Tuesday, 10 March 2020

Student Name/ID Number	CELIS VASQUEZ SONIA PATRICIA
Unit Number and Title	**12: Data Analytics**
Academic Year	2020
Unit Tutor	Eduardo Caro
Assignment Title	**Data Analytics: Descriptive, inference, and predictive techniques**
Issue Date	**January 29th, 2020**
Submission Date	March 27th, 2020
IV Name & Date	Javier Cara

DATA ANALYTICS: DESCRIPTIVE, INFERENCE, AND PREDICTIVE TECHNIQUES

A. Introduction to data analytics

1. Give an example for the following terms:
    a. Data. - Representación de un atributo o variable cuantitativa o cualitativa.
        i. Nombres de las calles -
    b. Information. - Datos procesados con significado.
        i. Mapa -
    c. Knowledge. - Integrar datos e información con experiencia para la toma de decisiones. - Inferencia.
        i. Ruta -

2. Go to the web https://www.teoalida.com/cardatabase/, section "European Car Models", and download the "Demo/sample free" Excel database. Please, organize the data in the following categories:
   - Categorical data
     - *Nominal data*
     - *Ordinal data*
   - Numerical data
     - *Discrete data*
     - *Continuous data*

Categorical data - Cualitativos - NO medibles - determinan modalidades

- ***Nominal data*** *- Nominales - no Ordinales - no se pueden ordenar, no tiene sentido ordenarlas, No puedo calcular la media, Los datos son principalmente alfabéticos, son datos "etiquetados" o "nombrados" que pueden dividirse en varios grupos*

  European / World classification
  Make
  Model
  Country of origin
  Country
  American classification
  Description
  Pre-1990 car models

- ***Ordinal data*** *- Ordinales - se pueden ordenar, tiene sentido ordenarlas, tienen un orden de categorías mientras que los nominales no.*

  Platform / generation number
  Sold in Western Europe
  Sold in North America
  Sold in Europe
  Sold in North America
  Sold in India
  Timeline included
  Class

Numerical data - Cuantitativos - medibles

- ***Discrete data*** *- Discretos - número finito de valores enteros. - barras*

  Units produced
  Production years

Model Years (US/Canada)
Model Years (North America)
Units produced
Models included (WORLD)
Models included (EURO)
Company Founded
First car produced
Last car produced
Years produced
Year
models
Number of car models by the year of launch

- ***Continuous data*** *- Continuos - cualquier valor real infinito dentro de un intervalo. - histogramas*

B. Descriptive Analytics

3. For the data file "*Tablet Computer Sales.txt*", find the average number, standard deviation, variance, and interquartile range of units sold per week.

```
(base) hadoop@ubuntu-hokkaido-3568:~/R/Data$ cat Tablet_Computer_Sales
.txt
Week     Units_Sold
1        88
2        44
3        60
4        56
5        70
6        91
7        54
8        60
9        48
10       35
11       49
12       44
13       61
14       68
15       82
16       71
17       50
(base) hadoop@ubuntu-hokkaido-3568:~/R/Data$
```

Leer el fichero
> TablaVentas=read.table("Tablet_Computer_Sales.txt", header=T)

El fichero contiene 17 pares de valores Units_Sold y Week:
> length(TablaVentas[[1]])
[1] 17

Los nombres de los campos de los ficheros son:
> names(TablaVentas)
[1] "Week"   "Units_Sold"

La media de las unidades vendidas es:
> mean(TablaVentas$Units_Sold)
[1] 60.64706

La varianza de las unidades vendidas es:

```
> var(TablaVentas$Units_Sold)
[1] 253.8676
```

La desviación estándar de las unidades vendidas es:

