

---

---

# MANUAL EN R IN L<sup>A</sup>T<sub>E</sub>X

---

---

REALIZADO POR:

- **GLENN NICOLAS RICO**
- **MAYKOLL GIL**
- **MIGUEL ANGEL BARRERA**

**GLENNN.RICOL@KONRADLORENZ.EDU.CO**

**MIGUELA.GOMEZB@KONRADLORENZ.EDU.CO**

**MAYKOLLJ.GILM@KONRADLORENZ.EDU.CO**

# Table of Contents

<b>1</b>	<b>Brief Introduction to Latex</b>	<b>2</b>
1.1	Downloading Tex Studio . . . . .	2
1.2	Source Code and Resume Templates . . . . .	2
1.3	LaTeX Files . . . . .	2
1.3.1	LaTeX Commands . . . . .	3
1.4	General Items . . . . .	3
1.5	Terminology to Know . . . . .	3
1.6	Tex Studio Commands . . . . .	3
1.7	Creating Templates . . . . .	3
<b>2</b>	<b>Probabilidad</b>	<b>5</b>
2.1	Condicionada . . . . .	5
2.2	Bayes . . . . .	6
2.2.1	Teorema de Bayes . . . . .	6
2.3	Distribución binomial . . . . .	8
2.4	Distribución de Poisson . . . . .	8
2.5	Distribución Z normal . . . . .	9
2.6	Intervalos de confianza . . . . .	10
2.6.1	Pruebas de hipótesis . . . . .	10
2.6.2	Prueba de normalidad . . . . .	10
2.7	Prueba de varianza . . . . .	11
2.8	Prueba de media . . . . .	12

## About This File

---

This file was created for the benefit of all teachers and students wanting to use Latex for tests/exams/lessons/thesis/articles etc.

The entirety of the contents within this file, and folder, are free for public use.

# Brief Introduction to Latex

LaTeX, which is pronounced «Lah-tech» or «Lay-tech», is a document preparation system for high-quality typesetting. It is most often used for medium-to-large technical or scientific documents but it can be used for almost any form of publishing.

When using Latex, ***make sure to keep your source code organized by indenting and using sections/chapters/subsections.*** Not keeping your source code organized makes it harder to fix possible errors in your code.

## 1.1 Downloading Tex Studio

Latex is free to download and is available on all types of operating systems. To download  $\text{\LaTeX}$ , click the url to the right: <https://www.texstudio.org/>

## 1.2 Source Code and Resume Templates

If interested in obtaining the source code for this guide, see:

<https://sourceforge.net/p/latex-source-code/wiki/Download/>

If interested in obtaining some of the resume templates I have created, see:

<https://sourceforge.net/p/latex-resume-template/wiki/download/>

## 1.3 LaTeX Files

LaTeX files have names that end with the extension .tex: for example, an acceptable file name might be myfile.tex. (Never use spaces in file names.) The input file contains both the text of

your document and the LATEX commands needed to format it. The first command in the file, `\documentclass`, defines the style of the document.

### 1.3.1 LaTeX Commands

To distinguish them from text, all LATEX commands (also called control sequences) start with a backslash `\`. A command name consists of letters only and is ended by a space or a non letter.

## 1.4 General Items

### `\usepackage`

Packages in latex allow for the user to include cool and unique features into ones document. Most packages for latex will be found on [www.ctan.org](http://www.ctan.org)

```
\usepackage{package}
```

## 1.5 Terminology to Know

- Preamble: All the code that precedes your `\begin{document}`. The preamble section of your document is where all the formatting takes place.
- Commands: Commands are special words that determine  $\text{\LaTeX}$  behavior.
- Environments: Environments are used to format blocks of text in a  $\text{\LaTeX}$  documents

## 1.6 Tex Studio Commands

- The button with the magnifying glass, labeled view, in the top part of Tex Studio allows you to view your document sidebyside with your code.
- The green arrows next to the magnifying class, labeled compile, allows you to update your document with any new code you include. Compiling also allows you to check if there are any errors in your code.

## 1.7 Creating Templates

One of the many great things about  $\text{\LaTeX}$  is that offers many ways to avoid some of the repetitive things that come with creating several documents. One of the features  $\text{\LaTeX}$  offers is to create personal user templates for files that a user often uses but does not want to continuously type out every time.

For example, lets say you wanted to save a general layout of a lesson document that included **both** the preamble and some of the actual text and commands in the document. To create the template, one would first proceed to the folder originally containing this pdf (The Contents You Downloaded from the Zip File)) and then go to the folder labeled **Files to make Templates out of**. Once there, click the file named: **Lesson Template**. Once you have opened that file in textstudio, go to **File**, then **Make Template**. Once that is done, give the template a name and then press ok.

