera forma se ha calculado la probabilidad de forma lógica la cual es bastante tedioso, en la segunda forma se hace uso de la probabilidad binomial con la cual los cálculos se han simplificado.

Ejercicio 59

Una máquina llena de cajas de palillos de fósforo. En una porción del 10% la máquina no llena las cajas por completo. Se toman al azar 25 cajas de fósforo. Calcule la probabilidad de que no haya más de 2 cajas incompletas

Desarrollo

Éxito = "cajas incompletas"

X = " número de cajas incompletas"

n = 25

P(éxito) = p = 0,1

$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$P(x = 0) = {}_{25}C_0\, 0{,}1^0 0{,}9^{25} = 0{,}07$

$P(x = 1) = {}_{25}C_1\, (0{,}1)^1 (0{,}9)^{24} = 0{,}199$

$P(x = 2) = {}_{25}C_2\, (0{,}1)^2 (0{,}9)^{23} = 0{,}265$

$P(x \leq 2) = 0{,}537$

Realización del ejercicio en R:

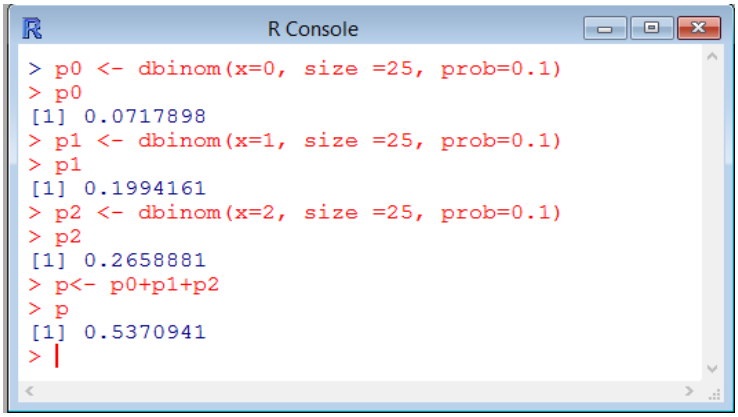


Ilustración 7 Tips R 19 Distribución Binomial

Ejercicio 60

Una encuesta revela que el 20% de la población es favorable a un político y el resto es desfavorable. Si se eligen 6 personas al azar se desea saber:

a.- La probabilidad de que las 6 personas sean desfavorables

b.- La probabilidad de que a 4 de las 6 personas sean favorables

Desarrollo parte a.

n=6 pruebas de Bernoulli

Éxito = "voto desfavorable"

$P(éxito) = p = 0{,}8$

X= "Número de personas desfavorables"

$P(x = 6) = 6C_6 (0{,}8)^6 (0{,}2)^0 = 0{,}26$

Desarrollo parte b

Éxito="voto favorable"

$P(exito) = p = 0{,}2$

X= “Número de personas favorables”

$P(X = 4) = 6C_4(0{,}2)^4(0{,}8)^2 = 0{,}015$

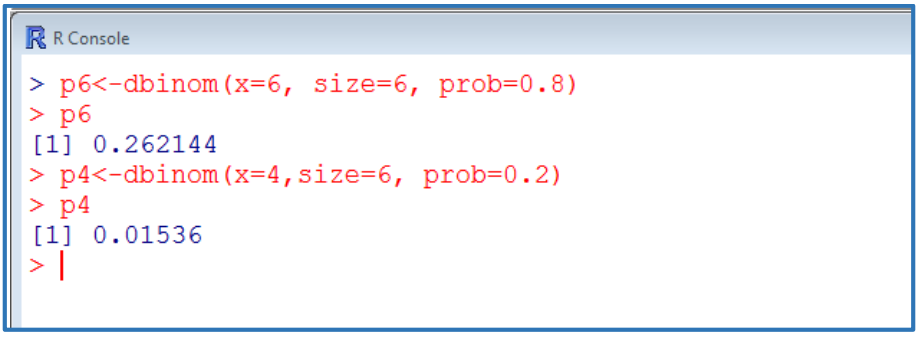


Ilustración 8 Tips R 20. Distribución Binomial

Ejercicio 61

Una aeronave dispone de 4 motores que funcionan independientemente la probabilidad de que falle 1 motor durante un vuelo sea 0,01. ¿Cuál es la probabilidad de que un vuelo dado?:

a) No se observen fallas
b) No se observe más de una falla
c) Si un avión puede seguir volando si al menos dos motores continúan funcionando ¿Cuál es la probabilidad de que el avión se accidente?

Desarrollo literal a

Éxito=” motores fallen”

X =”Número de motores fallen”

$P(éxito) = p = 0{,}01$

$P(x = k) = nC_k p^k q^{n-k}$

$P(\text{"}no\ se\ observen\ fallas\text{"}) = P(x = 0) = 4C_4(0{,}01)^0(0{,}99)^4 = 0{,}96$

Desarrollo literal b

Éxito=”motores fallen”

X =”Número de motores fallen”

$P(éxito) = p = 0{,}01$

$P(\text{"}no\ se\ observa\ mas\ de\ una\ falla\text{"}) = p(x \leq 1)$

$= P(x = 0) + P(x = 1)$

$$= 4C_0(0{,}01)^0(0{,}99)^4 + 4C_1(0{,}01)^1(0{,}99)^3 = 0{,}9994$$

Desarrollo literal c

Éxito=”motores funcionen”

X =”Número de motores funcionen”

$P(éxito) = p = 0{,}99$

$P(\text{"}por\ lo\ menos\ 2\ motores\ funcionen\text{"}) = p(x \geq 2)$

$= P(X = 2) + P(X = 3) + P(X = 4)$

$$= 4C_2(0{,}99)^2(0{,}01^2) + 4C_3(0{,}99)^3(0{,}01) + 4C_4(0{,}99)^4(0{,}01^0)$$

$= 0{,}99999603$ es la probabilidad que no se accidente, entonces la

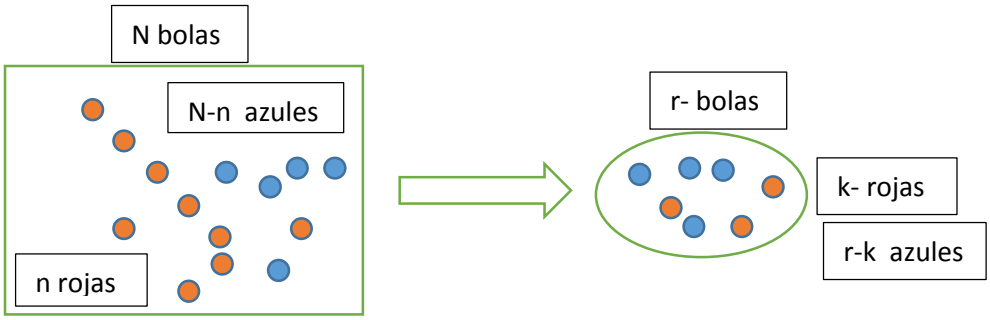
P(accidente el avión) = 1- P(no se accidente el avión) = $1 - \ 0{,}99999603 = \ 3{,}96X10^{-6}$

5.1.3 Distribución Hipergeométrica

La distribución hipergeométrica es útil en aquellos casos en los que se extraigan muestras.

Es importante recalcar que tanto la distribución binomial como la distribución hipergeométrica persiguen el mismo objetivo “el número de éxitos en una muestra que contiene n observaciones”, la diferencia radica en que que la hipergeométrica considera no solo a los elementos de la muestra, sino también a los elementos de la población.

La principal aplicación es en el muestreo de aceptación y control de calidad donde de un lote de artículos se toma una muestra y se analiza para decidir si se acepta o rechaza el lote.



La distribución Hipergeométrica calcula la probabilidad de que haya k bolas rojas en una muestra de r bolas de una población de N bolas y n bolas rojas.

$$P(X = k) = \frac{nC_k\, N - n\; C_{r-k}}{NC_r}\;; \quad k = 0,1,\dots,\min\{n.r\}$$

Propiedades:

- $E(X) = \frac{rn}{N} = rp$
- $Var(X) = rpq\frac{N-r}{n-1}$
- X ~ H (N,n,r)

Ejercicio 62

En una línea de control de calidad se revisan 10 artículos determinando que hay 3 que no cumplen con las especificaciones. Si se escogen al azar 2 artículos identifique los parámetros de la ley y halle la esperanza de la variable aleatoria, que describe el número de piezas correctas de las dos escogidas.

X= "Número de piezas correctas"

X = {0,1,2}

$N = articulos$

r= 2

n= 7

$$P(X = 0) = \frac{7C_0\; 3C_2}{10C_2} = 0{,}06$$

$$P(X = 1) = \frac{7C_1\; 3C_1}{10C_2} = 0{,}46$$

$$P(X = 2) = \frac{7C_2 3C_0}{10C_2} = 0{,}46$$

$$E(X) = \sum_{i=1}^{3} x_i p_i$$

$$= 0 + 0{,}46 + 2(0{,}46)$$

$= \mathbf{1,38}$

$$E(X) = \frac{rn}{N} = rp = 2\left(\frac{7}{10}\right) = \frac{7}{5} = 1{,}4$$

$$Var(X) = 2\left(\frac{7}{10}\right)\left(\frac{3}{10}\right)\left(\frac{10-2}{10-1}\right) = \left(\frac{21}{50}\right)\left(\frac{8}{9}\right) = 0{,}373$$

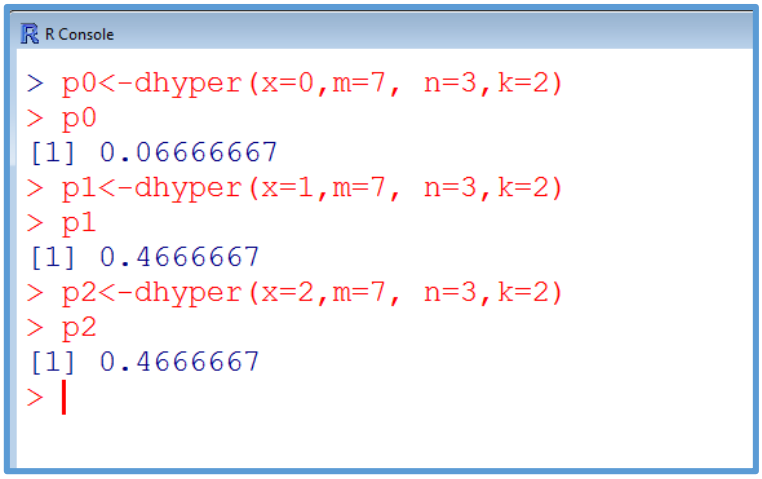


Ilustración 9 Tips R 22. Distribución Hipergeométrica.

Ejercicio

Para llenar 4 vacantes de contador se presentan 10 personas, 7 hombres y 3 mujeres. Salen seleccionado 3hombres y 1 mujer. Las mujeres acusan al empleador de discriminación sexual por que los llevan a juicio. Si el juez supone que la elección fue al azar, ¿ puede decirse que existió discriminación al hacer la elección?.

Solución:

Para poder solucionar el altercado, se debería calcular la probabilidad que salga 3 hombres y también la probabilidad de que salga 3 mujeres, si la probabilidad de que salgan 3 mujeres en el grupo de 4 vacantes es mayor a la probabilidad de que salgan 3 hombres en el grupo de 4 vacantes se podría decir que hubo una discriminación debido al sexo de la persona.

Para los hombres

r = 4 personas

N=10

n= 7

X= ”Número de hombres”

$$P(X = 3) = \frac{7C_3 3C_1}{10C_4} = 0{,}5$$

La probabilidad de salir 3 hombres en el grupo de 4 personas es 0,5.

Para las mujeres

N=10

n= 3

X =”Número de mujeres”

$$P(X = 3) = \frac{3C_3 7C_1}{10C_4} = 0{,}033$$

La probabilidad de salir 3 mujeres en el grupo de 4 personas es 0,033.

Se concluye que no habido descriminación.

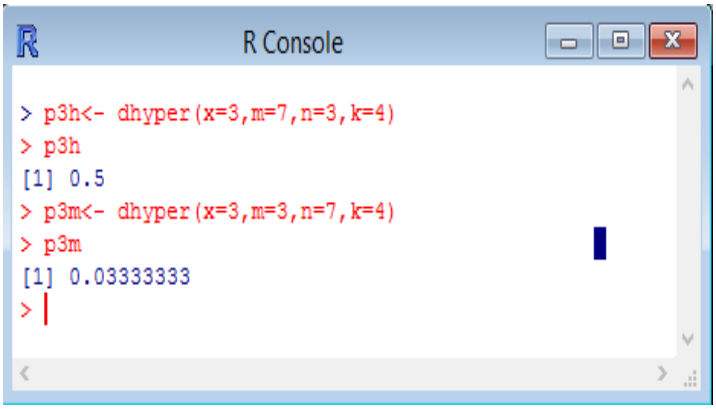


Ilustración 10lustración 11Tips R 23. Distribución Hipergeométrica.

5.1.4 Distribución Geométrica

Sea una sucesión de pruebas de Bernoulli con probabilidad de éxito p, en lugar de contar el número de éxitos, nos interesa conocer el número de intentos hasta obtener el primer éxito, tal sucesión se dice que es un experimento geométrico.

X = "Número de pruebas hasta obtener el primer éxito"

$P(X = k) = q^{k-1}p$; $k = 1,2 \ldots$

Propiedades

- $E(x) = \frac{1}{p}$
- $Var(x) = \frac{q}{p^2}$
- X ~ G(p)

Ejercicio 64:

Cuando se graba un comercial de tv la probabilidad de que un actor recite correctamente el dialogo de su toma es de 0.3. ¿Cuál es la probabilidad que el actor recite correctamente su dialogo en la 6ta vez, su esperanza y intervalo de confianza al 95%?

Desarrollo

Éxito="recite correctamente"

P(éxito)=p= 0,3

X = "Número de intentos hasta que recite correctamente"

$$P(X = 6) = (0,7)^5(0,3) = 0,05$$

$E(X) = \frac{1}{p} = \frac{1}{0,3} = 3,33$, indica que en promedio a la tercera vez recita correctamente

$$Var(x) = \frac{q}{p^2} = \frac{0,7}{0,3^2} = 7,78$$

$$\sigma(X) = 2,789$$

Intervalo de confianza al 95%

$$\left[3.3 - 1.96 * \left(\frac{2.78}{\sqrt{6}}\right); 3.3 + 1.96 * \left(\frac{2.78}{\sqrt{6}}\right)\right] = [1.08; 5.51]$$

Ejercicio

En un examen el profesor realiza varias preguntas a un estudiante. La probabilidad que de que el estudiante responda correctamente a cualquier preguntas es =0,9. El profesor interrumpe el examen apenas el estudiante manifiesta el desconocimiento de la pregunta hecha, calcular.

a.- Formar la ley de distribución de la variable aleatoria que describe el número de preguntas que realiza el profesor

b.- Hallar el número esperado de preguntas que ha de realizar el profesor.

Desarrollo

Éxito =”responda incorrectamente”

P(éxito) =p = 0,1

X= “número de preguntas a realizar por el profesor”

$$P(x = k) = q^{k-1}p$$

$$P(x = 1) = (0.9)^0(0,1) = 0,1$$

$$P(x = 2) = (0.9)^1(0,1) = 0,09$$

$$P(x = 3) = (0.9)^2(0,1) = 0,081$$

$$P(x = 4) = (0.9)^3(0,1) = 0,073$$

X	1	2	3	4	…
p	0,1	0,09	0,081	0,073	

Parte b

$$E(X) = \frac{1}{p} = \frac{1}{0,1} = 10$$

En promedio se espera que el profesor realice 10 preguntas

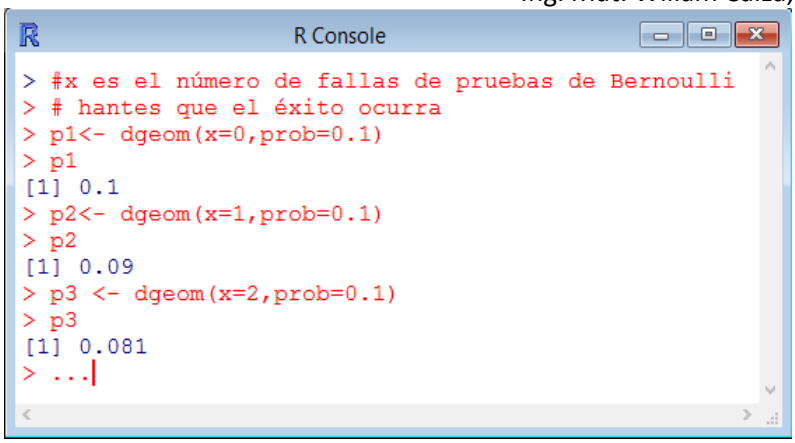


Ilustración 12 Tips R 25. Distribución Geométrica.

Ejercicio

Se sabe que aproximadamente el 20% de los usuarios de Windows no cierra el programa adecuadamente. Supongamos que Windows está instalado en una computadora pública que es utilizada aleatoriamente por personas que actúan independientemente.

a) ¿Cuál es la probabilidad de que el tercer usuario sea el primero que cierre adecuadamente Windows?
b) ¿Cuál es el número medio de personas que usan la computadora desde el momento que se enciende hasta que alguien no cierra adecuadamente?