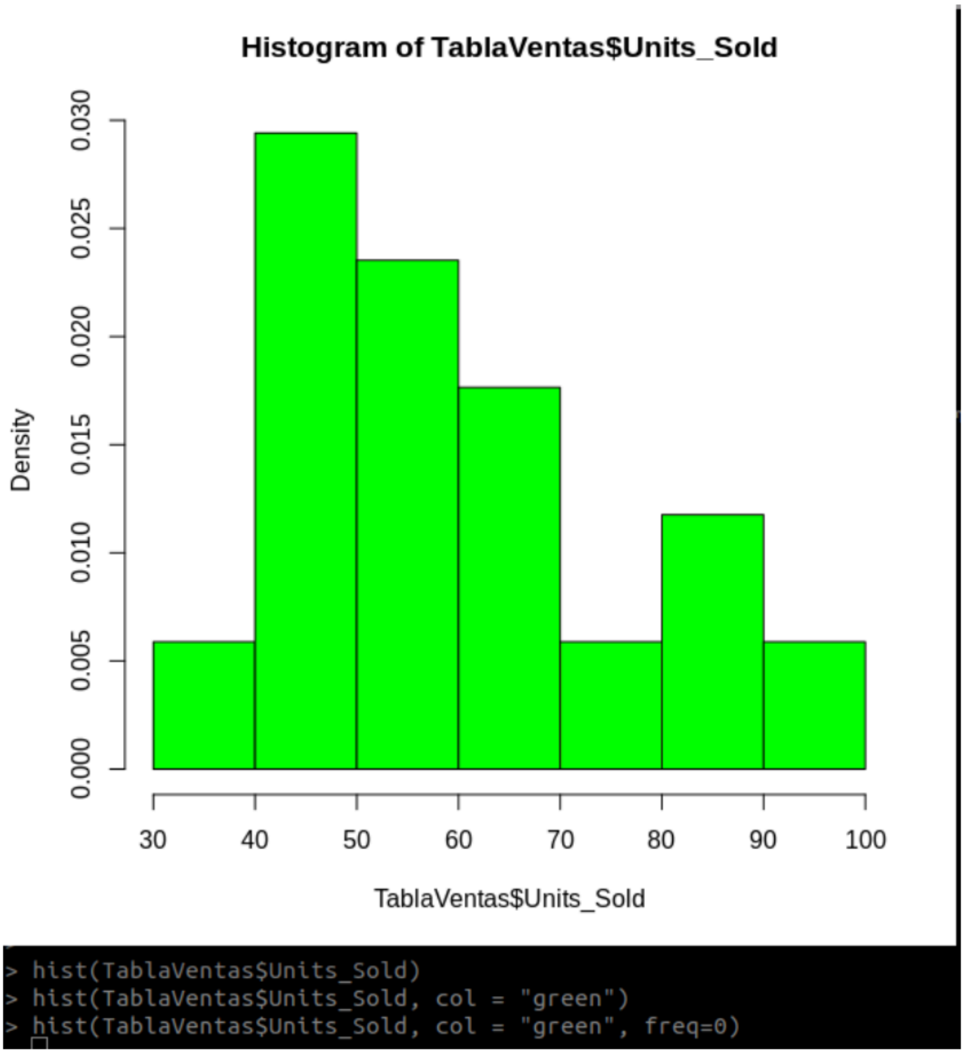
```
> sqrt(var(TablaVentas$Units_Sold))
[1] 15.93322
```

La mediana de las unidades vendidas es:

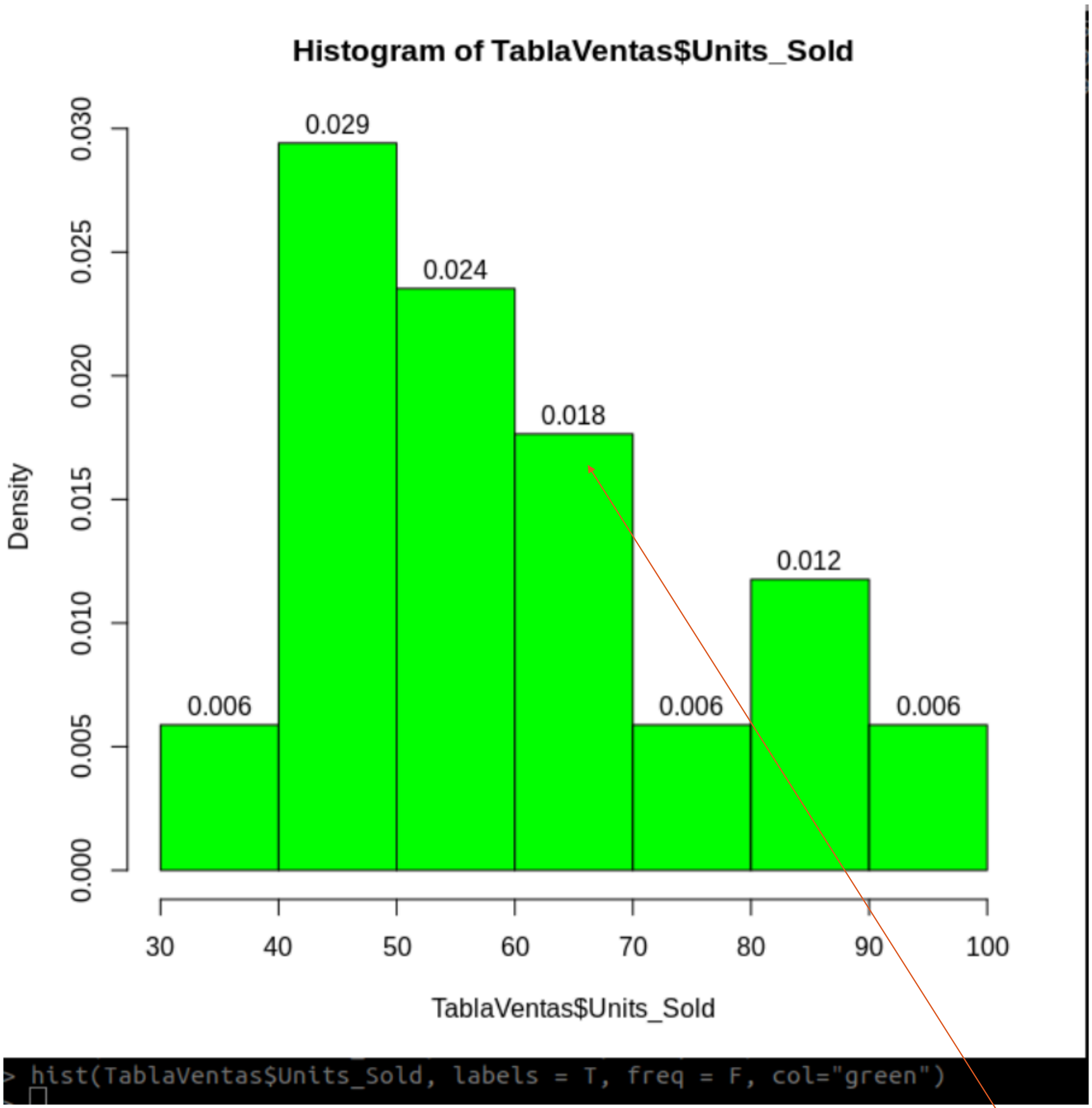
```
> median(TablaVentas$Units_Sold)
[1] 60
```