After creating a template, all a user needs to do to use that template is to go back to **File**, and then **New From Template** and then click the template you created.

# Probabilidad

La probabilidad es simplemente qué tan posible es que ocurra un evento determinado cuando interviene el azar.

## 2.1 Condicionada

La probabilidad condicionada mide la probabilidad de un determinado suceso conociendo información previa sobre otro suceso.

Dado dos sucesos **A** y **B** tales que **B** ≠ **0**, se denomina probabilidad de **A** condicionada de **B** que se escribe **P(A/B)**, definida:

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

De la fórmula de la probabilidad condicionada se puede derivar una expresión que nos resulta:

$$P(A \cap B) = P(A/B) \cdot P(B)$$

Esta es la expresión llamada principio de la probabilidad compuesta.

Ejemplo vamos a calcular la probabilidad del rango en el lanzamiento de 3 dados de tres formas diferentes.

```
# primera forma
rango.dados1<- function(n){
  b1<-matrix(0,n,3)
  b1<-sapply(1:n,function(x){sample(1:6,3,T)})
  r<-apply(b1,2,max)-apply(b1, 2, min)
  prob<-table(r)/n
```

```

    return (prob)
  }
  rango.dados1(100)

```

Para ejecutar este código tenemos en cuenta las aplicaciones de **apply**, **table** y se va a ejecutar en 100 pasos y el resultado sera de la siguiente manera.

```

r
  0    1    2    3    4    5
0.05 0.13 0.24 0.24 0.22 0.12

```

Ahora vamos a utilizar la siguiente linea de código para calcular los vectores usando las simulaciones.

```

# segunda forma
rango.dados2<- function(n){
  b1<-rep(0,n)
  for (i in 1:n) {
    b1[i]= diff(range(sample(1:6,3,T)))
  }
  prob<-table(b1)/n
  return(b1)
}
rango.dados2(100)

```

El código debería ejecutado debería ejecutar lo siguiente:

```

[1] 1 4 2 5 4 3 4 0 2 1 4 3 1 4 2 3 4 2 1 5 3 5 2 4 2 4 3 2 3 5 3 2 2 5 5 4 4 3 1
[40] 4 5 1 1 2 4 3 5 5 1 4 3 1 2 3 4 3 5 2 2 3 4 4 4 4 3 3 2 2 1 1 2 5 4 2 5 4 2 3
[79] 3 0 2 2 2 2 2 3 3 3 3 2 4 4 1 4 4 5 4 2 1 4

```

Y por ultimo la ultima sentencia de código que vamos a utilizar va hacer de la siguiente manera

```

# tercera forma
rango.dados3<-function(n){
  b1<-c()
  b1<-replicate(n,sample(1:6,3,T))
  r1<-apply(b1,2,max)-apply(b1,2,min)
  prob<-table(r1)/n
  return(prob)
}
rango.dados3(100)

```

Ejecutado debe aparecer los vectores que se evalúan en las posiciones del 1 al 6.

```

r1
  0    1    2    3    4    5
0.03 0.13 0.14 0.27 0.20 0.23

```

## 2.2 Bayes

### 2.2.1 Teorema de Bayes

El teorema de Bayes o regla de Bayes sea  $B_1, \dots, B_n$  donde es una partición de de los posibles resultados, talque cada una de estas particiones la probabilidad sea distinta de cero (0) Sea B un

suceso cualquiera del que se conocen las probabilidades condicionales  $P(B|A)$ . Entonces, la probabilidad  $P(A|B)$  viene dada por la expresión:

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

Donde cada probabilidad se define de la siguiente manera:

- $P(A)$  son las probabilidades a priori
- $P(B|A)$  es la probabilidad de  $B$  en la hipótesis  $A$
- $P(A/B)$  son las probabilidades a posteriori.

Ahora vamos a ejecutar un código para Bayes que funciona para cualquiera que sea el problema que se quiere ejecutar.

```
print ("ingrese el porcentaje ")
porcentaje<-c(scan())
print ("ingrese la probabilidad ")
prob<-c(scan())

funcionb<- function (porcentaje ,prob){
  suma <- 0.00
  for(i in 1: length(porcentaje)){
    suma<-suma+porcentaje[i]*prob[i]
  }
  resta <- 0.00
  probabilidad<-c()
  for(i in 1:length(porcentaje)){
    resta<-porcentaje[i]*prob[i]/suma
    probabilidad<-c(probabilidad , resta)
  }
  return (probabilidad)
}

print (funcionb (porcentaje ,prob))
```

Para probar que si funciona vamos a ejecutar un problema.