Desarrollo

Parte a:

Éxito = cierre correctamente

P(éxito) = p= 0.8

X =”Número de personas que usan incorrectamente la computadora hasta que el primero cierre correctamente”

P(el tercer usuario sea el primero en cerrar correctamente) = $P(X = 3) = 0.2^2 0.8 = 0.032$

Parte b:

X=”número de personas que usan correctamente hasta el primero que cierre incorrectamente”

$$E(X) = \frac{1}{0.2} = 5$$

5.1.5 Distribución de Poisson

Sea X una variable aleatoria discreta, se dice que X sigue una distribución de Poisson con parámetro λ, con λ promedio de los sucesos en una unidad de tiempo.

La distribución de Poisson calcula la probabilidad de sucesos en una unidad de tiempo, definida por

$$P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{\mathrm{k!}} \quad ; \quad k = 0,1,2,3, \ldots$$

Propiedades:

- E(X)= λ
- $\mathrm{Var(x)} = \lambda$
- X∼ P (λ)

Comúnmente esta ley depende del tiempo (t), es conveniente escribirla de la siguiente forma:

$$P(X = k) = \frac{e^{-\lambda \mathrm{t}}(\lambda \mathrm{t})^{k}}{\mathrm{k!}} \quad ; \quad k = 0,1,2, \ldots$$

Existe una relación entre la distribución de Poisson y la distribución Binomial, cuando n >30 y p ≤ 0.05 y $\lambda = \mathrm{np}$. Lo anterior describe cuando se podría utilizar la distribución Binomial si los parámetros de Poisson tienen esos valores.

Ejercicio:

El promedio de llamadas que recibe una central telefónica en un minuto es de 1.5, halle la probabilidad de que en 4 minutos se reciban:

a) 3 llamadas

$$\lambda = 1,5\left[\frac{llamadas}{minuto}\right]$$

$$t = 4\, minutos$$

$X =$ "Número de llamadas en 4 minutos"

$$P(X = 3) = \frac{e^{-(1,5(4))}\left(1,5(4)\right)^{3}}{3!} = 0,089$$

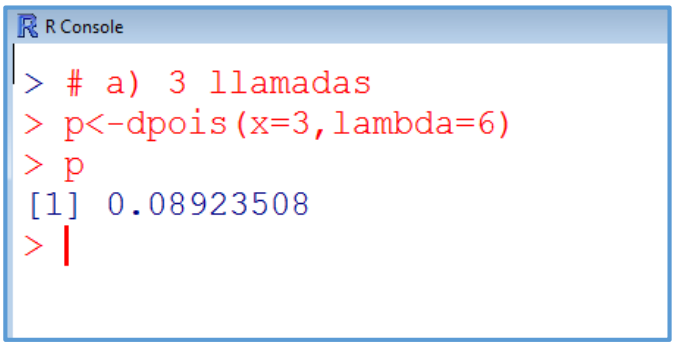


Ilustración 13 Tips R 26. Distribución Poisson.

b) Menos de 3 llamadas

$$P(x < 3) = P(x = 2) + P(x = 1) + P(x = 0)$$

$$= \frac{e^{-(1{,}5(4))}(1{,}5(4))^2}{2!} + \frac{e^{-(1{,}5(4))}(1{,}5(4))^1}{1!} + \frac{e^{-(1{,}5(4))}(1{,}5(4))^0}{0!}$$

$$= 0{,}0446 + 0{,}0148 + 2{,}47X10^{-3}$$

$$= 0{,}0619$$

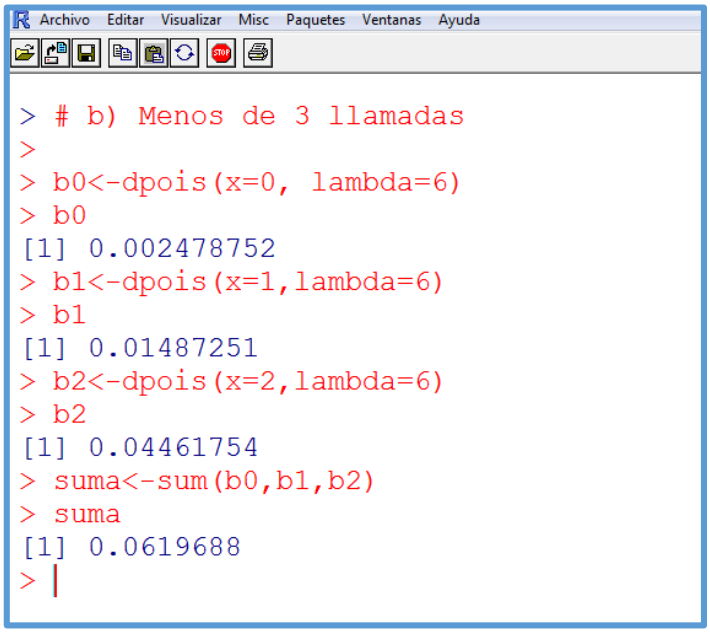


Ilustración 14 Tips R 27. Distribución Poisson.

c) No menos de 4 y no más de 7

$$P(4 \leq X \leq 7) = P(x = 4) + P(x = 5) + P(x = 6) + P(x = 7)$$

$$= \frac{e^{-(1{,}5(4))}(1{,}5(4))^4}{4!} + \frac{e^{-(1{,}5(4))}(1{,}5(4))^5}{5!} + \frac{e^{-(1{,}5(4))}(1{,}5(4))^6}{6!} + \frac{e^{-(1{,}5(4))}(1{,}5(4))^7}{7!}$$

$$= 0{,}133 + 0{,}16 + 0{,}16 + 0{,}14$$

$$= 0{,}593$$

```
R Console
> #c) No menos de 4 y no mas de 7
>
> c4<-dpois(x=4,lambda=6)
> c4
[1] 0.1338526
> c5<-dpois(x=5,lambda=6)
> c5
[1] 0.1606231
> c6<-dpois(x=6,lambda=6)
> c6
[1] 0.1606231
> c7<-dpois(x=7,lambda=6)
> c7
[1] 0.137677
> suma<-sum(c4,c5,c6,c7)
> suma
[1] 0.5927759
>
```

Ilustración 15 Tips R 28. Distribución Poisson.

Ejercicio:

En un hotel el promedio de pedidos a la habitación es igual a dos cada media hora. Halle la probabilidad de que en una hora se reciba

$$\lambda = 2\left[\frac{pedidos}{30min}\right]\left[\frac{60\ min}{1\ hora}\right] = 4\left[\frac{pedidos}{1\ hora}\right]$$

Si $t = 1h$

X= "Número de pedidos en una hora"

a) Tres pedidos

$$P(X = 3) = \frac{e^{-(4(1))}(4(1))^3}{3!} = 0{,}195$$

b) Menos de 3 pedidos

$$P(x < 3) = P(x = 2) + P(x = 1) + P(x = 0)$$

$$\frac{e^{-(4(1))}(4(1))^2}{2!} + \frac{e^{-(4(1))}(4(1))^1}{1!} + \frac{e^{-(4(1))}(4(1))^0}{0!}$$

$$= 0{,}146 + 0{,}073 + 0{,}018$$

$$= 0{,}237$$

Ejercicio

Una fábrica de gaseosas recibió 100 botellas vacías. La probabilidad de que al transportarlas resulte una botella rota es 0.03. Halle la probabilidad de que la fábrica reciba rotas:

a) Exactamente dos botellas

X=”Número de botellas”

$$\lambda = np = 100(0.03) = 3$$

$$P(X = 2) = \frac{e^{-3}(3)^2}{2!} = 0{,}22$$

Mediante la distribución binomial

$$P(X = 2) = 100C2\ \ 0.03^2\ 0.97^{98} = 0{,}22$$

b) Más de dos

$$P(X \geq 3) = 1 - \big(P(X = 0) + P(X = 1) + P(X = 2)\big)$$

$$= 1 - \left(\frac{e^{-3}(3)^0}{0!} + \frac{e^{-3}(3)^1}{1!} + \frac{e^{-3}(3)^2}{2!}\right)$$

$$= 1 - (0{,}04978 + 0{,}14936 + 0{,}224) = 1 - 0{,}42318 = 0{,}5768$$

c) Por lo menos una

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = 0{,}95$$

5.1.6 Distribución Binomial Negativa

Sea X una variable aleatoria Binomial Negativa con parámetros r y p. Se requiere conocer el número de pruebas de Bernoulli necesarias hasta obtener r éxitos, tomando en cuenta que el último éxito es r, su función de probabilidad es

$$P(X=k)=\binom{r-1}{k-1}p^r(q)^{k-r};\quad k=\ r,r+1,r+2,\dots$$

Propiedades

- $E(X)=\frac{r}{p}$
- $Var(X)=\frac{rq}{p^2}$
- $X\sim BN(r,p)$

Ejercicios

De un grupo de 50 esquizofrénicos, 12 padecen alteraciones cerebrales. calcule la probabilidad de que padezcan tal alteración:

a) el primero con alteración cerebral se encuentre en la octava consulta de historiales.
b) el tercero con alteración cerebral se encuentre en la décima consulta de historiales.

Literal a:

Para r = 1 éxito, ya que un éxito debe estar en la octava consulta tenemos

P(éxito)= P(padezcan alteraciones)= 12/50 = 0.24

$$P(X=8)=\binom{8-1}{1-1}\ 0.24^1\ 0.76^7=0.0351$$

Literal b:

Para r =3 éxitos, ya que un éxito debe estar en la décima consulta tenemos

P(éxito)= P(padezcan alteraciones)= 12/50 = 0.24

$$P(X=10)=\binom{10-1}{3-1}\ 0.24^3\ 0.76^7=0.0728$$

Ejercicio

Disponemos de una caja que contiene: 5 triángulos (T), 3 círculos (C) y 2 rectángulos (R). Realizando extracciones de figuras con reemplazamiento:

a) Calcule la probabilidad de que, al realizar 8 extracciones, se obtenga en 4 ocasiones un círculo.

b) Calcule la probabilidad de que se necesiten 8 extracciones para obtener 4 círculos.

Literal a.

P(éxito) = P(salga circulo)= 3/10= 0.3

X= "Número de círculos "

$$P(X = 4) = \binom{8}{4} 0.3^4 0.7^4 = 0.1361$$

Literal b:

Para r=4 éxitos, y como el último éxito debe ser el octavo, tenemos

$$P(X = 8) = \binom{8-1}{4-1} 0.3^4 0.7^4 = 0.068$$

EJERCICIOS PROPUESTOS

Fuente: GALINDO Edwin, Estadística - Métodos y Aplicaciones, ProCiencia Editores 2011, Ejercicios 4.6, literales (28, 29, 31, 32, 34) pág. 130

1) **Una jugadora de tenis gana el 30 % de los partidos que realiza. Ella jugara en un torneo mientras no sea eliminada en el partido.**

- Halle la probabilidad de que sea eliminada en el segundo partido;
- Si para ganar el torneo se debe ganar 5 partidos consecutivos, ¿cuál es la probabilidad de que la jugadora pierda en la final del torneo?
- ¿Cuántos partidos se espera que llegue a jugar durante el torneo?

Respuestas: a) 0.2211 b) 0.007946 c) 3 partidos

2) **Una marca de refrescos tiene impresas, en cada una de las tapas, una de las figuras de los cuatro jinetes del apocalipsis, y quien reúna la colección completa ganara un premio. Si un comprador cree que hay igual número de figuras de cada uno de los personajes en la proporción, ¿Cuántos refrescos a de esperar para ganar el premio?**

Respuesta: E(X)= 8.33, es decir 9 refrescos

3) **Un lepidopterista solo está interesado en los ejemplares de una clase de mariposas, que constituyen el 15% de todas las mariposas de la zona. Halle la probabilidad de que esta persona tenga que cazar 8 mariposas de las que no le interesan antes de comprar.**

- un ejemplar de clase deseada;
- tres ejemplares de clase deseada;

Respuestas: a) 0.04087 b) 0.03564

4) **En una fábrica, el departamento de control de calidad, revisa los lotes de piezas que entran, de acuerdo con el siguiente criterio: se van extrayendo piezas sucesivamente y el lote es rechazado si se encuentra la primera pieza defectuosa antes de la vigésima extracción. Si conocemos que el 2% de piezas son defectuosas, ¿cuál es la probabilidad que un lote sea rechazado?**

Respuestas: 0.3188

5) **Se sabe que, aproximadamente, el 20% de usuarios de Windows no cierran el programa adecuadamente. Supongamos que el Windows está instalado en una computadora pública que es utilizada aleatoriamente por personas que actúan independientemente unas de otras?**

- ¿Cuál es la probabilidad que el tercer usuario sea el primero que cierra adecuadamente el Windows?;
- ¿Cuál es el número medio de personas que usan la computadora desde el momento en que se enciende hasta que alguien no cierra el programa adecuadamente?;

Respuestas: a) 0.032 b) 5

5.2 Distribución Variable Aleatoria Continuas

5.2.1 Distribución Uniforme

Sea X una variable aleatoria continua se dice que X sigue una distribución Uniforme en un intervalo [a, b], si su función de densidad es:

$$f(x) = \begin{cases} \frac{1}{b-a}, & si\ x \in [a,b] \\ 0, & en\ otro\ caso \end{cases}$$

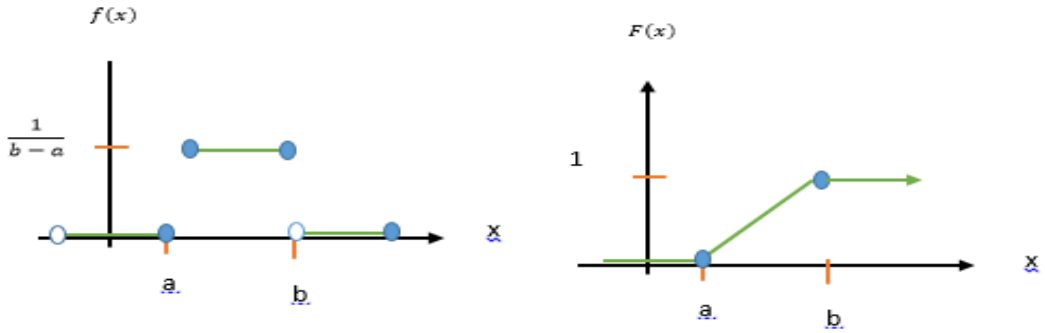
Y función de distribución:

$$F(x) = \begin{cases} 0, & si\ x < a \\ \frac{x-a}{b-a}, & si\ x \in [a,b] \\ 1, & si\ x > b \end{cases}$$

Propiedades:

- $E(x) = \frac{a+b}{2}$
- $Var\ (x) = \frac{(b-a)^2}{12}$
- $X \sim U[a,b]$

Sus gráficas son:



Ejercicio

Demostrar esperanza y varianza de la distribución Uniforme

$$E(x) = \int_a^b x\, f(x)$$

$$E(x) = \int_a^b x \left(\frac{1}{b-a}\right) dx$$

$$E(x) = \left(\frac{1}{b-a}\right) \frac{x^2}{2} |_a^b$$

$$E(x) = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right)$$

$$E(x) = \frac{1}{b-a}\left(\frac{b^2 - a^2}{2}\right)$$

$$\boldsymbol{E(x) = \frac{a+b}{2}}$$

$$Var(x) = E(x^2) - (E(X)^2)$$

Encontremos $E(x^2)$

$$E(x^2) = \int_a^b x^2 \left(\frac{1}{b-a}\right) dx$$

$$= \frac{1}{b-a} \frac{x^3}{3} |_a^b = \frac{1}{3(b-a)} (b^3 - a^3) = \frac{b^2 + ab + a^2}{3}$$

$$Var(x) = \frac{1}{3}(b^2 + ab + a^2) - (\frac{\boldsymbol{a+b}}{\boldsymbol{2}})^{\boldsymbol{2}}$$

$$= \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4}$$

$$= \frac{4b^2 + 4ab + 4a^2 - 3(a^2 + 2ab + b^2)}{12}$$

$$= \frac{b^2 - 2ab + a^2}{12}$$

$$\boldsymbol{Var(x) = \frac{(b-a)^2}{12}}$$

Ejercicio 58

En una práctica de precisión aérea se deja caer una bomba a lo largo de la línea de un kilómetro de longitud, el blanco se encuentra en el punto medio de la línea. El blanco se destruirá si la bomba cae a una distancia menor que 75 metros del centro. Calcule la probabilidad de que el blanco se destruya si la bomba cae al azar a lo largo de la línea.

X= “longitud de la práctica aérea”, x ϵ [0, 1000]

1ra Forma:

$$P(\text{Blanco sea destruido}) = P(425 \leq X \leq 575)$$

$$P(\text{Blanco sea destruido}) = \int_{425}^{575} \frac{1}{1000}\, dx$$

$$P(\text{Blanco sea destruido}) = \frac{1}{1000} x|_{425}^{575}$$

$$P(\text{Blanco sea destruido}) = \frac{1}{1000}(575 - 425)$$

$$P(\text{Blanco sea destruido}) = 0.15$$

2da Forma:

$$P(425 \leq X \leq 575) = F(575) - F(425)$$

Su función de distribución es:

$$F(x) = \begin{cases} 0, & si\ x < 0 \\ \dfrac{x}{1000}, & si\ x \in [a, b] \\ 1, & si\ x > 1000 \end{cases}$$

$$P(425 \leq X \leq 575) = \frac{575}{1000} - \frac{425}{1000}$$

$$P(425 \leq X \leq 575) = 0.15$$

Ejercicio 59

Una variable aleatoria X sigue una distribución uniforme sobre [-2,3].

a) Calcular Pr(X=1)

b) Pr(X<1.3)

c) Pr(|X|<1.5)

$$F(x) = \begin{cases} 0, & si\ x < -2 \\ \dfrac{x+2}{5}, & si\ x \in [-2,3] \\ 1, & si\ x > 3 \end{cases}$$

Parte a:

La probabilidad P(X=1) = 0, ya que la probabilidad en un punto es cero.