El eje x de la gráfica nos muestra las Unidades Vendidas y en el eje y podemos ver la frecuencia relativa de estas ventas por semana.

```
> hist(TablaVentas$Units_Sold, col = "green", freq=0)
```



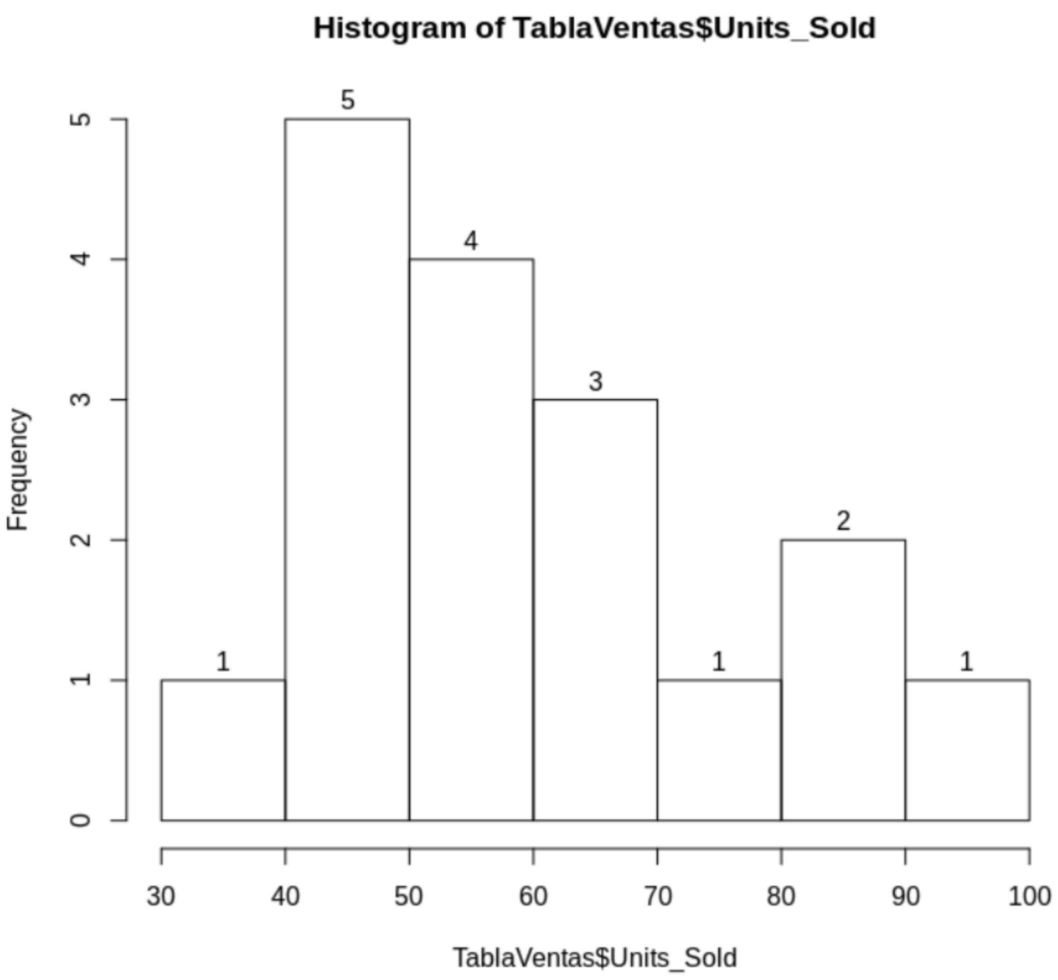
> hist(TablaVentas$Units_Sold, labels = T, freq = F, col="green")



esta frecuencia se multiplica por la amplitud del rango (el ancho de la base) y nos indica que el 18% de las unidades vendidas esta entre 60 y 70.

En el Histograma podemos ver el comportamiento de las ventas

> hist(TablaVentas$Units_Sold, labels = T)




```
> TablaVentas=read.table("Tablet_Computer_Sales.txt", header=T)
> hist(TablaVentas$Units_Sold, labels = T)
```