Una empresa que fabrica celulares tiene dos maquinas A y B cada maquina fabrica el 60% y 40% de los celulares respectivamente y cada maquina tiene un 5% y 10% respectivamente de celulares defectuosos entonces ¿Cual es la probabilidad de que el celular haya sido fabricado por la maquina A sabiendo que es defectuoso?

Usando el código simple vamos a digitar los porcentajes que van hacer **0.06** y **0.04** que son los porcentajes de fabricación de cada maquina, y después digitamos la probabilidad que es el defecto de cada maquina seran **0.05** y **0.10** nos va quedar como solución:

```
[1] 0.4285714 0.5714286
```



## 2.3 Distribución

## binomial

Una distribución binomial es una distribución de probabilidad discreta que describe el número de éxitos al realizar en  $n$  experimentos independientes entre sí, acerca de una variable aleatoria.

Se entiende como una serie de pruebas o ensayos en la que solo podemos tener 2 resultados (sea éxito o fracaso), siendo el éxito nuestra variable aleatoria.

$$P_{(x)} = \binom{n}{r} p^x q^{n-x}$$

$$P_{(x)} = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Donde:

$p$  = Probabilidad de éxito

$q$  = Probabilidad de fracaso

$x$  = Variable aleatoria

$n$  = Número de ensayos

En R ya viene la función binomial y se puede usar de la siguiente manera **binom** pero no se usa así de simple se tiene que poner **dbinom** para usar la función binomial simple, probabilidad acumulada se usa **pbinom**, la inversa de la función de distribución, es decir, los percentiles **qbinom** y **rbinom** se usa para los valores binomiales generados.

**Ejemplo:** Imaginemos que un 80% de personas en el mundo han visto el partido del final del último mundial de fútbol. Tras el evento, 4 amigos se reúnen a conversar. ¿Cuál es la probabilidad de que 3 de ellos hayan visto el partido?

```
n <- 4
p <- 0.8
x <- 3
dbinom(x, n, p)
```

nos va quedar como solución:

```
[1] 0.4096
```

## 2.4 Distribución

## de

## Poisson

Esta distribución es una de las más importantes distribuciones de variable discreta, los experimentos que resultan en valores numéricos de una v.a  $X$  y que representan el número de resultados durante un intervalo de tiempo dado o en una región específica frecuentemente se conocen como experimentos Poisson.

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!}$$

Veamos el siguiente ejemplo: Suponga que los accidentes de una calle siguen un proceso de Poisson con una tasa de 2 accidentes por semana.

- Hallar la probabilidad de que se comentan 5 accidentes durante la próxima semana.

Vamos a calcular mediante la función **pois** al igual que la función binomial tiene las mismas características, **dpois**, **ppois**, **qpois** y **rpois** **dpois** es para la  $x = X$  probabilidad acumulada se

usa **ppois**, la inversa de la función de distribución, es decir, los percentiles **qpois** y **rpois** se usa para los valores binomiales generados.

Ahora para solucionar el ejemplo se usa el siguiente código:

Ejecutado el problema primero se halla lambda teniendo en cuenta que lambda se halla mediante, si 2 accidentes son por 1 (una) semana entonces por 5 accidentes en la siguiente semana entonces

$$2 \rightarrow 1$$

$$x \rightarrow 1$$

Entonces calcular la media y se encuentra que es dos, ese es el valor de lambda.

$$x < -5$$

$$\text{lambda} < -2$$

$$\text{dpois}(x, \text{lambda})$$

$$[1] \quad 0.03608941$$

## 2.5 Distribución

## Z

## normal

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

Una variable aleatoria continua, X, sigue una distribución normal de media  $\mu$  y desviación típica  $\sigma$ , y se designa por  $N(\mu, \sigma)$ , si se cumplen las siguientes condiciones:

- La variable puede tener valores desde  $(-\infty, \infty)$
- La función de densidad, es la expresión en términos de ecuación matemática de la curva de Gauss

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Para estandarizar un valor de X:

$$Z = \frac{x - \mu}{\sigma}$$

Donde  $\mu$  es la media y  $\sigma$  es la distribución estándar.