Parte b:

$$P(X < 1.3) = P(-2 < X < 1.3) = F(1.3) - F(-2) = \frac{3.3}{5} - 0 = 0.66$$

Parte c:

$$P(|X| < 1.5) = F(1.5) - F(-1.5) = \frac{3.5}{5} - \frac{0.5}{5} = 0.6$$

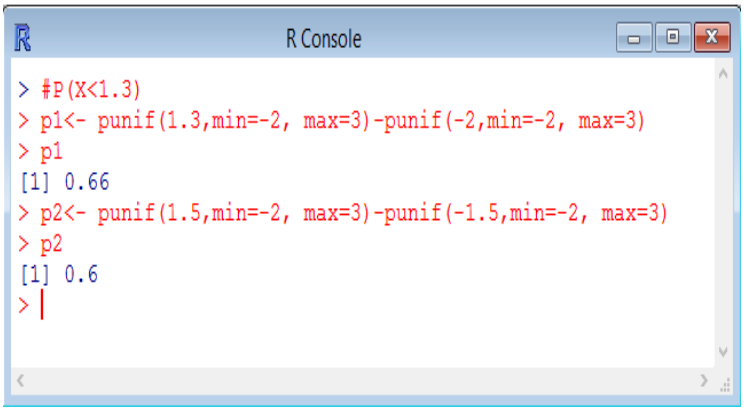


Ilustración 16 Tips R 30. Distribución Uniforme.

5.2.2 Distribución Exponencial

Sea X una variable aleatoria continua, se dice que X sigue una distribución exponencial de parámetro λ, con función de densidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \; ; x \geq 0 \\ 0; \; en \; otro \; caso \end{cases}$$

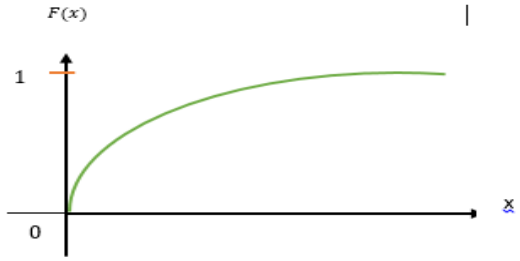
con función en distribución

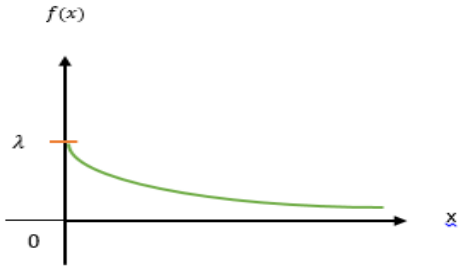
$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} \; ; \quad x \geq 0 \\ 0 \; ; \; en \; otro \; caso \end{cases}$$

Propiedades

- E(X) = $\frac{1}{\lambda}$
- Var(X) = $\frac{1}{\lambda^2}$
- X $\sim \varepsilon(\lambda)$

Las gráficas de las funciones de distribución y densidad





Ejercicios

Calcular la esperanza de la distribución Exponencial.

$$\boldsymbol{E(x)} = \int_0^{\infty} \mathrm{x}\lambda e^{-\lambda \mathrm{x}} dx$$

$$= \lambda \lim_{b\to\infty} \int_0^b \mathrm{x}\, e^{-\lambda \mathrm{x}} dx \quad (0)$$

Mediante integración por partes

$u = x$

$du = dx$ (1)

y $dv = e^{-\lambda \mathrm{x}} dx$ (2)

Cambio de variable

$m = -\lambda \mathrm{x}$

$\frac{dm}{dx} = -\lambda$

$-\frac{1}{\lambda} dm = dx$ (3) en (2)

$$v = -\frac{1}{\lambda} \int e^m dm$$

$v = -\frac{1}{\lambda} e^{-\lambda x}$ (4)

(1) Y (4) en (0)

$$= \lambda \lim_{b\to\infty} \left[-\frac{x}{\lambda} e^{-\lambda x} + \frac{1}{\lambda} \int e^{-\lambda x}\, dx \right]$$

$$= \lambda \lim_{b\to\infty}\left[-\frac{xe^{-\lambda x}}{\lambda} - \frac{1}{\lambda^2}e^{-\lambda x}\right]|_0^b$$

$$= \lambda \lim_{b\to\infty}\left[\left(-\frac{b}{\lambda}e^{-b\lambda} - \frac{1}{\lambda^2}e^{-b\lambda}\right) - \left(-\frac{e^{-\lambda(0)}}{\lambda^2}\right)\right]$$

$$= \lambda \lim_{b\to\infty}\left[-\frac{b}{\lambda}e^{-b\lambda} - \frac{1}{\lambda^2}e^{-b\lambda} + \frac{1}{\lambda^2}\right]$$

$$= \lambda[-\frac{b}{\lambda}\lim_{b\to\infty}\frac{1}{e^{b\lambda}} - \frac{1}{\lambda^2}\lim_{b\to\infty}\frac{1}{e^{b\lambda}} + \frac{1}{\lambda^2}] \text{ ; pero } \lim_{b\to\infty}\frac{1}{e^{b\lambda}} = 0$$

$$= \lambda[0 + 0 + \frac{1}{\lambda^2}]$$

$$= \lambda\left[\frac{1}{\lambda^2}\right]$$

$$= \frac{1}{\lambda}$$

Concluimos $E(X) = \frac{1}{\lambda}$

Ejercicio

Una variable aleatoria está distribuida según una ley exponencial del parámetro λ=3.

a) Hallar P (0.13< X <0.7)

Desarrollo

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

$$f(x) = \begin{cases} 0, & x < 0 \\ 3e^{-3x}, & x \geq 0 \end{cases}$$

P (0.13<X <0.7)

$$\int_{0.13}^{0.7} 3e^{-3x}\,dx$$

$$= -\frac{1}{3}\,3\,(e^{-3x})|_{0.13}^{0.7}$$

$$= \left[-e^{-3(0.7)} + e^{-3(0.13)}\right]$$

$$= -0.12245 + 0.677$$

$$= 0.55$$

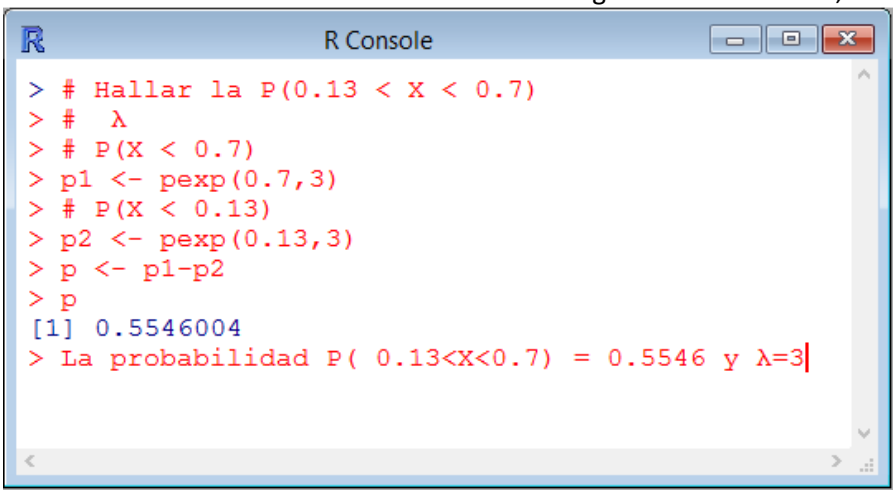


Ilustración 17Tips R 31. Distribución Exponencial.

b) Determine su esperanza

Desarrollo

$E(x) = \int_0^{\infty} 3e^{-3x}\, x\, dx$

$= 3 \lim_{b\to\infty} \int_0^{b} e^{-3x}\, x\, dx$

u=x dv= $\int e^{-3x}\, dx$

du=dx v=$-\frac{e^{-3x}}{3}$

$3 \lim_{b\to\infty} \left[x(\frac{-e^{-3x}}{3}) + \frac{1}{3}\int_0^{b} e^{-3x} dx\right]$

$= 3 \lim_{b\to\infty} \left[-x\frac{e^{-3x}}{3}\right] + \frac{1}{3}\left[-\frac{e^{-3x}}{3}\right] |_0^{b}$

$=3 \lim_{b\to\infty} \left[-b\frac{e^{-3b}}{3} - \frac{e^{-3b}}{9} + \frac{e^{0}}{9}\right]$

$=3\left[\lim_{b\to\infty} -b\frac{e^{-3b}}{3} - \lim_{b\to\infty}\frac{e^{-3b}}{9} + \lim_{b\to\infty} 1/9\right]$

$=3\left[-1/3 \lim_{b\to\infty}\frac{1}{e^{3b}} - 1/9 \lim_{b\to\infty}\frac{1}{e^{3b}} + 1/9\right]$

$=3\left[-\frac{1}{3}(0) - \frac{1}{9}(0) + 1/9\right]$

$=\frac{1}{3}$

Ejemplos 58

El tiempo durante el cual las baterías para teléfono celular trabajan en forma efectiva hasta que fallan se distribuye según un modelo exponencial, con un tiempo promedio de falla de 500h.

a) Calcular la probabilidad de que una batería funcione por más de 600h.

Desarrollo

$E(x) = 500\ h$

X= "tiempo de duración de batería"

$$E(x) = \frac{1}{\lambda} = 5000$$

$$\lambda = \frac{1}{500}$$

$$P(x \geq 600) = \int_{600}^{\infty} \lambda e^{-\lambda x}\ dx$$

$$= \int_{600}^{\infty} \frac{1}{500} e^{-\frac{\lambda}{500}x}\ dx$$

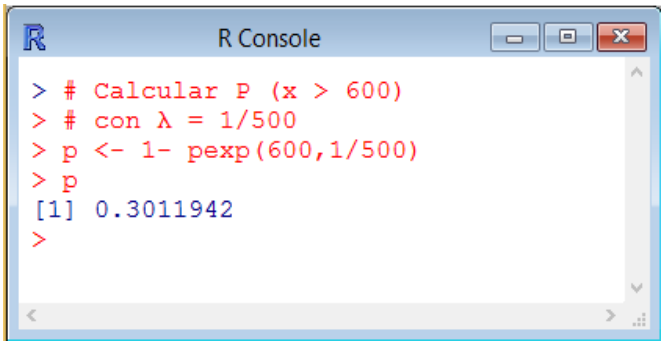
$$= \frac{1}{500}\left[-500e^{-\frac{\lambda}{500}x}\right]|_{600}^{\infty}$$

$$= -e^{-\frac{\infty}{500}} + +e^{\frac{-600}{500}}$$

$$= -\frac{1}{e^{\frac{\infty}{500}}} + e^{-\frac{6}{5}}$$

$$= 0 + e^{-\frac{6}{5}}$$

=0.3011



b) Si la batería ya trabajo 350h, ¿cuál es la probabilidad de que trabaje más de 300h adicionales?.

$P(X \geq 650 \quad \backslash X > 350)$

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

$$= \frac{P(X \geq 650 \cap X > 350)}{P(x \geq 350)}$$

$$= \frac{P(X \geq 650)}{P(X \geq 350)}$$

$$= \frac{1 - P(X \leq 650)}{1 - P(X \leq 350)}$$

$$= \frac{1 - P(0 \leq X \leq 650)}{1 - P(0 \leq X \leq 350)}$$

$$= \frac{1 - [F(650) - F(0)]}{1 - [F(350) - F(0)]}$$

Su función de distribución es

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\frac{1}{500}x} ; & x \geq 0 \\ 0 \ ; \ en\ otro\ caso \end{cases}$$

$$= \frac{1 - \left(1 - e^{\frac{-650}{500}}\right) + \left(1 - e^{\frac{-0}{500}}\right)}{1 - \left(1 - e^{\frac{-350}{500}}\right) + \left(1 - e^{\frac{-0}{500}}\right)}$$

$$\frac{1 - \left(1 - e^{\frac{-650}{500}}\right)}{1 - \left(1 - e^{\frac{-350}{500}}\right)}$$

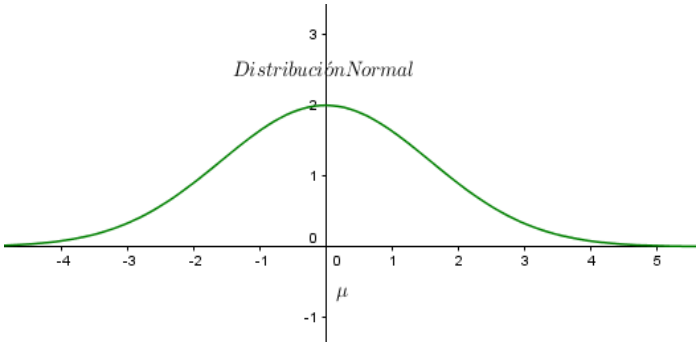
0.5488

5.2.3 Distribución Normal

Sea X una variable aleatoria continúa, se dice que X sigue una distribución normal con media μ y varianza σ^2, notado $X \sim N(\mu, \sigma^2)$.

La función de densidad de una distribución Normal es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{-(x-\mu)^2}{2\sigma^2}} ; \ \forall x \in \mathbb{R}$$

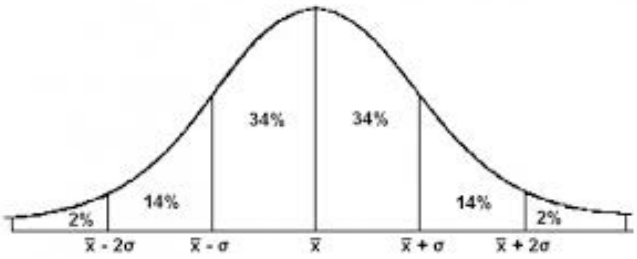


Y la función de distribución

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2\sigma^2}(t-u)^2}\,dt$$

Propiedades:

1.- La distribución Normal es simétrica, las áreas son iguales a los lados del eje de simetría.



2.- El eje de simetría es igual a la media, moda y mediana $(Mo = \mu = Md)$.

3.-- $E(X) = \mu$

4.- $Var(X) = \sigma^2$

5.- Si $Z \sim N(0,1)$; Z sigue una distribución normal estándar.

5.- Z es una variable centrada $(X - \mu)$.

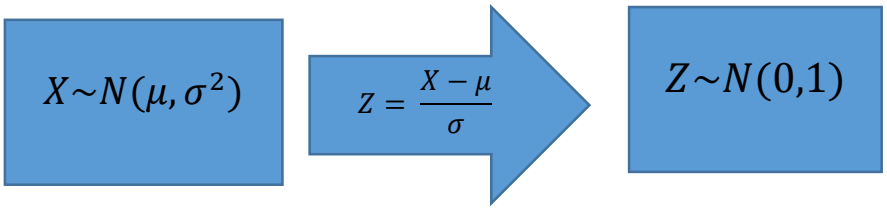
6.- Z es una variable normalizada $\frac{X-\mu}{\sigma}$

7.- Z es una variable con la cual podemos realizar comparaciones entre datos de diferentes muestras, sin estar necesariamente en las mismas unidades.

8.- El área comprendido entre $[\bar{x} - \sigma;\ \bar{x} + \sigma] = 68\%$

9.- El área comprendido entre $[\bar{x} - 2\sigma;\ \bar{x} + 2\sigma] = 96\ \%$

Relación entre X y Z



Ejercicio:

El perímetro craneal de los hombres medidos en centímetros en una ciudad, es una variable aleatoria con μ=60 cm y σ^2= 4.

a) Qué porcentaje de los hombres tiene un perímetro craneal entre 57 y 64cm.

Desarrollo

X = "Perímetro craneal de los hombres"

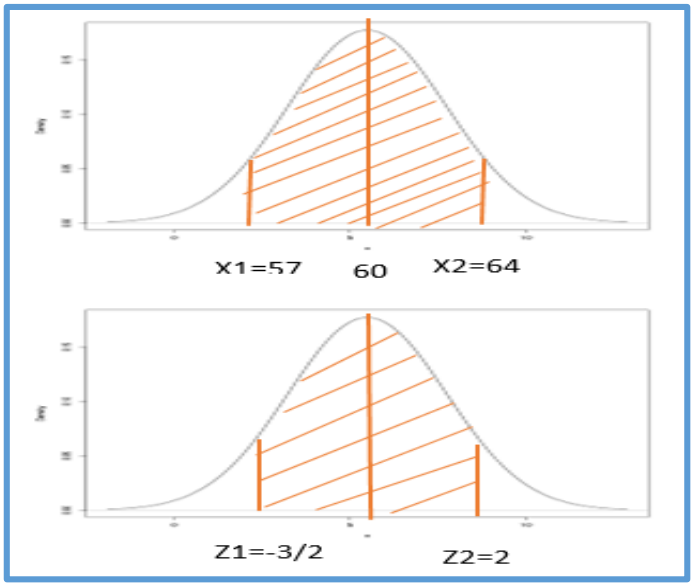
$X \sim N(60, 4)$

$\sigma^2 = 4$

$\sqrt[2]{\sigma} = \sqrt{4}$

$\sigma = 2$

$$P(57 \leq x \leq 64)$$



$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{57 - 60}{2} = -\frac{3}{2}$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{64 - 60}{2} = 2$$

$$P(57 \leq X \leq 64) = P\left(-\frac{3}{2} \leq Z \leq 2\right)$$

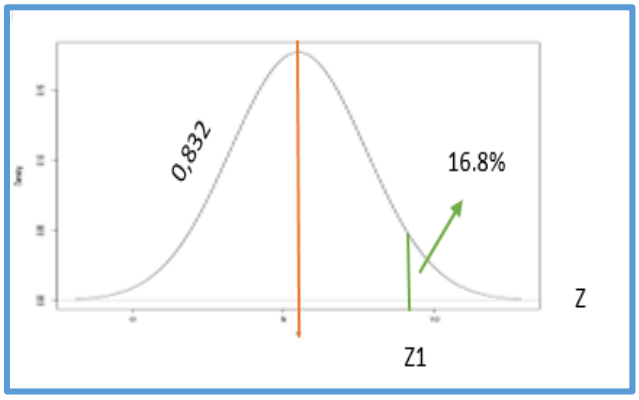
= P (Z≤ 2) - P (Z≤ −1.5)

= 0.9752-0.0668

= 0.9084

b) Qué perímetro craneal debe tener un hombre para que el 16.8% de sus paisanos tengan más cabeza que él.