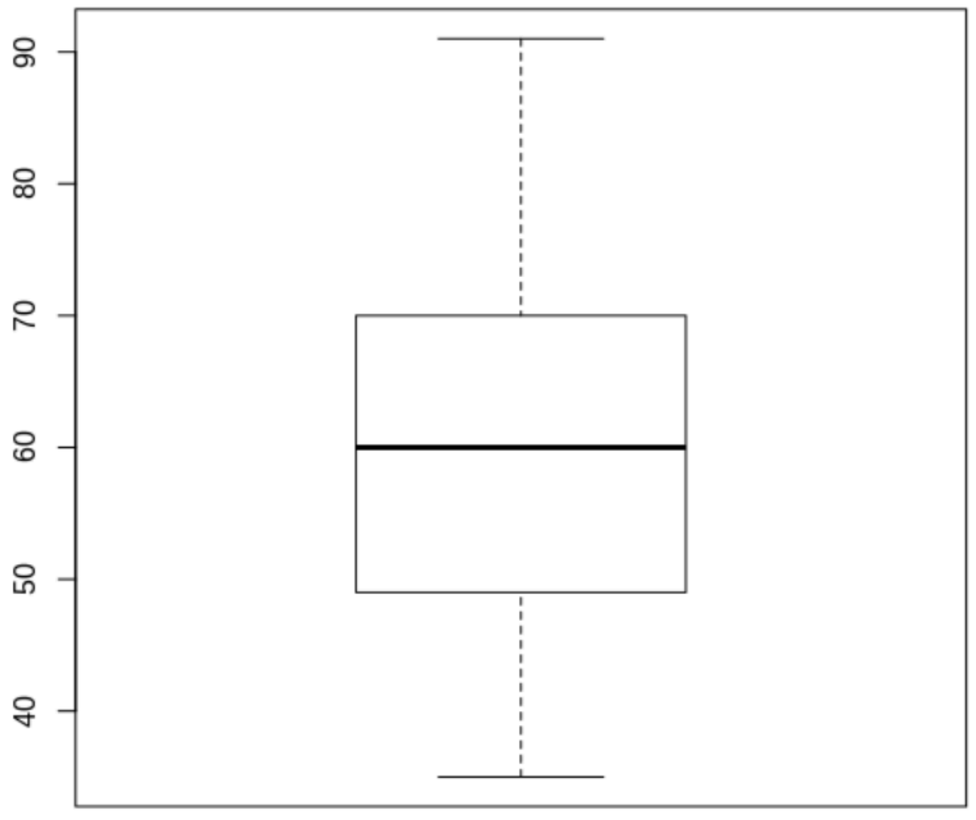
El eje x de la gráfica nos muestra las Unidades Vendidas y en el eje y podemos ver la frecuencia absoluta de estas ventas.

El 50% de las unidades vendidas están por debajo de la mediana, entre 30 y 60, El 50% de las unidades vendidas están por encima de la mediana, entre 60 y 100.

Por debajo de la mediana se han hecho 1+5+4 = 9 ventas, el 25%.
Por encima de la mediana se han hecho 3+1+2+1 = 7 ventas, el 25%

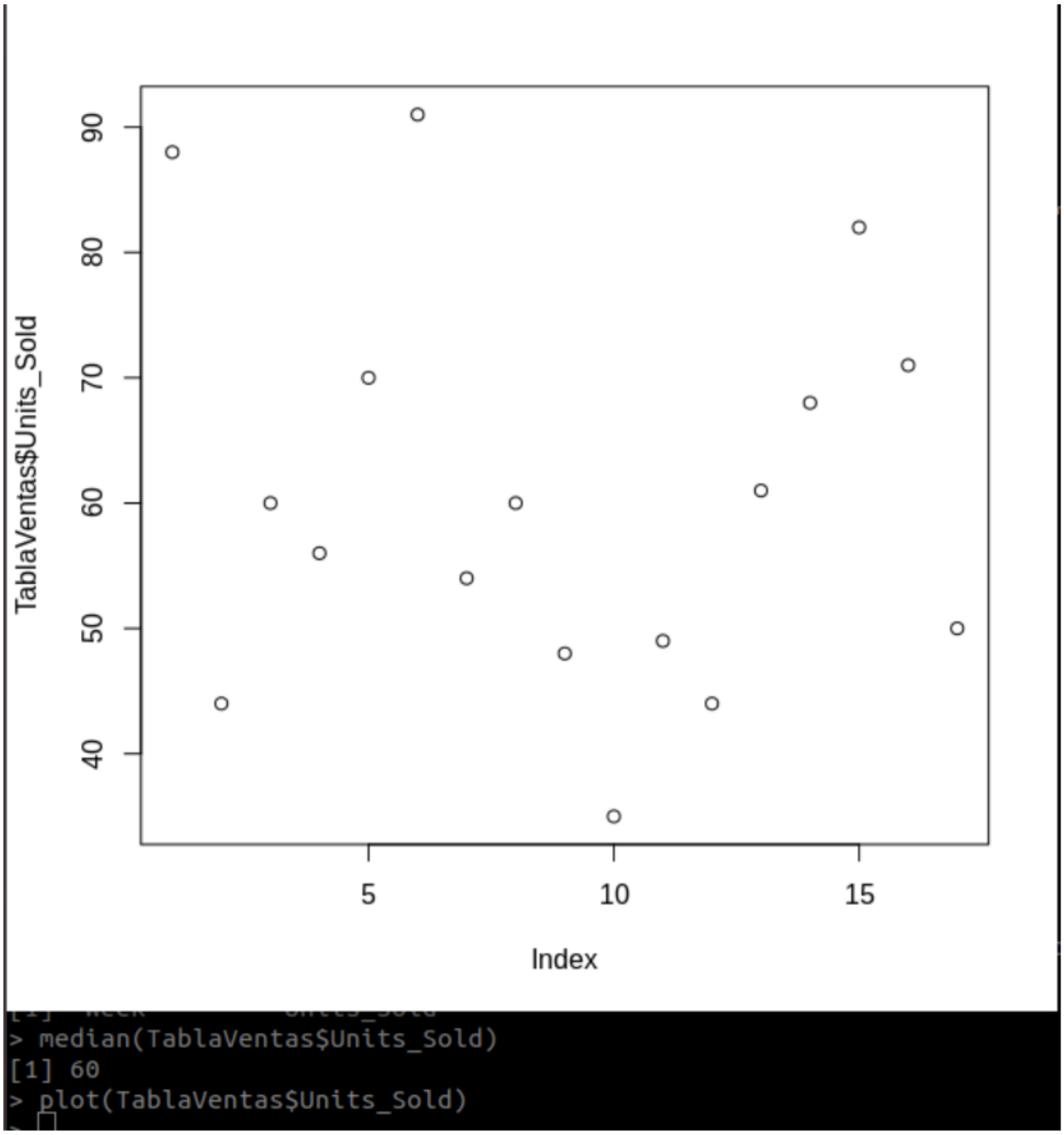
> quantile(TablaVentas$Units_Sold)
 0% 25% 50% 75% 100%
 Q1 60 Q3
 35 49 60 70 91