En R para hacer una distribución normal podemos usar la siguiente función **norm** al igual que las distribuciones de Poisson y binomial tiene:

**dnorm, pnorm, qnorm y rnorm.**

**dnorm** es para la  $x = X$  probabilidad acumulada se usa **pnorm**, la inversa de la función de distribución, es decir, los percentiles **qnorm** y **rnorm** se usa para los valores binomiales generados.

Vamos hacer el siguiente problema.

La temperatura durante septiembre está distribuida normalmente con media 18,7°C y desviación estándar 5°C. Calcule la probabilidad de que la temperatura durante setiembre esté por debajo de 21°C

```
x<-21
media<-18.7
desvia<-5
pnorm(x, media, desvia)
```

Y la solución para el siguiente ejercicio es:

```
[1] 0.6772419
```

## 2.6 Intervalos de confianza

### 2.6.1 Pruebas de hipótesis

Son procedimientos de decisión basados en datos que puedan producir una conclusión acerca de algún sistema científico.

Una **hipótesis estadística** es una afirmación o conjetura acerca de una o más poblaciones.

Tener en cuenta:

- Hipótesis nula  $H_0: \mu = \mu_0; \mu \leq \mu_0; \mu \geq \mu_0$
- Hipótesis Alternativa:  $H_1: \mu = \mu_1; \mu < \mu_1; \mu > \mu_1; \mu \neq \mu_1$

Para probar las hipótesis se tiene la prueba de normalidad de Kolmogorov-Smirnov, pruebas de variabilidad y pruebas de medias.

### 2.6.2 Prueba de normalidad

Los resultados de la prueba indican si usted debe rechazar o no puede rechazar la hipótesis nula de que los datos provienen de una población distribuida normalmente.

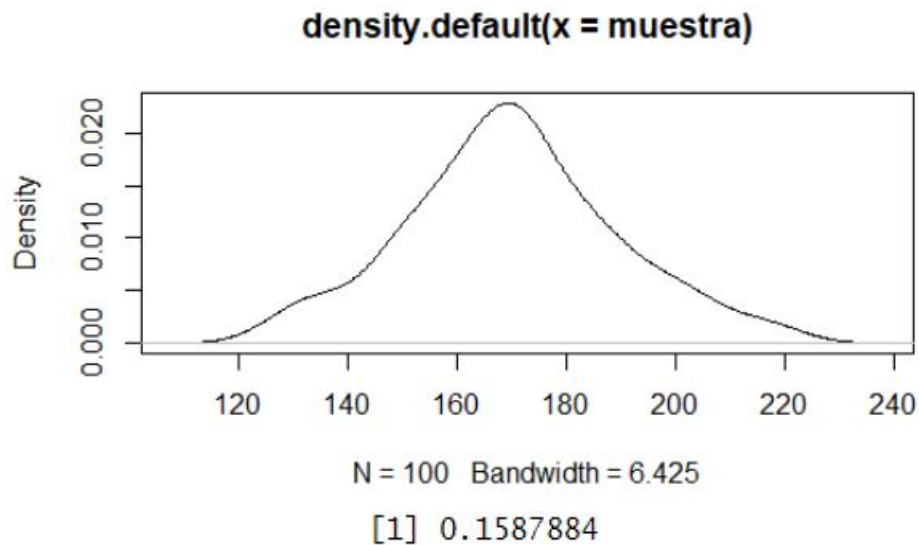
Vamos a usar la prueba de normalidad de Kolmogorov-Smirnov y de Shapiro Wilk

- **Prueba de normalidad de Kolmogorov-Smirnov:**

Esta prueba compara la función de distribución acumulada empírica de los datos de la muestra con la distribución esperada si los datos fueran normales. Se va a usar el comando **lillie.test** hace el cálculo de Kolmogorov-Smirnov mejorado para usar **lillie.test** tener en cuenta que toca instalar **nortest** y vamos a usar un ejemplo propio para eso, teniendo en cuenta que para Kolmogorov-Smirnov se usa para datos mayores a 50.

```
set.seed(100)
muestra<- rnorm(n=100,mean=170,sd=19)
plot(density(muestra))
library(nortest)
lillie.test(muestra)$p.value
```

y su respuesta y gráfica serían las siguientes:

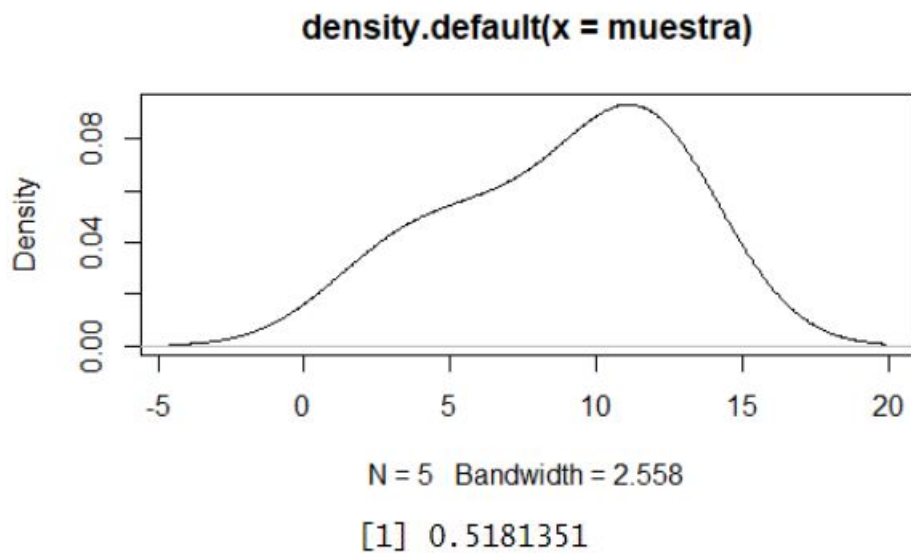


respuesta Kolmogorov-Smirnov

- **Prueba de normalidad Shapiro Wilk:**

se usa para contrastar la normalidad de un conjunto de datos. Se plantea como hipótesis nula que una muestra  $x_1, \dots, x_n$  proviene de una población normalmente distribuida

```
set.seed(50)
muestra <- rnorm(n=5, mean=10, sd=4)
plot(density(muestra))
library(nortest)
shapiro.test(muestra)$p.value
```



respuesta Shapiro Wilk

## 2.7 Prueba de varianza

Pruebas de hipótesis para una varianza es un procedimiento para juzgar si una propiedad que se supone cumple una población estadística es compatible con lo observado en una muestra de dicha población en este caso la varianza, para ello formularemos dos Hipótesis (llamada

"Hipótesis Nula") y (llamada "Hipótesis Alternativa"), con ellas realizaremos una o más pruebas, para tratar de encontrar cual se debe rechazar.  
cuando se trabaja con varianza se utiliza  $\chi^2$  que es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos., por lo tanto nos vamos a concentrar en  $\chi^2$ .

Entonces vamos a R para usar crear una hipótesis para usar  $\chi^2$  en R es simplemente usar **chisq.test.**

```
fumar=matrix(c(51,43,22,59,56,45,48,78,95),ncol=3,byrow = TRUE)
colnames(fumar) = c("mucho","poco","medio")
rownames(fumar) = c("actual","pasado","nunca")
fumar=as.table(fumar)
fumar
chisq.test(fumar)
```

En el código hemos creado nuestra hipótesis de fumadores donde creamos una tabla de contingencia, datos inventados.

	mucho	poco	medio
actual	51	43	22
pasado	59	56	45
nunca	48	78	95

Y la respuesta de  $\chi^2$ :

Pearson's Chi-squared test

```
data: fumar
X-squared = 28.766, df = 4,
p-value = 8.722e-06
```

## 2.8 Prueba de media

Se utiliza una prueba de una muestra para probar una afirmación con respecto a una media de una población única, en vez de estimar el valor de un parámetro, a veces se debe decidir si una afirmación relativa a un parámetro es verdadera o falsa.

En este caso Vamos a usar la prueba t que es muy buena que se utiliza cuando deseamos comparar dos medias (las cuentas se deben medir en una escala de intervalo o de cociente) Vamos A ejecutar un código en R de una prueba t student entonces para una prueba t student en E se necesita escribir el código con el siguiente código **t.test.**

```
set.seed(10)
x1 <- rnorm(100,10)
x2 <- rnorm(100,10.5)
test <- t.test(x1,x2)
print(test)
```

Ejecutando el código.

```

Welch Two Sample t-test

data: x1 and x2
t = -4.0081, df = 197.83,
p-value = 8.665e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8080508 -0.2751220
sample estimates:
mean of x mean of y
 9.863451 10.405037

```

Como el p-value es  $< 0.05$  podemos afirmar que las muestras difieren en su media, es decir, las dos variables son diferentes con un gráfico de cajas se puede ayudar a interpretar este resultado, las medias se indican mediante un punto azul:

```

boxplot(x1, x2, names=c("X1", "X2"))
medias <- c(mean(x1), mean(x2))
points(medias, pch=18, col="blue")

```