Desarrollo

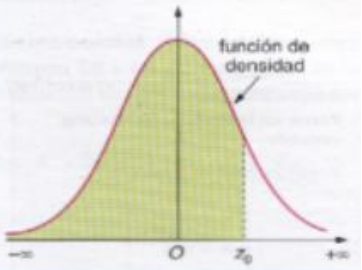


La expresión siguiente indica la probabilidad de las X (perímetros craneales) mayores a un x_1 (valor especifico) sea igual a un 16.8%.

$$P(X \geq x_1) = 0.168$$

TABLA DE LA DISTRIBUCIÓN NORMAL $N(0, 1)$

$P(Z \leq z_0) = \{$ área del recinto coloreado $\}$



z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9646	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

En la tabla se puede observar el valor de Z correspondiente al 83.2%

$Z_1 = 0.96$ (1)

$Z_1 = \frac{x_1 - \mu}{\sigma}$ (2)

Resolviendo (1) y (2)

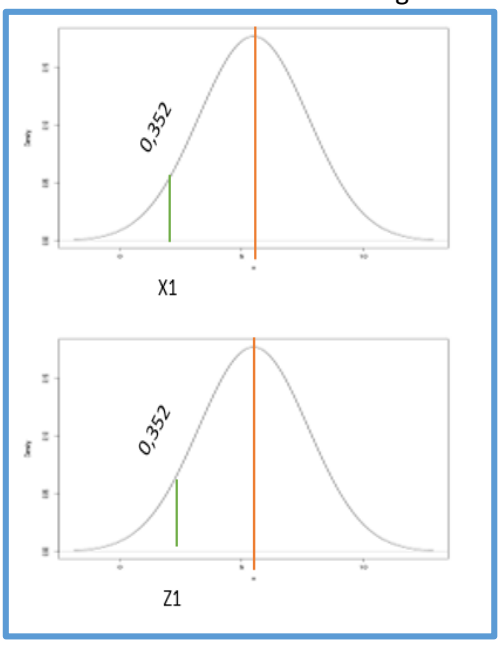
$x_1 = \sigma Z_1 + \mu$

$x_1 = 2(0.96) + 60$

$x_1 = 61.92$

c) Y cuanto para que el 35.2% tenga menos.

Desarrollo



La ecuación siguiente indica, la probabilidad de las X (perímetros craneales) menores a uno dado x_1, sea igual al 35.2 %.

$P(X < x_1) = 0.352$

$Z_1 = -0.38$ (1)

$z_1 = \frac{x_1 - \mu}{\sigma}$ (2)

Resolviendo (1) y (2)

$x_1 = \sigma z_1 + \mu$

$x_1 = -0.38(2) + 60$

$x_1 = 59.24$

Ejercicio

Los errores de medición de peso de una balanza obedecen a una ley normal con desviación estándar 20 mg y esperanza 0 mg. Halle la probabilidad de que tres mediciones independientes, el error de por lo menos una de ellas no sea mayor, en valor absoluto, que 4mg.

X = ”errores de medición”

$X \sim N(0, 20^2)$

En general el error de medición es

$\mu - 4 \leq x \leq \mu + 4$

$-4 \leq x - \mu \leq 4$

$|x - \mu| \leq 4$

$P(|x - \mu| \leq 4)$

$$P\left(\frac{-4}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{4}{\sigma}\right) = \mathrm{P}\left(\frac{-4}{\sigma} \leq \mathrm{Z} \leq \frac{4}{\sigma}\right) = \mathrm{P}(\frac{-4}{20} \leq \mathrm{Z} \leq \frac{4}{20})$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{-4-0}{20} = -0.2$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{4-0}{20} = 0.2$$

$$P(-4 \leq x \leq 4) = P(-0.2 \leq Z \leq 0.2)$$
$$= \mathrm{P}\,(\mathrm{Z} \leq 0.2) - \mathrm{P}\,(\mathrm{Z} \leq -0.2)$$
$$= 0.5793 - 0.4207$$
$$= 0.1586$$

La probabilidad que una medición no exceda 4mg es 0.1586.

La probabilidad de que el error de por lo menos una de ellas no sea mayor, en valor absoluto, que 4mg es:

P(por lo menos una de ellas)= P(un error no sea mayor o dos errores no sean mayor o tres errores no sean mayor)= P(un error no sea mayor)+ P(dos errores no sean mayor) +P(tres errores no sean mayor). (1)

Aplicando la probabilidad Binomial a (1), se tiene

P(un error no sea mayor)= 3(0.1586)(0.8414)(0.8414)= 0.3368

P(dos errores no sean mayor)= 3(0.1586)(0.1586)(0.8414)=0.06349

P(tres errores no sean mayor)= (0.1586)(0.1586)(0.1586)=0.00398

concluyendo la P(por lo menos una de ellas)= 0.4043

Ejercicio

Se toman dos exámenes sobre 100 puntos en el primero se obtuvo una media de 80 y desviación estándar de 4 y en el segundo una media de 65 y desviación estándar de 5.

Un estudiante saco 84 en el primer examen y 75 en el segundo. ¿En cuál de los dos resultados, obtuvo mejor desenvolvimiento?

Desarrollo

X_1~N (80,16)

X_2~N (65,25)

$$P(X < 84) = \mathrm{P}\left(\mathrm{Z} < \frac{84-80}{4}\right) = P(Z < 1) = 0.84$$

$$P(X < 75) = \mathrm{P}\left(\mathrm{Z} < \frac{75-65}{5}\right) = P(Z < 2) = 0.97$$

Obtuvo mejor desenvolvimiento cuando obtuvo la nota de 75 ya que el 97% del curso obtuvo notas menores a él.

Definición de función Gamma

Se define la función Gamma de x y x>0 como

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

Ejercicio

$$\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}dt = \int_0^\infty e^{-t}dt = -e^{-t}|_0^\infty = -(e^{-\infty} - e^{-0}) = -(0-1) = 1$$

Ejercicio

Demuestre $\Gamma(x+1) = x\Gamma(x)$

$$\Gamma(x+1) = \int_0^\infty t^{x}e^{-t}dt$$

Integrando por partes

$$u = t^x$$

$$du = xt^{x-1}dt$$

$$dv = e^{-t}dt$$

$$v = -e^{-t}$$

$$\Gamma(x+1) = \lim_{b\to\infty}(-t^x e^{-t}|_0^b + x\int_0^\infty e^{-t}t^{x-1}dt) \quad (1)$$

$\lim_{b\to\infty} -\frac{t^x}{e^t}|_0^b$, el valor del límite es cero, aplicando x veces L'Hopital se tiene: en el numerador un número y en el denominador se tiene la misma expresión, aplicando el límite se tiene cero. (2)

(2) en (1)

$$\Gamma(x+1) = x\Gamma(x)$$

Ejercicio

Demuestre: $\Gamma(n+1) = n!$

$$\Gamma(n+1) = n\Gamma(n)$$
$$= n(n-1)\Gamma(n-1)$$
$$= n\,(n-1)(n-2)\Gamma(n-2)$$
$$= n\,(n-1)(n-2)(n-3)\Gamma(n-3)$$
$$= \cdots$$
$$= n\,(n-1)(n-2)(n-3)\ldots(1\,)\Gamma\big(n-(n-1)\big) = n(n-1)(n-2)\ldots 1\Gamma(1) = \mathrm{n!}$$

Ejercicio

Demuestre:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{\frac{1}{2}-1}e^{-t}dt = \int_0^\infty t^{\frac{-1}{2}}e^{-t}dt$$

Cambio de variable

$$t = z^2; z = \sqrt{t}; dt = 2zdz$$

$$= \int_0^\infty (z^2)^{\frac{-1}{2}}e^{-z^2}2zdz = 2\int_0^\infty z^{-1}e^{-z^2}zdz = 2\int_0^\infty e^{-z^2}dz$$

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty e^{-z^2}dz \ (1)$$

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty e^{-w^2}dw \ (2)$$

Multiplicando

$$(\Gamma(1/2))^2 = (2\int_0^\infty e^{-z^2}dz)(\,2\int_0^\infty e^{-w^2}dw)$$

$$(\Gamma(1/2))^2 = \,4\int_0^\infty\int_0^\infty e^{-w^2}e^{-z^2}dzdw = 4\int_0^\infty\int_0^\infty e^{-(z^2+w^2)}dzdw$$

Cambio a coordenadas polares, el área de integración es el primer cuadrante

$$r^2 = z^2 + w^2 \;\; y \;\; jacobiano \;\; = \;\; r$$

$$(\Gamma(1/2))^2 = \,4\int_0^{\frac{\pi}{2}}\int_0^\infty e^{-r^2}rdrd\theta$$

Pero

$$\int \; e^{f(x)}f'(x)dx = \; e^{f(x)} + c$$

$$\int_0^\infty e^{-r^2}rdr = \frac{-1}{2}\int_0^\infty e^{-r^2}(-2r)dr = \frac{-1}{2}e^{-r^2}|_0^\infty = \frac{1}{2}$$

Reemplazando en la ecuación anterior

$$(\Gamma(1/2))^2 = \,4\int_0^{\frac{\pi}{2}}\frac{1}{2}d\theta = \,2\int_0^{\frac{\pi}{2}}d\theta = 2\theta|_0^{\frac{\pi}{2}} = \pi$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

Funciones gamma incompletas.

Dadas las funciones Gamma incompletas

$$\Upsilon(x,\alpha) = \int_0^\alpha e^{-t}t^{x-1}dt$$

$$\Gamma(x,\alpha) = \int_\alpha^\infty e^{-t}t^{x-1}dt$$

$$\Gamma(x) = \Upsilon(x,\alpha) + \Gamma(x,\alpha)$$

Distribución Gamma

La distribución gamma (generalización de una exponencial) es adecuada para modelizar el comportamiento de variables aleatorias continuas con asimetría positiva, la misma que es dependiente de dos parámetros α y β.

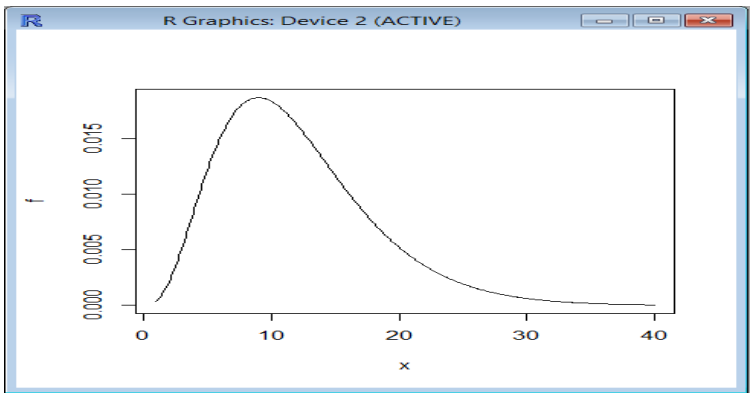
El primer **parámetro α** sitúa la máxima intensidad de probabilidad y por este motivo en algunas fuentes se denomina **"la forma"** de la distribución: cuando se toman valores próximos a cero aparece entonces un dibujo muy similar al de la distribución exponencial. Cuando se toman valores más grandes de α el centro de la distribución se desplaza a la derecha y va apareciendo la forma de una campana de Gauss con asimetría positiva.

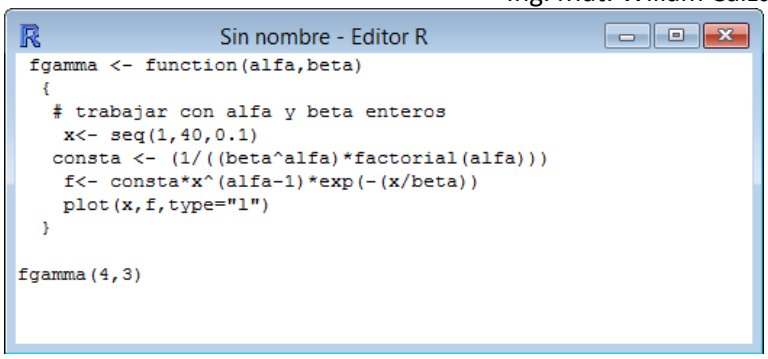
El **parámetro β** determina la forma o alcance de esta asimetría positiva desplazando la densidad de probabilidad en la cola de la derecha. Para valores elevados de β la distribución acumula más densidad de probabilidad en el extremo derecho de la cola. Al dispersar la probabilidad la altura máxima de densidad de probabilidad se va reduciendo; de aquí que se le denomine **"escala".** Valores más pequeños de β conducen a una figura más simétrica y concentrada, con un pico de densidad de probabilidad más elevado.

Una forma de interpretar β es el "tiempo promedio entre ocurrencia de un suceso", relacionándose con el parámetro de la Poisson como β=1/λ.

La función de densidad gamma está definida

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} & ; \ x, \alpha, \beta > 0 \\ 0, \quad en\ otro\ caso \end{cases}$$





La función f(x) es función de densidad ya que:

- f(x) ≥ 0
- si $h = \frac{1}{\beta}$

$$\int_0^\infty f(x)dx = \frac{h^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-xh} dx$$

Sea

$$u = hx;\ \frac{du}{dx} = h\ ;\ \frac{du}{h} = dx$$

Reemplazando en la ecuación anterior

$$\int_0^\infty f(x)dx = \frac{h^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{u}{h}\right)^{\alpha-1} e^{-u} \frac{du}{h}$$

$$= \frac{h^\alpha}{\Gamma(\alpha)h^\alpha} \int_0^\infty u^{\alpha-1}\, e^{-u} du = \frac{h^\alpha}{\Gamma(\alpha)h^\alpha} \Gamma(\alpha) = 1$$

Concluyendo que f(x) es función de densidad.

Función de distribución Gamma

$$\boldsymbol{F(x,\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} dy}$$

Ejercicio

Demuestre: $E(X) = \alpha\beta$

$$E(X) = \int_0^\infty x\, f(x)dx = \frac{h^\alpha}{\Gamma(\alpha)} \int_0^\infty x\, x^{\alpha-1} e^{-xh} dx = \frac{h^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-xh} dx$$

Con

$$u = hx;\ \frac{du}{dx} = h\ ;\ \frac{du}{h} = dx$$

$$= \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \frac{u}{h^{\alpha}}^{\alpha} e^{-u} \frac{du}{h} = \frac{h^{\alpha}\Gamma(\alpha+1)}{\Gamma(\alpha)h^{\alpha+1}} = \frac{h^{\alpha}\alpha\Gamma(\alpha)}{\Gamma(\alpha)h^{\alpha+1}} = \frac{\alpha}{h}$$

$$= \frac{\alpha}{h} = \frac{\alpha}{\frac{1}{\beta}} = \alpha\beta$$

Ejercicio

Demuestre $Var(X) = E(X^2) - (E(X))^2$

$$E(X^2) = \int_0^{\infty} x^2 f(x)dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^2 x^{\alpha-1} e^{-xh} dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1} e^{-xh} dx$$

$$= \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \frac{u}{h^{\alpha+1}}^{\alpha+1} e^{-u} \frac{du}{h} = \frac{h^{\alpha}\Gamma(\alpha+2)}{\Gamma(\alpha)h^{\alpha+2}} = \frac{h^{\alpha}(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)h^{\alpha}h^2}$$

$$= \frac{(\alpha+1)\alpha}{h^2}$$

$$Var(X) = \frac{(\alpha+1)\alpha}{(\frac{1}{\beta})^2} - (\alpha\beta)^2 = (\alpha+1)\alpha\beta^2 - (\alpha\beta)^2 = \alpha\beta^2$$

Ejercicio

En un estudio con ratas se usa una investigación de respuesta a la dosis para determinar el efecto de la dosis de un tóxico en su tiempo de sobrevivencia. El tóxico es uno que se descarga con frecuencia en la atmósfera desde el combustible de los aviones. Para cierta dosis del tóxico el estudio determina que el tiempo de sobrevivencia, en semanas, tiene una distribución gamma con α= 5 y β=10.¿Cuál es la probabilidad de que una rata no sobreviva más de 60 semanas?