El 25% de las unidades vendidas están por debajo de el Q1= y
El 25% de las unidades vendidas están por encima de el Q3
La mitad de las ventas se hacen entre el Q1=50 y el Q3=70.
El Rango InterQuartil = Q3 - Q1 = 20 que nos indica la dispersión de los datos.




```
[1] 60
> plot(TablaVentas$Units_Sold)
> boxplot(TablaVentas$Units_Sold)
```

No tenemos puntos atípicos.


TablaVentas$Units_Sold
90
80
70
60
50
40
5
10
15
Index
> median(TablaVentas$Units_Sold)
[1] 60
> plot(TablaVentas$Units_Sold)

C. Probability -

4. Let us define X as a random variable Gaussian-distributed with $\mu = 4$ and $\sigma = 3$.
   - Compute $P(X \leq 6)$
   - Compute $P(3 \leq X \leq 6)$
   - Compute a such as $P(X \leq a) = 0.85$
   - Generate 1000 random numbers distributed as X.
     - Plot a histogram with these numbers, and superimpose the theoretical density function
     - Compute the proportion of the numbers that has been generated in the interval [3, 6], and compare with the probability computed above in $\mathrm{P}(3 \leq \mathrm{X} \leq 6)$
     - Check if the generated random numbers have a Gaussian distribution.

RNORM -> Crea n valores aleatorios con un mu y un sigma dado.
rnorm random de la distribución normal
rnorm(n, mean, sd)

Generate 1000 random numbers distributed as X.

```
> x = rnorm( 1000, 4, 3)
> x
  [1]  9.722570951  6.211427198  2.508262913  4.139724814  4.677790110
  [6]  1.230995381  1.965396672  6.488571367  3.072921287  6.790270359
 [11]  8.320486977  7.315415223  9.037359923  6.458714038  6.009824193
 [16]  4.607721569  4.207768344  4.719899337  2.479265433  2.602571266
 [21]  5.671424456  5.444647257  3.723626253  1.837756028  2.930583570
 [26]  7.917282936  3.012074310  5.188248465  8.024267729  5.421459376
 [31]  3.981487017  6.361560591  4.271581798  1.243004533  1.944055671
 [36]  1.530621602  2.480864685  4.240728436  1.882294691  1.337529812
 [41] -1.882293641  3.585676118  6.929947789  7.849017666  0.339912836
 [46]  1.983010350  2.730427892  3.896994672  5.918259763  7.162565008
 [51] -0.287221541  5.187820980  3.548050820  4.830088653  3.311645221
 [56]  0.915211273  3.612222944  9.842090769  4.742765099 -0.835973072
 [61]  1.766909070  1.671513668  6.495753925  5.129725893  3.918217479
 [66]  6.768604611  5.834482601  1.516918895  6.140512728  8.874400451
 [71]  4.514966635  2.532810399  5.432233809  0.726443464  4.271291944
 [76]  1.774256967  7.597992599 13.055332535  3.576914728  1.751239379
 [81]  4.952006568  2.370382055  1.079241610  6.795152411  3.321105427
```

```
 [86] -0.576366567  4.409198008 -1.238873051 -0.335542623  2.932252428
 [91] -0.822976474  5.862416781  5.224077734  1.298535381  5.560786148
 [96]  1.726152640  4.405386415  6.062141376  6.933085281  0.083971564
[101]  5.939252510  0.712851474  8.258705038  6.061144946  4.468130686
[106]  0.289010919  0.525983395  2.998358336  0.514076221  3.643954712
[111]  7.995448922  3.230590080  4.657879228  2.641335486  3.178088515
[116]  7.152351631  4.917588796  4.092624350  3.149892615  0.869055959
[121]  8.543357113  9.895700597  9.641745611 -0.903208267  1.698466801
[126]  3.741987027  2.600176448  8.851360217  5.114223753  3.378683645
[131] -5.468752355  0.219529107  4.177014481  7.445974180  1.282261997
[136]  6.364841480  4.158220335  2.429991719  7.112274650  1.523045448
[141]  0.177278636  2.817009377  3.879597499  1.458014354  8.364717483
[146]  2.626763080  5.992223951  1.497974660  1.471208855  2.472373746
[151]  2.140080913 -0.173138318  0.738701084  1.100190538  3.093848425
[156]  2.614752807 -1.269680625  4.113679107  0.630165691  8.225266756
[161]  0.390878421  1.328138590  4.312852878  4.624936995 -0.638777809
[166]  4.066598692  2.215765286  5.690783357 -0.910915638  4.446947131
[171]  0.811697070 10.765247879  1.208862971  4.663853639  6.174358129
[176]  5.210056644  2.756640294  5.905663114  6.637139488 -1.526801276
[181]  4.850419670  2.454648589  6.042974894  5.672352385  3.957725545
[186]  1.770495991 -0.236739938  3.628012973  3.027004404  0.254988469
[191]  5.506717494  1.850413858 -0.563487826  4.751301635  9.866927720
[196] -1.535536387  5.071618159  6.144802877  3.687932530  5.955797143
[201]  4.215097928  5.822850543  7.654388015  1.647258955  3.539643139
[206] -0.233199808  1.037190762  1.558937496 -0.686512648  3.841825701
[211]  2.392496050  5.991446739  8.566035726  4.111370956  4.371670295
[216]  8.282059268 -6.113036001 11.402304920  3.474051749  4.239006719
[221] -0.114112348  5.079776356 -3.809169347  7.036561787 -1.514828920
[226]  3.360214337  4.365131538  4.461060799  7.383020658 11.699878215
[231]  6.186755272  6.123775530  6.539739394  1.794553599  2.485610905
[236]  2.608139050  3.566808815  0.497721853  2.866089227  5.828394493
[241]  2.178345540  5.093038819  4.961607481  2.437286659  1.363423225
[246]  7.491729676  2.414379528  0.106438284  4.265413101  0.681171206
[251]  3.482857725  2.574357004  4.548963722 10.896553245  0.959729015
[256]  6.357540064 -2.745009234  6.829742809  0.721179865  6.199110413
[261]  3.765789212  1.705341167  0.278667849  1.696325044  4.077781824
[266]  3.432137284  7.761687800  3.724041081  8.545637001  9.422028957
[271]  5.415925932  4.638817511  8.386504230  6.678765443  1.897813242
[276]  3.423741539  1.793430722  2.329396411  6.039721013  2.264599874
[281]  7.161069995  1.220731111  1.314265577  7.696737849 12.913710838
[286]  7.666838966  4.385567897 -0.025855143  8.174583388  7.421116900
[291]  1.936413434  5.222448047  5.433769895  6.152562759 -2.393373931
[296] 10.550173858 -2.381103694 -0.259253982  2.798480173  2.904275533
[301]  1.619783402  5.069009040  5.284249074  5.996968546  4.768729185
[306]  7.487155127  6.990478337  0.558350666  7.996831624  9.180178705
[311]  1.096819799  0.708577377  3.173212465  3.401918301  1.099976911
[316]  1.728649864  5.812703433  0.959233117  2.754107096 10.343180638
[321]  6.002077412  4.806700449  5.513053738  2.441421099  4.266799649
[326]  6.718228634  3.976782440  7.893239060  8.171056814  5.700936398
```

```
[331]  5.715773094  5.605739732  5.261658432  8.457835117 -1.772544152
[336]  0.313950185  0.139520466 -1.984582319 -0.358700320 -0.263102154
[341]  0.586050407  1.285688944  6.010951138 -1.509690832  9.145090283
[346]  3.405104049  0.008460258  7.067317943  2.072413652  1.409091624
[351]  6.388530766  3.294254067  5.777705081  3.862223343  1.765908491
[356]  2.014202425  6.227685870  5.164518611  4.327541490  2.750391887
[361]  4.017437234 -3.419899582  4.577251596 -0.219982295  6.666242077
[366]  3.754189859  9.072962750  2.681111516  5.663441872  6.332012385
[371] 10.220133183  5.714244418  3.474496491  3.372059874  2.454697039
[376]  4.194899577  5.468510852 -0.783665169  1.734988793  2.897535699
[381]  2.008500508  0.747606127  6.645002321  2.958695091  6.995609988
[386]  1.099431493  2.561533184  3.565269981  2.181268612  4.715741589
[391]  3.334422765  6.078309463  6.479170667  5.641301272  2.351223746
[396]  6.748727574  3.438648621 -0.984403643  3.258230305  3.916405463
[401]  6.523336276  2.742001492  4.888083652  1.221543887  5.796572221
[406]  2.615260506  2.020880629  8.423752873  7.753089644  3.649085955
[411]  2.144595410  3.583692428  3.418949356  2.614139292  8.880841274
[416]  3.715306563 -0.644387593  2.174673450  5.619982301  0.554610257
[421]  3.664605347  6.653025570  6.306438095 -3.125562697  7.915400069
[426]  3.681465943  0.023515992  5.496227963  1.912540843  2.301665062
[431]  1.425172971  6.289315635  4.137596070  8.652135426  4.357716195
[436]  9.688389615 11.097061609  2.177468042  6.773638776  3.236352421
[441]  3.424033871 -0.529310312  4.871063188 -1.641336992  4.631817378
[446]  9.633210482 12.245507500  4.226492924 -2.446737578  4.645199327
[451]  5.072334883  1.975459207  6.545352273  1.981379664  1.460565884
[456]  4.631088143  3.675709009  4.379561070  3.192372357  5.061189990
[461]  4.586050603  5.205766084  3.129277073  0.829500224  1.100961436
[466]  2.699422596  6.403515983  4.059579128  4.595158721  2.128305872
[471]  1.337695568  9.376709660  6.116801496  6.095803707  1.756079230
[476] -1.994034725  0.374882106  1.544116064  4.085145067  6.410087771
[481]  7.399036990  3.916130322  4.679054670  0.979837858  4.908624780
[486]  1.890664785  4.833422507  9.889101494  7.741111531  3.285925742
[491]  4.964946817  3.393052994  6.082636278  3.123121926 11.364334899