Desarrollo:

Sea X =”Número semanales de sobrevivencia”

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}}$$

$$f(x) = \frac{1}{10^5\Gamma(5)} x^{5-1} e^{\frac{-x}{10}} = \frac{1}{10^5\Gamma(5)} x^4 e^{\frac{-x}{10}}$$

$$P(X \leq 60) = \int_0^{60} \frac{1}{10^5\Gamma(5)} x^4 e^{\frac{-x}{10}} \, dx$$

Cambio de variable

$$y = \frac{x}{10}; \quad \frac{dy}{dx} = \frac{1}{10}$$

$$P(X \leq 60) = \int_0^6 \frac{1}{10^5\Gamma(5)}(10y)^4 e^{-y}\, 10\, dy = \int_0^6 \frac{(10)^5}{10^5\Gamma(5)}(y)^4 e^{-y}\, dy$$

$$P(X \leq 60) = \int_0^6 \frac{1}{\Gamma(5)}(y)^4 e^{-y}\, dy \quad \text{(a)}$$

La expresión (a), tiene la forma de la función acumulada

FUNCION DE LA GAMMA INCOMPLETA

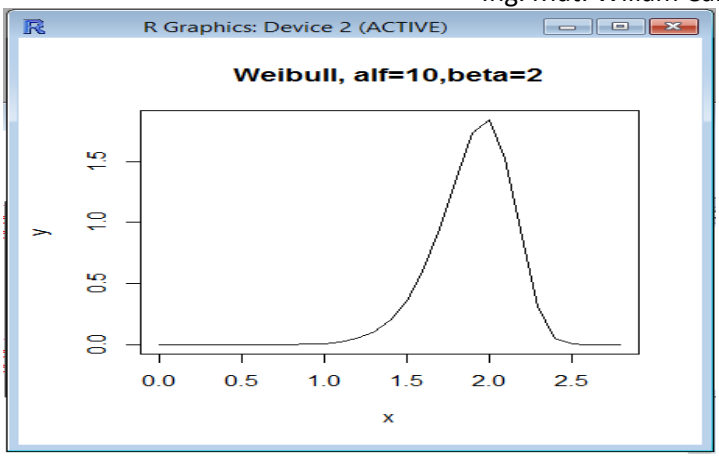
x	α									
	1	2	3	4	5	6	7	8	9	10
1	0,632	0,264	0,080	0,019	0,004	0,001	0,000	0,000	0,000	0,000
2	0,865	0,594	0,323	0,143	0,053	0,017	0,005	0,001	0,000	0,000
3	0,950	0,801	0,577	0,353	0,185	0,084	0,034	0,012	0,004	0,001
4	0,982	0,908	0,762	0,567	0,371	0,215	0,111	0,051	0,021	0,008
5	0,993	0,960	0,875	0,735	0,560	0,384	0,238	0,133	0,068	0,032
6	0,998	0,983	0,938	0,849	0,715	0,554	0,394	0,256	0,153	0,084
7	0,999	0,993	0,970	0,918	0,827	0,699	0,550	0,401	0,271	0,170

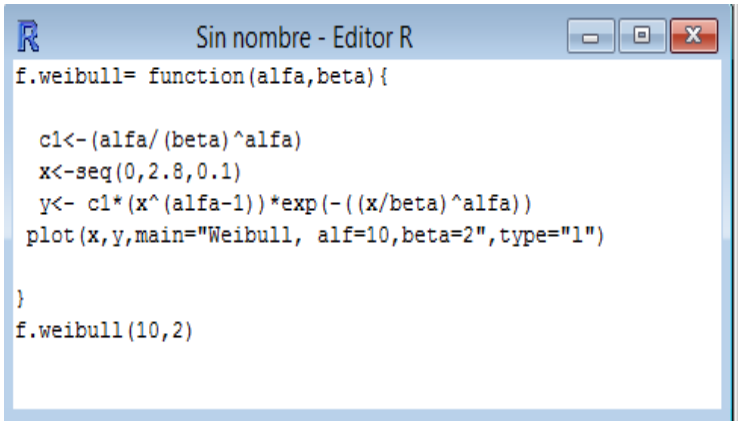
$$P(X \leq 60) = F(6{,}5) = 0.715$$

Función de distribución Weibull

Sea $X \sim weibull(\alpha, \lambda)$ y $\alpha > 0$, $\lambda > 0$, con función de densidad

$$f(x) = \begin{cases} \frac{\alpha}{\beta^\alpha}\, x^{\alpha-1} e^{-(\frac{x}{\beta})^\alpha}, & si\ x > 0 \\ 0, & si\ x \leq 0 \end{cases}$$





Y la función de distribución

$$F(x) = P(X \leq x) = \int_0^x \frac{\alpha}{\beta^{\alpha}}\, t^{\alpha-1} e^{-(\frac{t}{\beta})^{\alpha}} dt$$

$$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; x \geq 0$$

Propiedades

a) $E(X) = \beta\, \Gamma(1 + \frac{1}{\alpha})$
b) $Var(X) = \beta^2(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2(1 + \frac{1}{\alpha}))$

Ejercicio

La distribución de Weibull ha sido utilizada para modelar emisiones de varios contaminantes de motores. Sea X la v.a cantidad de emisiones de un cierto gas de un motor de cuatro tiempos de un tipo seleccionado aleatoriamente y suponga que X tiene distribución de Weibull con $\alpha = 2\ y\ \beta = 10$. ¿Calcule la probabilidad de que la cantidad de emisiones sea menor a 10?

$$P(X \leq 10) = 1 - e^{-\left(\frac{10}{10}\right)^2} = 1 - e^{-(1)^2} = 1 - e^{-1} = 1 - 0.36788 = 0.632$$

5.2.4. Teorema del Límite Central

Sea $X_1, X_2, \ldots, X_n$, variables aleatorias independientes con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ y que siguen una ley de probabilidad cualquiera no necesariamente la misma y

$Y = X_1 + X_2 + \cdots + X_n$, con $E(Y) = n\mu$ y $Var(Y) = n\sigma^2$, entonces:

$$Z = \frac{Y - E(Y)}{\sqrt{Var(Y)}} = \frac{Y - n\mu}{\sigma\sqrt{n}} \quad ; \quad \text{si } n \to \infty \quad Z \sim N(0,1^2)$$

Se puede utilizar cundo n ≥ 25.

Ejercicio

De la definición anterior demuestre $E(Y) = n\mu$ y $Var(Y) = n\sigma^2$.

Sea

$Y = X_1 + X_2 + X_3 + \cdots + X_n$ (1), calculando la esperanza matemática

$E(Y) = E(X_1 + X_2 + X_3 + \cdots + X_n)$

$E(Y) = E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_n)$

$E(Y) = \mu_1 + \mu_2 + \cdots + \mu_n$

$\boldsymbol{E(Y) = nu}$

En (1), calculando la varianza

$Var(Y) = Var(X_1 + X_2 + X_3 + \cdots + X_n)$

$Var(Y) = Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_n)$

$\mathbf{Var(Y) = \sigma^2 + \sigma^2 + \sigma^2 + \cdots + \sigma^2 = n\sigma^2}$

Ejercicio

El costo diario de operación de un autobús tiene un costo fijo de 30 dólares y un valor variable de 30 % de los ingresos. El ingreso tiene una distribución uniforme entre 50 y 250 dólares.

a) Calcular la probabilidad de que el costo de operar un autobús durante 31 días, supere los 2500 dólares.

b) Cuantos días de operación serán necesarios para que con una probabilidad del 0.95, el costo de operación sea al menos 2350 dólares.

Literal a:

X= "Ingreso diario de la operación del autobús"

$X \sim U[50,250]$

El costo de operación es C(X)= 30 + 0.3 X

$$C(X_1) = 30 + 0.3X_1$$

$$E(C(X_1)) = 30 + 0.3E(X_1)$$

$$E\big(C(X_1)\big) = 30 + 0.3\left(\frac{50+250}{2}\right) = 30 + 45 = 75$$

Lo que sería la misma esperanza para los 31 días.

La varianza Var(C(X))=0.09 Var(X)

$$Var\big(C(X_1)\big) = 0.09\, Var(X_1) = 0.09\, \frac{(250-50)^2}{12} = 300$$

Lo que sería la misma varianza para los 31 días.

Sea $Y = C_1 + C_2 + C_3 + \cdots + C_{31}$

$$E(Y) = E(C_1) + E(C_2) + E(C_3) + \cdots + E(C_{31}) = 31(75) = 2325$$

$$V(Y) = V(C_1) + V(C_2) + V(C_3) + \cdots + V(C_{31}) = 31(300) = 9300$$

$$P(C(x_1) + C(x_2) + C(x_3) + \cdots + C(x_{31}) > 2500)$$

$$P(Y > 2500) = P\left(\frac{Y - E(Y)}{\sqrt{Var(Y)}} > \frac{2500 - 2325}{\sqrt{9300}}\right) = P(Z > 1.814)$$

$P(Z > 1.81) = 1 - P(Z < 1.81) = 1 - 0.9649 = 0.0351$, lo que indica que de cada 100 veces en 3 veces el costo de operar sobrepasa los 2500 dólares.

Literal b:

$P(Y > 2350) = 0.95$

$Z_1 = \frac{2350-75n}{\sqrt{300}\sqrt{n}}$ (1)

$Z_1 = -1.65$, por tabla (2)

Igualando (1) y (2)

$2350 - 75n = -1.65(17.32)\sqrt{n}$

$2350 - 75n = -28.58\sqrt{n}$

$(2350 - 75n)^2 = \left(-28.58\sqrt{n}\right)^2$

$5522500 - 352500n + 5625n^2 = 816.8164n$

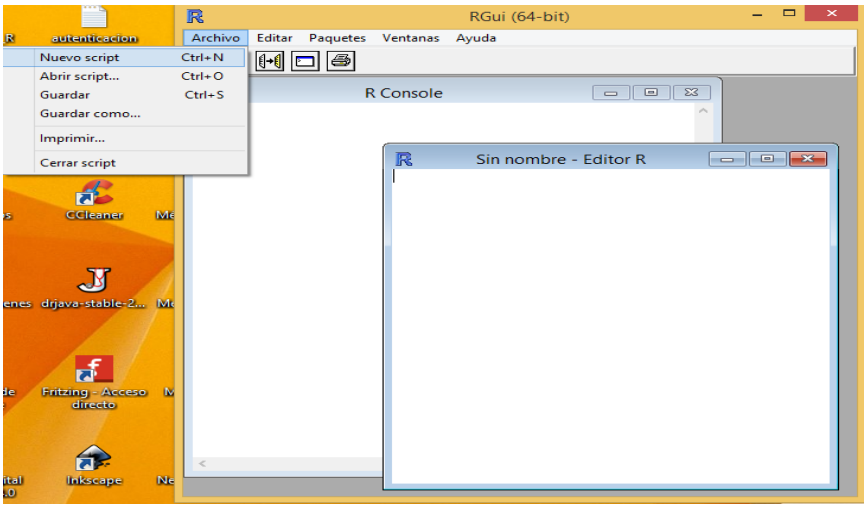
$5522500 - 353316.82n + 5625n^2 = 0$

$n_1 = 33.54$

$n_2 = 29{,}27$

Indica que con n= 34 días serán necesarios.

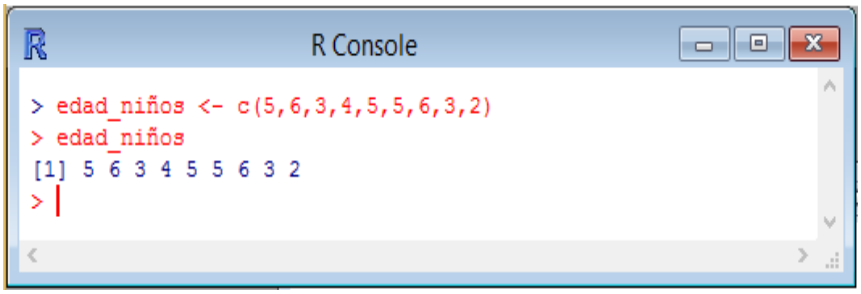
ANEXO 1: Programación con R.



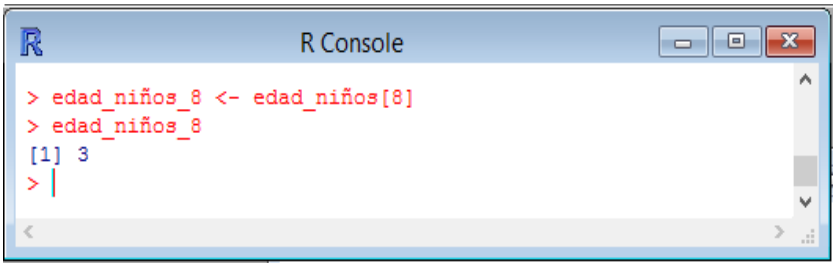
Inicialmente se debe abrir R, luego se hace click en "**Archivo"** y click en "**Nuevo script"**, en el script se puede probar los comandos o programar.

Manejo de datos

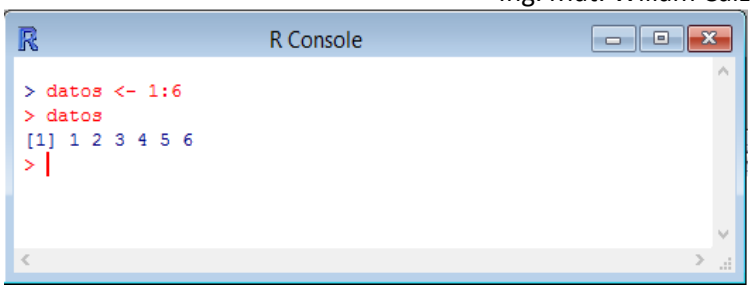
Para la creación de un vector de datos, se hace uso del comando **"c"**, para formar un vector de datos y realizar la asignación (**<-**) a la variable, mediante la expresión ***variable <- c(datos)***.



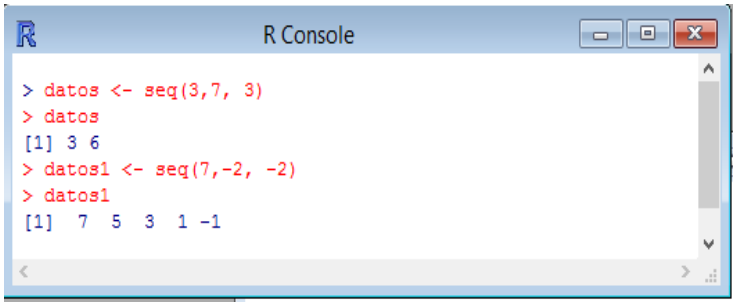
Extraer el dato que se encuentra en la posición 8 del vector **"edad_niños"**, se hace uso **"edad_niños[8]".**



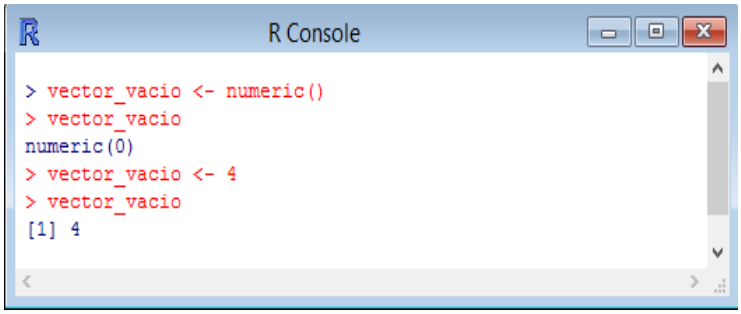
Crear un vector de datos del 1 al 6 con paso de 1, **"1:6"**.



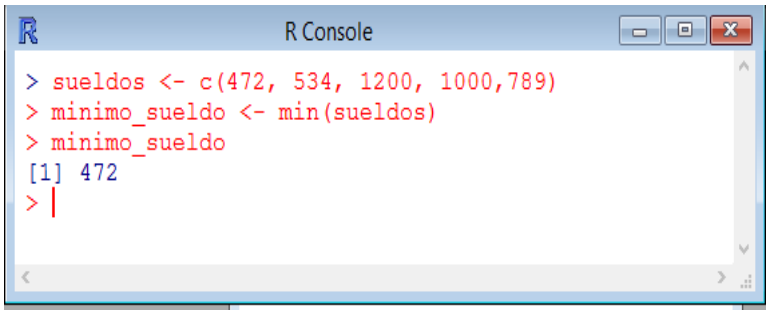
Crear un vector de 3 a 7 con paso de 3; crear un vector de 7 a -2 con pasos de -2.