[496] 13.422952710  1.817754779  9.795658809  2.738309621  5.964871948
[501]  2.493937434  4.029730951 -1.428867514  6.566986311  6.859171481
[506]  2.785994358 -1.177333526  1.043135236  5.571551876  4.071726459
[511]  5.802865302  6.458471352  4.494384196  0.672108448  8.339207597
[516] -0.694187917  5.753471120  6.161758453  3.813362365  9.026642099
[521]  4.920723262  3.017662911  5.758058290  3.776638176  5.713066712
[526]  6.717520510  8.571852809  8.356271972  8.425171910  7.384317832
[531]  4.607066626 -1.833863776  2.120880818  7.261128695  4.810269152
[536]  2.199796221  3.540722374  1.206079789  0.586302636  0.909893323
[541]  5.924383252  6.127433614  7.475842428  2.671882303  1.786570889
[546]  3.134906336  2.636938477  7.564451876  3.783923923  7.074039609
[551]  0.747876358  5.029983622  3.123774359  1.567869069  3.924607543
[556]  1.696671031  7.280866206  0.265342285  4.240395833  5.916836214
[561]  2.303462153  3.203209429  7.468445875  3.684980977  5.530153561
[566]  4.160483592 -0.617730561 -0.559640075  0.267257765  5.437319434
[571]  5.829950364  1.057053985 -0.021887972  4.504516478  6.898440572
```

```
[576]  0.049168339  4.865983276  5.728722430  8.603317305  3.228644984
[581]  1.643562299  6.563363507  6.713575327  1.866850151  4.396785748
[586]  2.329852467  5.867410206  3.218018960  5.706537515  2.155291382
[591]  5.186967216 -4.343854430  6.875426612  3.481984436  3.245776003
[596] -0.619837891  3.967382207  0.785835133  2.896925374  0.447481059
[601]  5.388519671  7.895643684  6.050256180  6.065210797  3.538405686
[606] -1.980923129  3.991151349  2.560672766  2.916720680  3.200498488
[611]  0.987199092  0.123714759  6.870618245 -1.362221688  7.088870504
[616] -0.952470179  9.134490592  5.195534531  9.434308379  2.378293089
[621]  9.949477513  2.720707511  0.399286765  7.737786508  0.845552136
[626]  4.532702079  4.275943296  9.973260229  5.462484125  5.209040066
[631]  3.190813404  7.695102069  3.203859661  3.221599937  6.647455026
[636]  1.522869919  5.247372258  4.808008084  3.261479908  3.869782233
[641]  9.082012558  6.879933416  6.675219604  8.247169701  6.772527837
[646]  3.317865782  0.172107422  6.462343934  7.089053815 12.100819190
[651]  1.448039004  3.427384559  7.074528827  0.752427279  3.489820925
[656]  4.902163835 -0.405603491  2.348873931  6.286238261  6.236264022
[661]  6.689746628 -2.311298212  1.150579425  1.052796893  1.666651591
[666]  6.484820202  2.614540236  4.447779956  1.834115823  8.222455673
[671]  9.478416393 11.807093745  9.066914003  6.480361868  0.331249637
[676]  3.116467422  1.607748409  3.011952100  6.851059334  2.288312692
[681]  1.318825090  3.113127101  7.149866598  6.353973581  4.041625919
[686]  6.540997495  5.657859450  5.784999260  4.921612570  0.569106047
[691]  6.899596937  8.278806612  9.418848635  4.667785765  5.010480439
[696]  6.849067150  2.998923418  0.194999569  0.830643034 -1.703665134
[701]  4.423620743  5.530128807  5.832168866  7.295113537  5.430939950
[706]  3.798712695 -0.659069454  5.064475439  8.981936789  3.448453669
[711]  5.931427932  7.547698140 -0.068651111  0.876545805  4.311691798
[716]  2.782839192  4.436149155  3.657116238  2.400586382  6.195525581
[721]  5.433768829 -0.419938931  4.776671529  6.977721657  9.252854790
[726]  1.188472478  8.622609738  3.676219656  2.918144141  3.462001367
[731]  3.192548547  4.633496694  6.686889015  2.423551924  4.858554560
[736] -3.826658525  2.016617221  6.294495910  0.742242710  5.513646668
[741]  3.682213357  2.432986208  1.714940422  8.536148499  1.727820631
[746]  7.400382993  6.268730238 -0.100547394  2.992598117 10.239088016
[751]  0.377079593  1.307163822  9.481052138 -0.610184480  3.409957928
[756] -0.157292930  6.301604399  4.216543625  9.533605964  8.633555938
[761]  6.096183524  3.638639635  0.975317044  5.125299917  7.725145457
[766]  2.039650909  3.473240677 -3.823699615  4.323735351  9.584445785
[771]  8.441958706  4.973996546  1.391446783  0.197915150  2.350267556
[776]  6.655826449  3.217182008  8.403025471  1.762704936  2.328387307
[781] -0.800270146  2.572068771  2.386047458  3.241964880 -0.366419675
[786]  2.256641122  2.875214488  7.341394895 -0.055212531  6.525665636
[791]  3.616757834  2.775267178  2.881864591  1.461219143  3.106719755
[796]  4.390606324  4.407141063  3.214795843  9.033697695  4.700564903
[801]  4.275098151  2.738830189  4.457127847 -0.201995279  7.033111677
[806]  6.893747737  0.298077240  0.112085491 -1.195432870  3.454119713
[811]  1.669384094  7.257897452  3.736462251  3.308560151  4.574995281
[816]  3.559881939  3.275774048  5.772927225  3.315348544  2.619204467
```

```
[821]  6.174814610  4.741439332  0.990673810 -0.609219512  2.734989995
[826] 10.145072186  7.147153262  4.832156640  5.850977952  0.599510927
[831]  5.947160676  0.666216323  6.676434553 -0.482537939 -1.900172763
[836]  3.048900638  7.838839390  5.851302513  6.823795786  7.622565426
[841]  2.155805800  4.454464392 -2.864232298  3.730787576  1.120894235
[846]  0.104634148 -2.498893482  8.550164250  5.903829403 -0.780464293
[851]  0.833347259  3.743421608  4.654216211  4.112404714  6.452046902
[856]  2.115153668  3.292777150  4.460443593  5.803095809  1.986178477
[861]  3.651534724  2.209781762  3.061074017  6.125712681  7.652292576
[866]  5.954147369  0.641963218 -2.030439149 -0.221400731  5.813850201
[871]  7.807039328  6.772659234  3.810651952  5.726720584 -0.127383625
[876]  0.220868545  3.839319307  9.391743706  3.294372582  7.543433686
[881]  6.240118606  9.815001498  4.164777359  2.643334865  4.923317586
[886]  3.086636336  3.322454892  8.085314508  3.169242515 -0.292071627
[891]  5.928410471  3.623712369  0.809973052  6.922605274  7.403139386
[896]  7.094559573  5.496784410  2.855290592  2.978759222 -1.044969746
[901]  2.410711240  3.870699755  6.478521724  1.187777273  5.772889807
[906]  7.154531467  4.284654211  3.201363701  4.896757445  2.732244214
[911]  7.676932284  3.289225515  2.501054648  4.199085588  8.234579629
[916]  8.959128329  3.620792197  6.739632937  1.195287454  6.378219374
[921]  1.169563754 -0.989844446  1.913297041 -2.248200791  3.369106795
[926]  5.229325213  6.424564818  7.271546502  0.306133700  5.245739608
[931]  6.541283057  0.532106227  8.508294597  9.272515031 -3.964967504
[936]  5.908047796  0.329210230  0.851137323  4.710201587  3.689213706
[941]  3.425921543  3.725367522 10.072711343  7.826463739  3.373082876
[946]  2.211918808  5.076166422  1.831917174  4.838812525  1.821511528
[951] 10.055279490  9.085432287  6.226849496  5.378771335  4.440647634
[956]  0.898951431  8.990225899  8.137731194 10.412555275  1.548484348
[961]  6.427600105  3.785829083  3.665892064  3.483239993  8.124939878
[966]  8.236620858  4.214459737  5.700270755  6.415499418  6.353337432
[971]  9.115362245  4.316094520  5.997749778  5.719000083  8.863206817
[976]  5.727793109  3.708372726  2.664170473  1.084940086 -0.424586902
[981]  2.033561889  2.115813583  4.310248773  1.551119406  4.185972246
[986]  2.867766666  3.386759483 -1.161971444  6.829082800  4.699670773
[991]  3.621348788  6.557779893  2.022140773  7.489283598 10.234863323
[996] -0.466739909  8.887104089  4.004537421  6.949013300  2.850796691
```