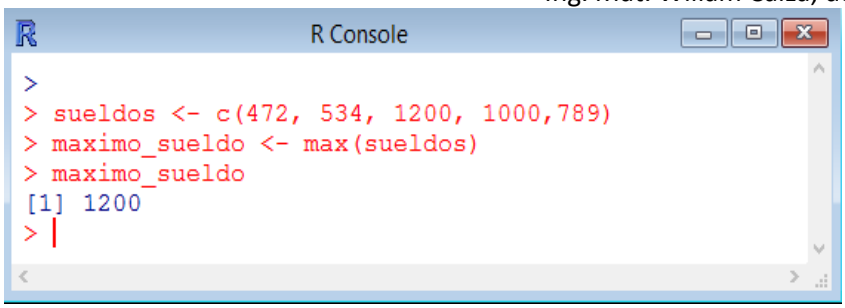
Crear un vector vacío **"numeric()"**, se lo utiliza para asignar datos durante la programación



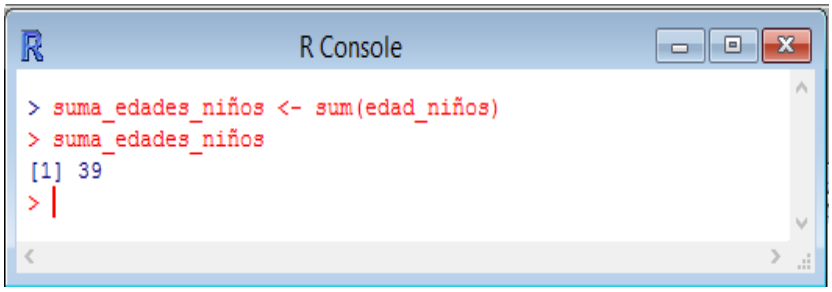
Encontrar el mínimo de los datos, se usa la función "**min(datos)".**



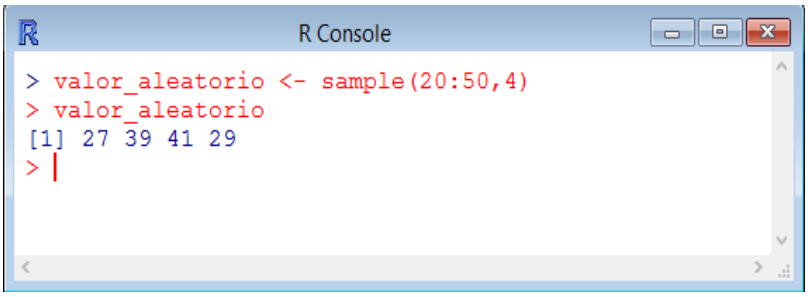
Encontrar el valor máximo de un conjunto de datos usamos el comando "**max(datos)".**



Utilizamos la función "**sum(datos)",** para obtener la suma los datos ingresados.



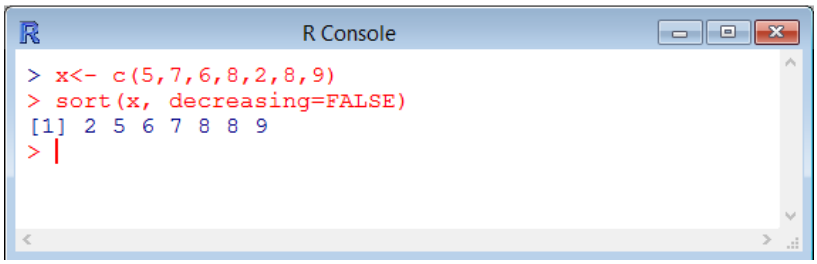
Elegir Aleatoriamente datos.



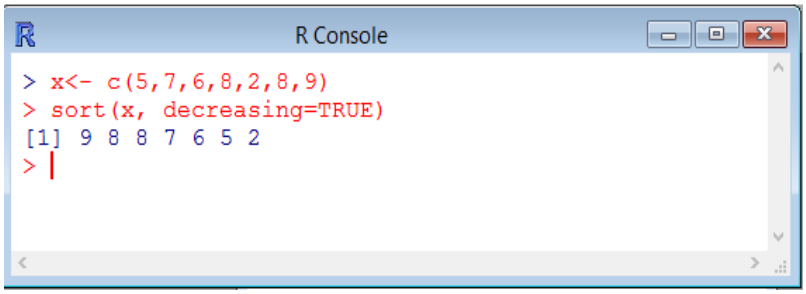
Utilizamos la función **sample(limites, número de términos aleatorios),**para obtener valores aleatorios entre los límites ingresados.

Ordenar datos.

Utilizamos la función **sort(datos, decreasing=FALSE)** para obtener datos en orden ascendente.



Utilizamos la función **sort(datos, decreasing=TRUE)** para obtener datos ordenados decrecientemente.



Permutaciones y Combinaciones

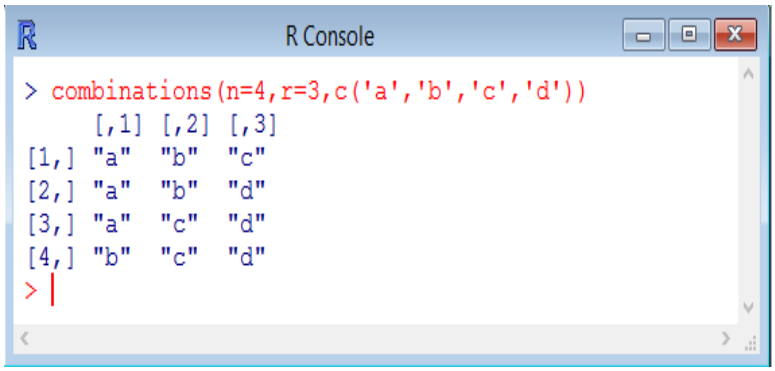
Para este tema necesitaremos instalar el paquete 'gtools', pueden ser instalados al digitar las siguientes expresiones:

install.packages('gtools')

library(gtools)

Combinaciones

Utilizando la función "**combinations(n=4, r=3, c(datos))"**, podemos encontrar las combinaciones de un conjunto de datos.



n=4, es el número de elementos.

r=3, es el número de elementos que debe tener la combinación o arreglo.

Permutaciones

Utilizando la función "**permutations(n=3,r=3,v=x,repeats.allowed=F**)", se encuentra el número de permutaciones de un conjunto.

n=3, es el número de elementos.

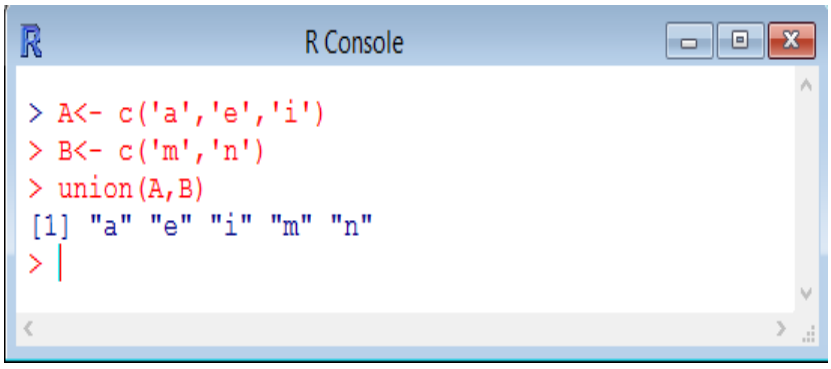
r=3, es el número de elementos que debe tener el arreglo o permutación.

Conjuntos

Se necesita haber instalado el paquete "gtools", puede ser instalados al digitar la siguiente expresion:

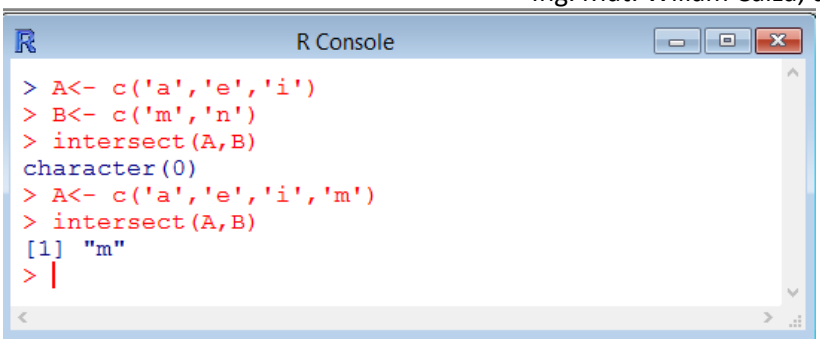
library(gtools)

Unión de conjuntos



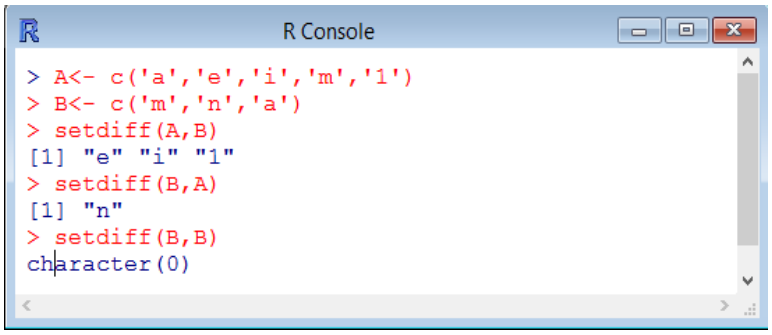
Utilizamos la función "**union (conjunto 1,conjunto2)" ,** obtendremos la unión entre los conjuntos especificados.

Intercesión de conjuntos



Utilizamos la función "**intersect(conjunto 1,conjunto2)",** obtendremos la intersección entre los conjuntos especificados.

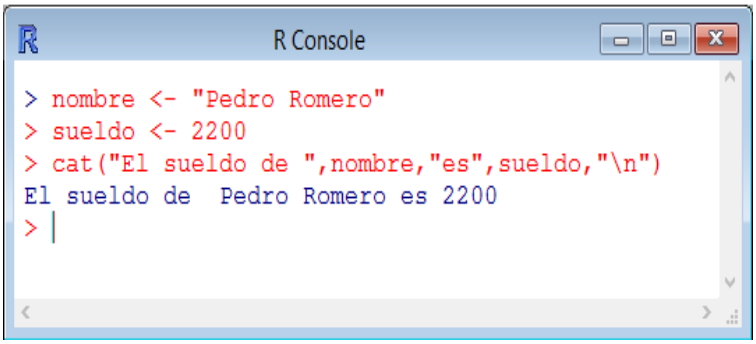
Diferencia de conjuntos.



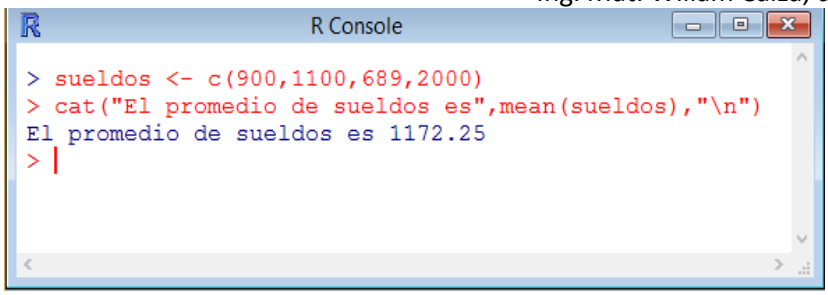
Utilizamos la función **setdiff(conjunto 1,conjunto2)** obtendremos la diferencia entre los conjuntos especificados.

Medidas de Tendencia Central.

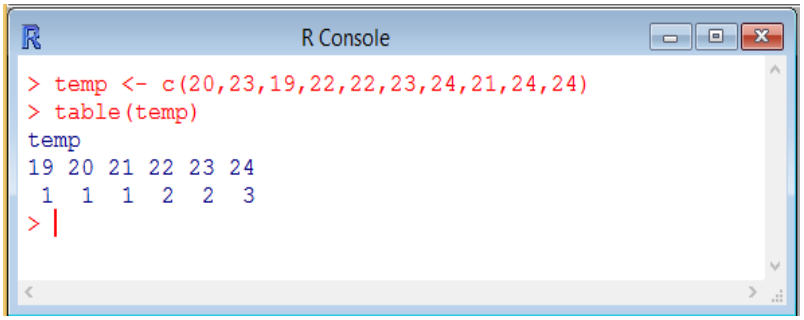
Se puede formatear la salida de información mediante la función "**cat**"



Utilizamos la función "**mean (datos)"** para calcular el promedio de los datos ingresados.

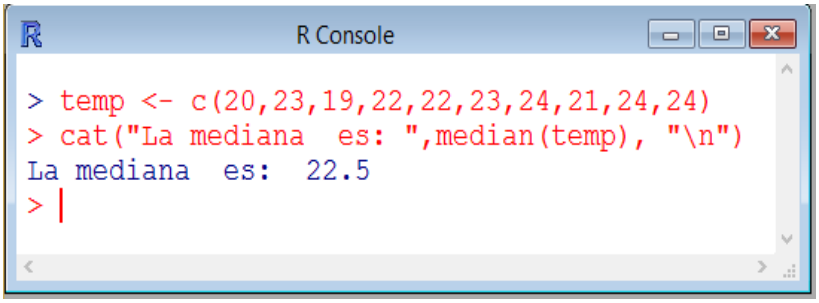


Frecuencia.



Utilizando la función "**Variable<-table(datos)**", se obtiene un vector (dato, frecuencia), en el ejemplo se puede verificar que 3 días se tiene una temperatura 24 grados.

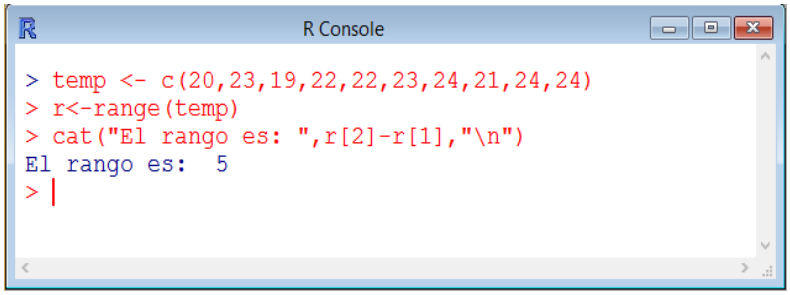
Mediana.



Con el comando "**Variable <-median(datos)",** se obtiene la mediana de los datos.

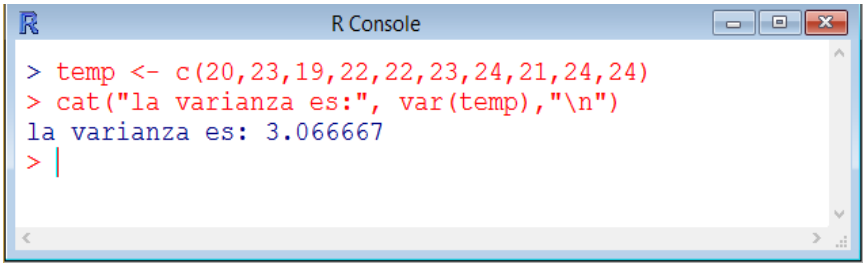
Medidas de Dispersión.

Rango.



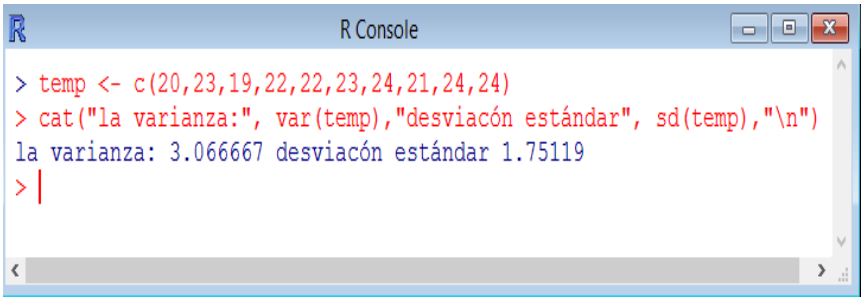
El comando **"range(datos)"** nos da como información el valor mínimo y el valor máximo de los datos.

Varianza.



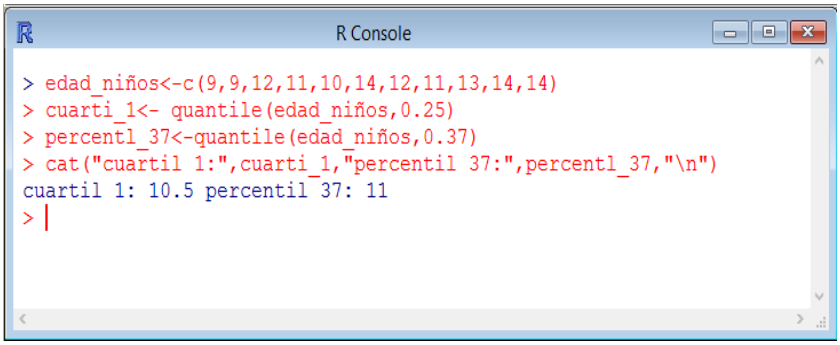
El comando " **var(datos)",** nos da la varianza de los datos.

Desviación Estándar.



El comando **"sd(datos)",** nos da la desviación estándar.

Medidas de posición

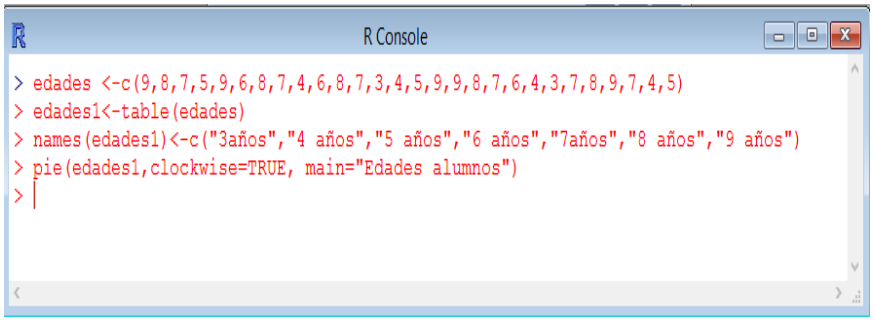


El comando "**quantile(datos, cuantil)",** nos da el valor del cuantil que se especifique, 0.25 será el primer cuartil , 0.1 será el primer decil.

Graficas

Diagrama de pastel

pie (datos, clockwise=sentido de las agujas del relog, main="Titulo del diagrama")



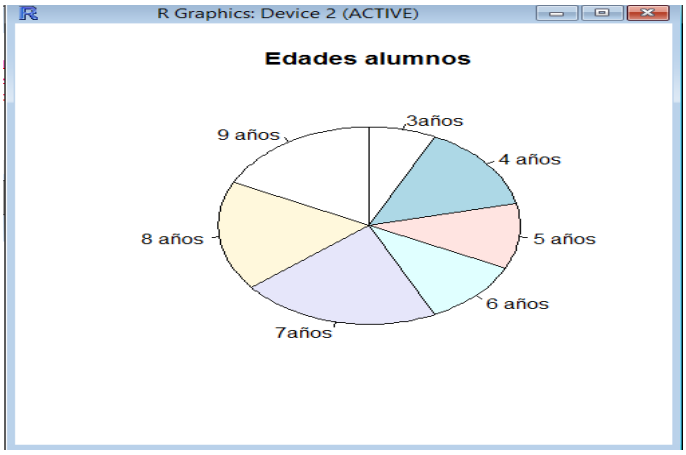
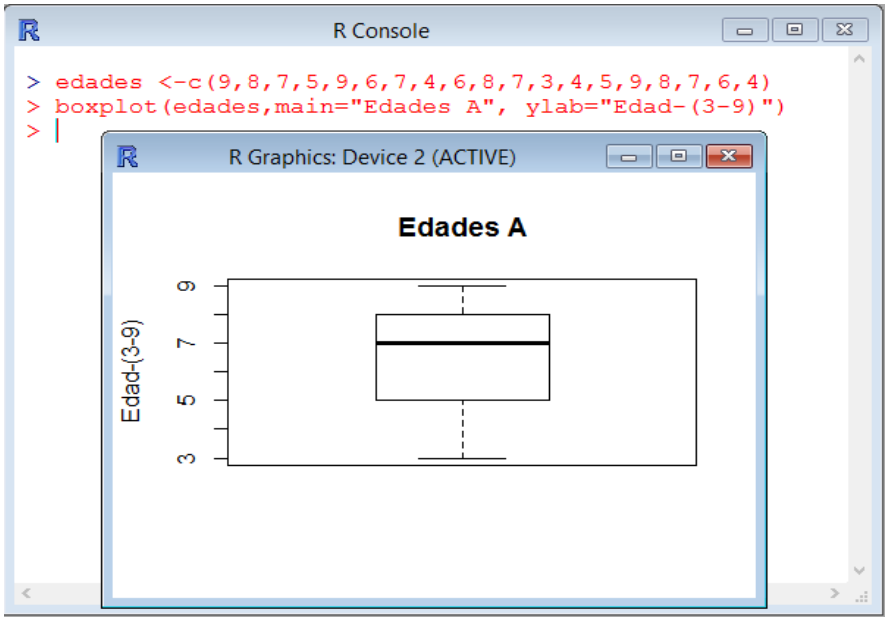


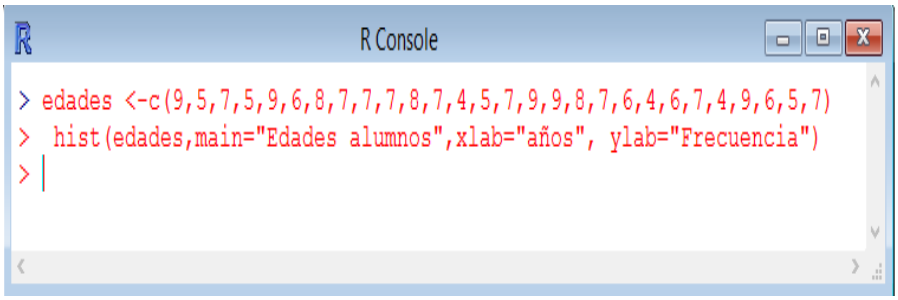
Diagrama de caja

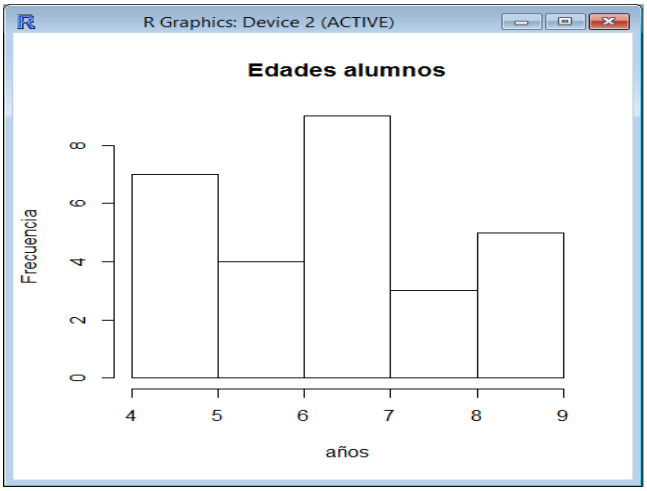
Mediante la función **"boxplot(datos, main,ylab)"**, grafica el diagrama de caja.



Histograma

Mediante la función **"hist(datos,main="titulo",xlab,ylab)"**, gráfica el histograma de los datos.



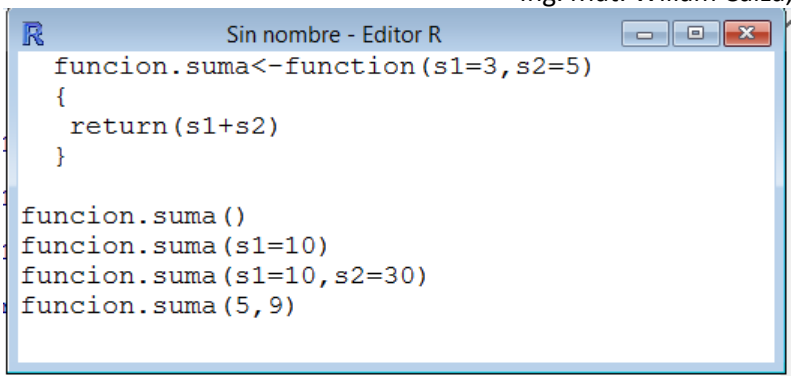


Creación de funciones en R

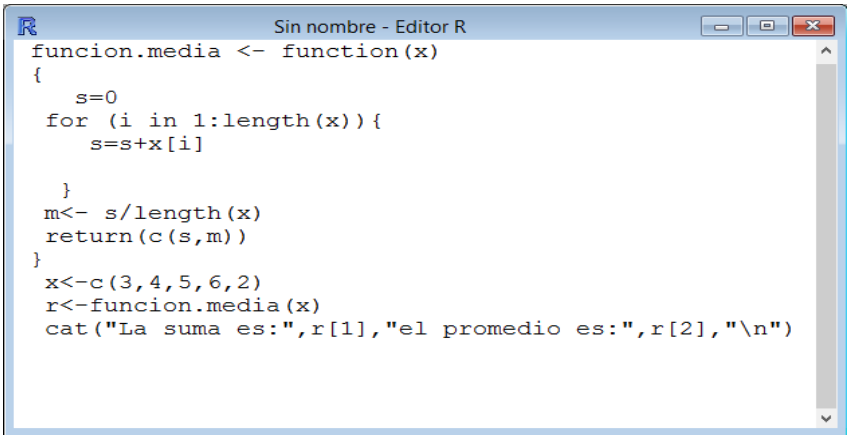
Una función se define asignando a un objeto la palabra "function", seguida de los argumentos que desee dar a la función, escritos entre paréntesis y separados por comas, y finalmente el cuerpo de la función entre llaves.

"Nombre_función <- function (argumento1,argumento2,…) { cuerpo de la función}"

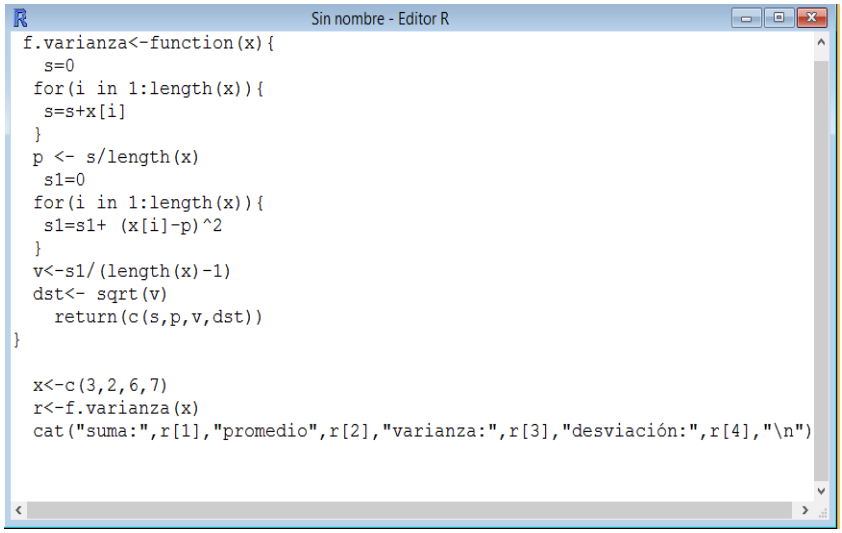
La siguiente función nos permite sumar dos datos, mediante la función ***return*** la cual recibe dos argumentos.



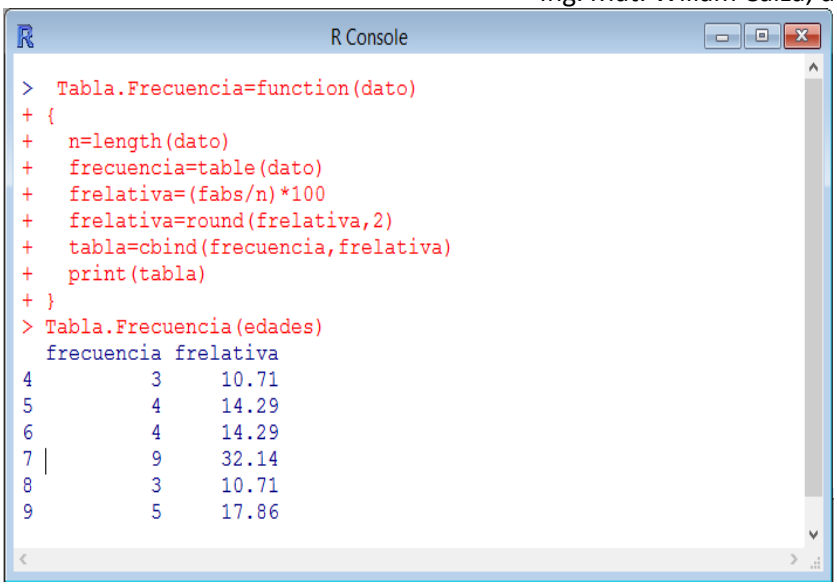
Mediante la siguiente función calculamos la media de un conjunto de datos, en el cual se observa el uso de "for" y "cat", retorna la suma y promedio de los datos. También se puede observar el uso de la función.



La función "f.varianza" hace el cálculo del promedio la suma de los cuadrados de los datos centrados, la varianza y la desviación estándar.

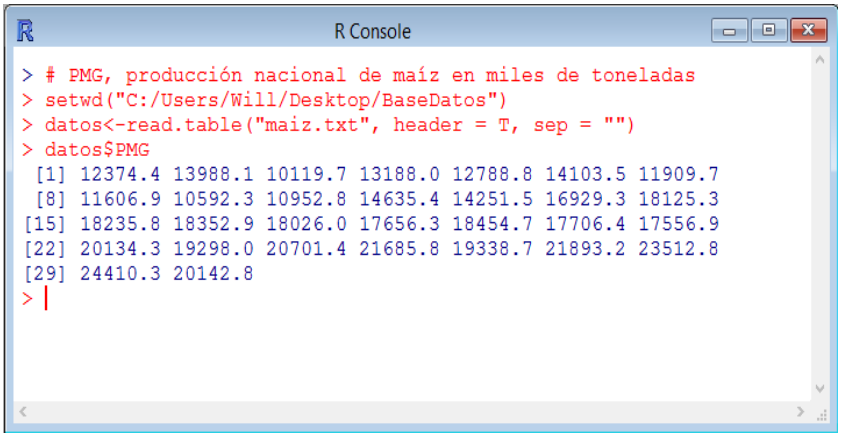


Con la función "Tabla.frecuencia", se imprime una tabla con tres columnas la una es los datos, la frecuencia y la frecuencia relativa, se puede observar el uso de la función "table".



Las siguientes funciones descritas a continuación nos permiten calcular la tabla de frecuencias.

La función “setwd”, indica la dirección por omisión de los archivos de R y además se lee el archivo “maíz.txt” para hacer uso las funciones.



La función “w”, calcula la longitud de la clase del histograma y el número de clases, se hace uso de “1+3.3log(n)” para el cálculo del número de clases.

```
> w=function(dato)
+     {
+      n=length(dato)
+      rango=max(dato)-min(dato)
+      k=1+3.3*log10(n)
+      k=round(k)
+      w=rango/k
+     l<- list(w=w,clases=k)
+     return(c(l))
+ }
> r1<- w(datos$PMG)
> r1
$w
[1] 2381.767

$clases
[1] 6

>
```

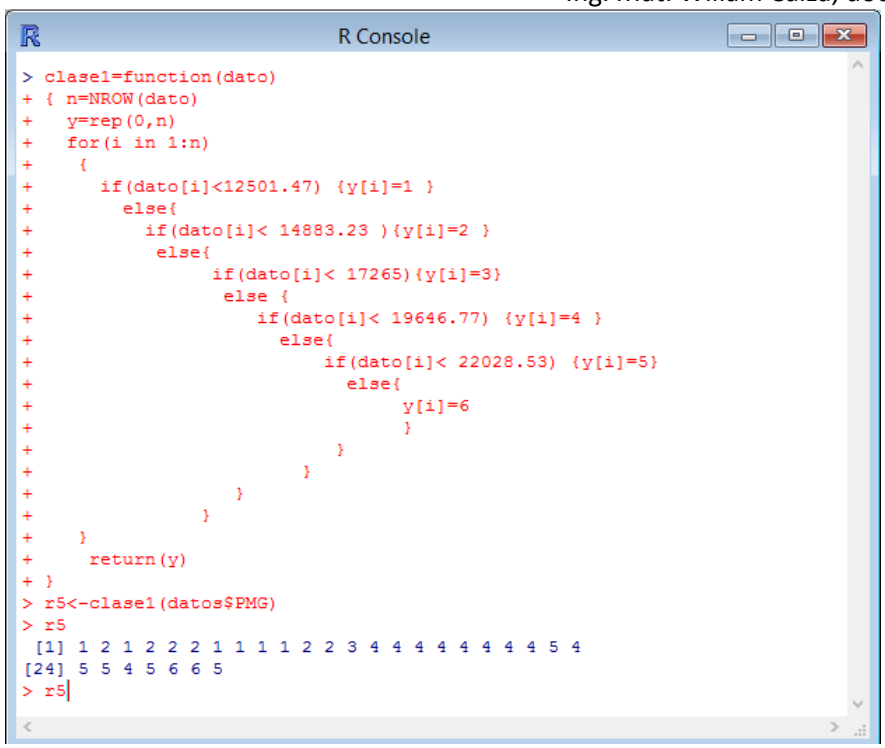
La función “limites”, calcula los limites inferior y superior de cada clase, por ejemplo clase 1 [10119.70; 12501.47], etc.

```
> limites=function(dato,w,clases)
+ {
+    LMin=min(dato)
+    LMax=LMin+clases*w
+    aa=seq(LMin,LMax,by=w)
+    return(aa)
+ }
>
>
> r2<-limites(datos$PMG,r1$w,r1$clases)
> r2
[1] 10119.70 12501.47 14883.23 17265.00
[5] 19646.77 22028.53 24410.30
>
```

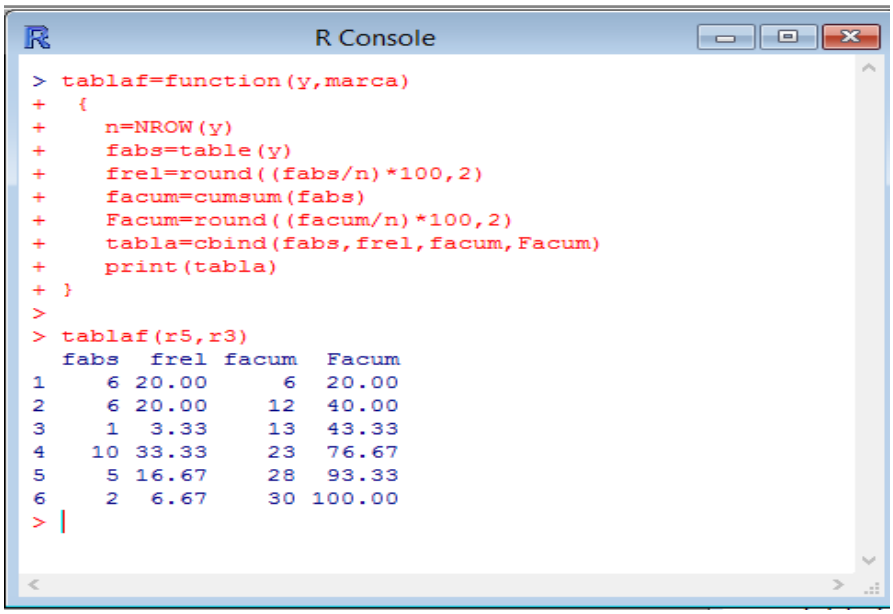
La función “marca”, encuentra los puntos medios de cada clase mediante el uso de la función “seq()”.

```
> marca=function(limites)
+ {
+    k=NROW(limites)
+    tic=limites[2]-limites[1]
+    ma=seq(limites[1]+tic/2,by=tic,length=k-1)
+    return(ma)
+
+ }
>
> r3<- marca(r2)
> r3
[1] 11310.58 13692.35 16074.12 18455.88
[5] 20837.65 23219.42
>
```

La función “clase1”, separa los datos para las clases existentes.



La función “tablaf”, calcula la frecuencia relativa, frecuencia acumulada y la frecuencia relativa acumulada.



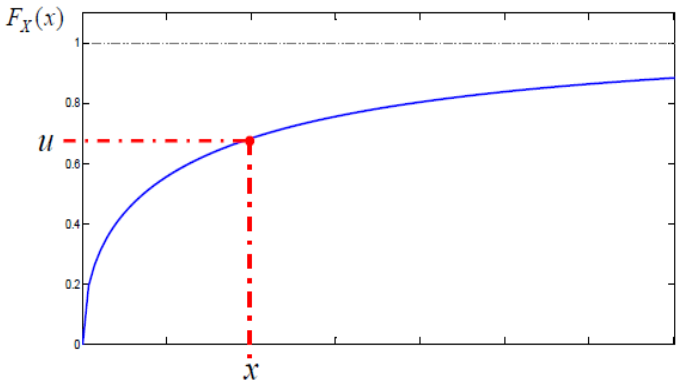
Anexo 2: Simulación

Uno de los temas especiales a tratarse es la aplicación de la estadística a la vida diaria, queremos comenzar con esto e iniciar con la simulación de las variables aleatorias.

Definición: Función de distribución inversa

Sea X una variable aleatoria con función de distribución $F_X(X = x)$ que admite inversa, entonces se verifica

$\mathcal{U} = F_X(X = x)$, $x = F_X^{-1}(\mathcal{U} = u)$ y $u \sim \mathcal{U}(0,1)$, $x = F_X^{-1}(\mathcal{U} = u)$ sigue la distribución de X.



Iniciemos con la variable aleatoria exponencial

$$F_X(x)\begin{cases}1 - e^{-\lambda x}, & x \geq 0\ y\ \lambda > 0 \\ 0, & si\ x < 0\end{cases}$$

Para un valor puntual tenemos $u = F_X(x)$

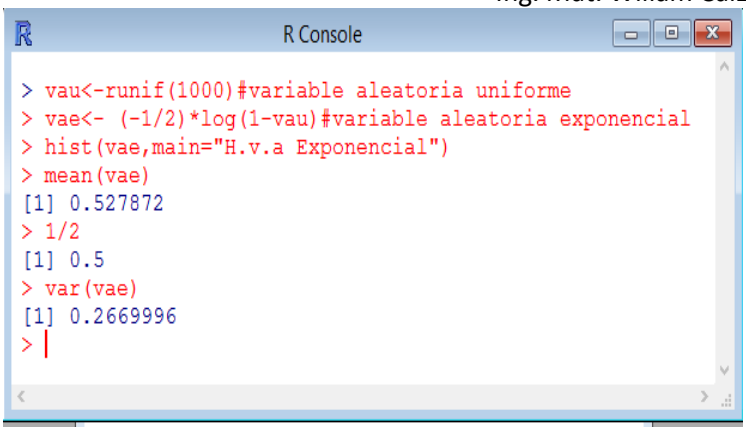
$$1 - e^{-\lambda x} = u$$

$$\ln(1 - u) = \ln(e^{-\lambda x})$$

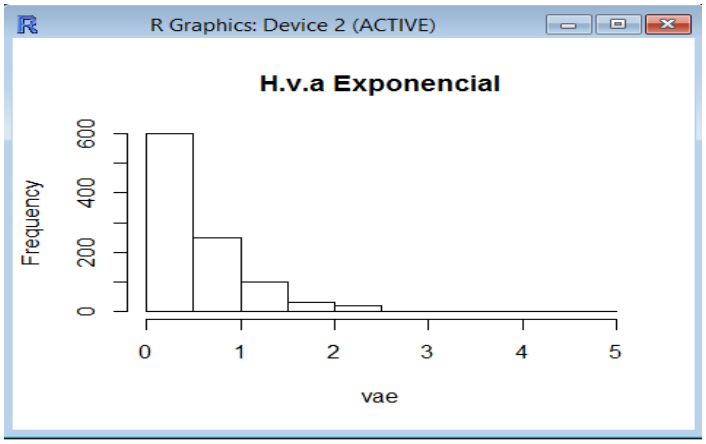
$$\ln(1 - u) = -\lambda x$$

$$x = \frac{-1}{\lambda}\ln(1 - u)$$

Al programar el algoritmo anterior, podemos observar que se realiza el cálculo de la media y varianza de los números generados.



Se puede observar en el histograma que las x (vae) generados por el algoritmo nos da un histograma parecida a la función de densidad exponencial por lo tanto la esperanza matemática (promedio) es 0.52 lo cual nos indica que λ= 2.



Generando valores de la variable aleatoria de Weibull tenemos,

$$X \sim W(\alpha, \beta)$$

$$F_X(x)\begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, & x \geq 0 \; y \;\; \alpha, \beta > 0 \\ 0, & si \;\; x \;\; < 0 \end{cases}$$

mediante el método anterior podemos obtener el algoritmo para calcular valore que sigan una distribución de Weibull

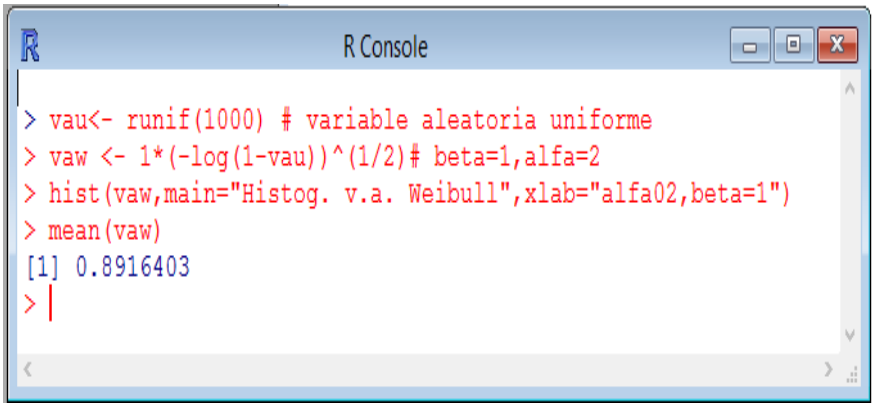
$$u = 1 - \; e^{-(\frac{x}{\beta})^{\alpha}}$$

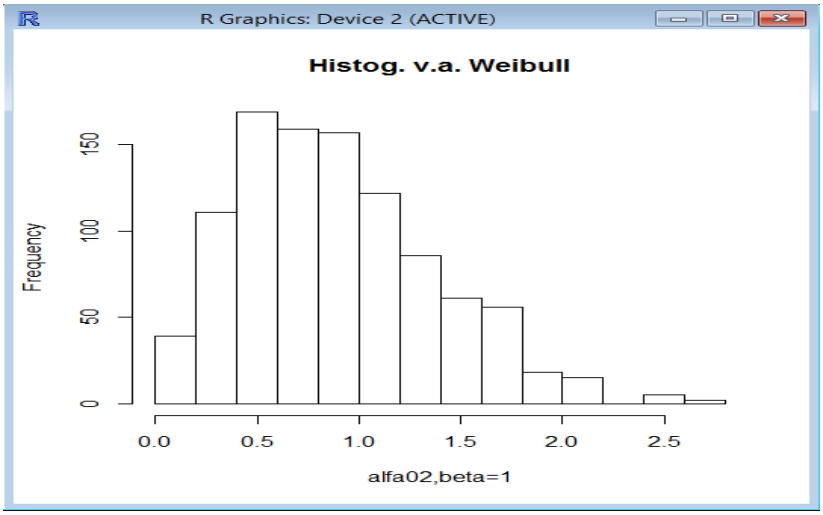
$$1 - u = \; e^{-(\frac{x}{\beta})^{\alpha}}$$

$$\ln(1-u) = -(\frac{x}{\beta})^{\alpha}$$

$$(-\ln(1-u))^{\frac{1}{\alpha}} = \frac{x}{\beta}$$

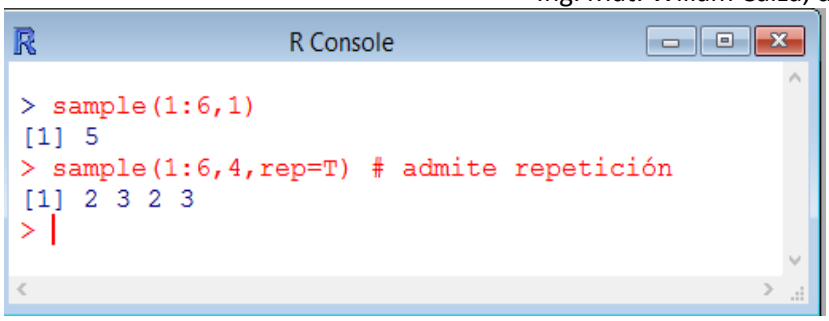
$$\beta(-\ln(1-u))^{\frac{1}{\alpha}} = x$$





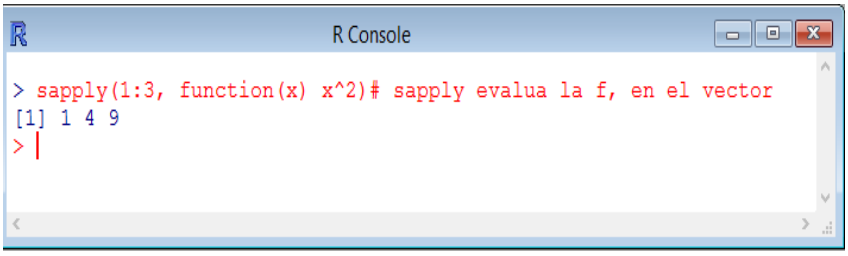
Ejercicio

Simular el lanzamiento de un dado y simular el lanzamiento de un dado cuatro veces con repetición.



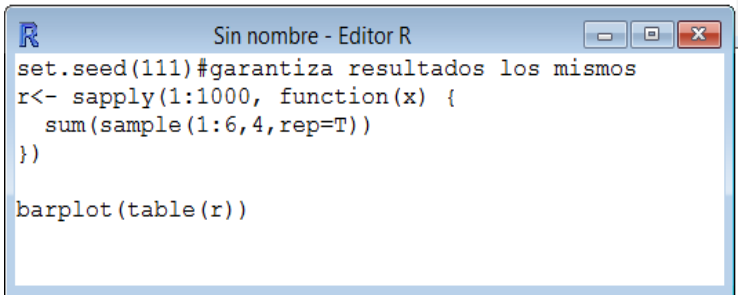
Ejercicio

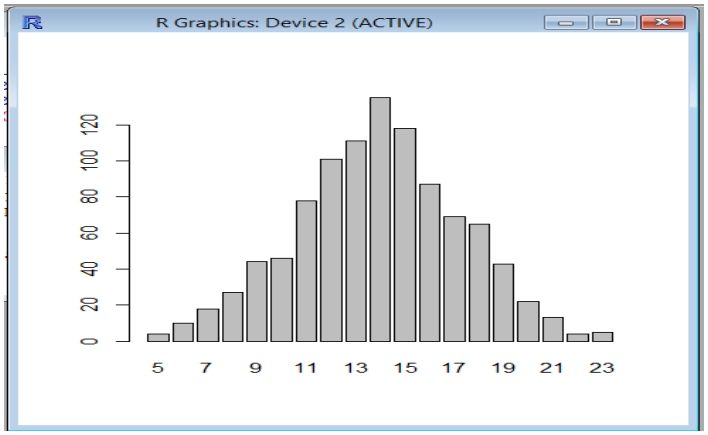
Evaluar una función para x= {1,2,3}, utilizando la función “sapply”.



Ejercicio

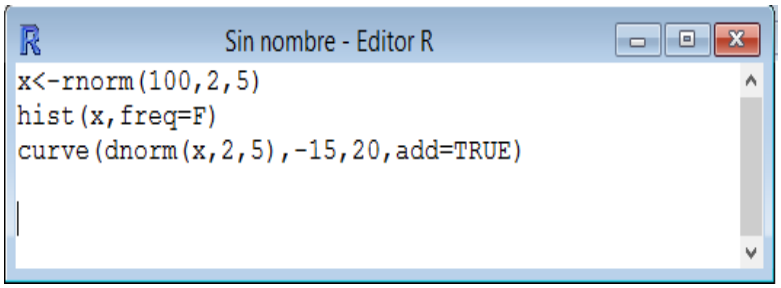
Simular la distribución de la suma de los números que salen al lanzar cuatro dados.

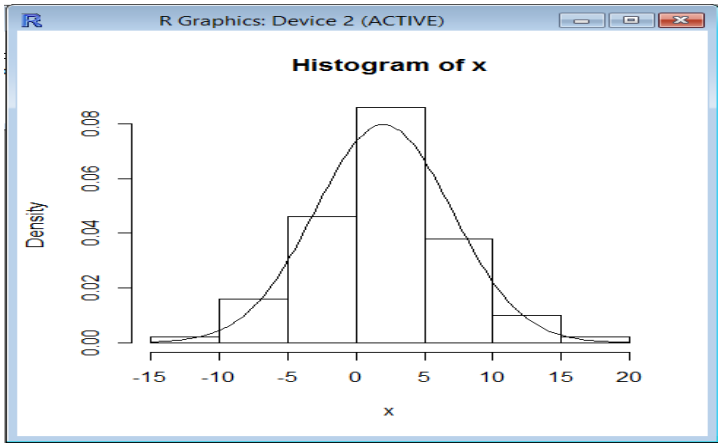




Ejercicio

Simular una distribución normal con media 2, desviación estándar 5 y su función de densidad



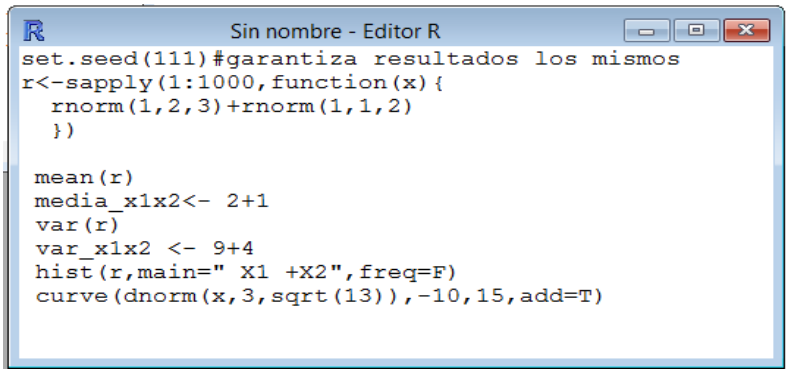


Ejercicio

Simular la distribución de dos distribuciones normales

$$X_1 \sim N(2,3)$$

$$X_2 \sim N(1,2)$$



Se ha verificado que la suma de dos distribuciones normales es una nueva distribución normal con parámetros

$$X_1 + X_2 \sim N(2+1, 3^2 + 2^2)$$

Ejercicio

Demuestre y realice una simulación de la "falta de memoria de una distribución exponencial".

Parte a)

Por definición se tiene

$$P(X > s+t \ \backslash X > t) = \frac{P(X > s+t \ \cap \ X > t)}{P(X > t)} = \frac{P(X > s+t)}{P(X > t)}$$

$$= \frac{1-Fx(s+t)}{1-Fx(t)} = \frac{1-1+e^{-\lambda(s+t)}}{1-1+e^{-\lambda t}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} \quad (1)$$

Tenemos

$$P(X > s) = 1 - P(X < s) = 1 - 1 + e^{-\lambda s} = e^{-\lambda s} \ (2)$$

(1) =(2)

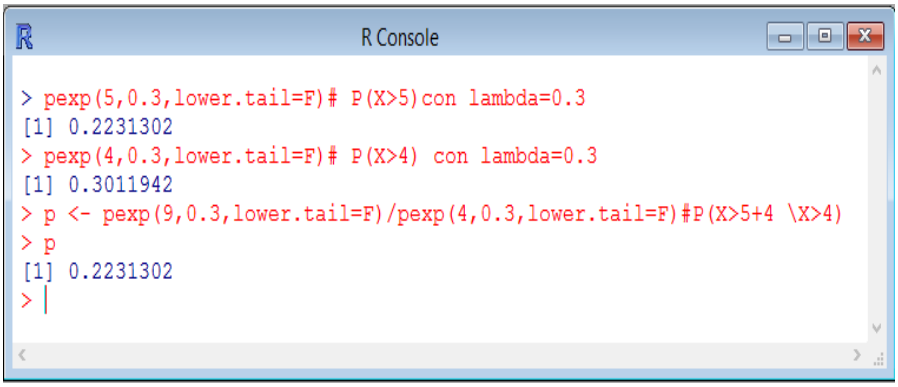
$$\boldsymbol{P(X > s+t \ \backslash X > t) = P(X > s)} \ \textbf{(a)}$$

La expresión (a) nos indica que la distribución exponencial, no tiene memoria.

Parte b)

$$P(X > 5+4 \ \backslash X > 4) = P(X > 9 \ \backslash X > 4) = \frac{P(X > 5+4 \ \cap \ X > 4)}{P(X > 4)}$$

$$= \frac{P(X > 5+4)}{P(X > 4)}$$

$$= \frac{P(X>9)}{P(X>4)} = P(X > 5) \ \text{por (a)}$$

Comprobemos