pnorm —> Calcula la probabilidad - AREA - dando un mu y un sigma - sólo calcula las areas a la izquierda

- Compute $P(X \leq 6)$

pnorm(6, 4, 3) = 0.7475

- Compute $P(3 \leq X \leq 6)$

pnorm(6, 4, 3) = 0.7475
pnorm(3, 4, 3) = 0.3694

pnorm(6, 4, 3) - pnorm(3, 4, 3) = 0.7475 - 0.3694 = 0,3780

- Compute a such as $P(X \leq a) = 0.85$

qnorm —> devuelve la altura de la curva, con un mu= 4 y un sigma = 3.

qnorm(0.85,4,3) = 7.1093 —> por tanto el 0,85% de los números están por debajo de 7.1093

Plot a histogram with these numbers, and superimpose the theoretical density function -

- Compute the proportion of the numbers that has been generated in the interval [3, 6], and compare with the probability computed above in $\mathrm{P}(3 \leq \mathrm{X} \leq 6)$

pnorm —> me devuelve p; la probabilidad

pnorm(1000, 4, 3) = 1 —> Teorico

Partiendo de que la muestra es de 1000 datos y sabiendo que se trata de una distribución normal…

Calculo la suma de los números que son mayores ó iguales a 3 de la muestra x
> sum(x >= 3)
[1] 646

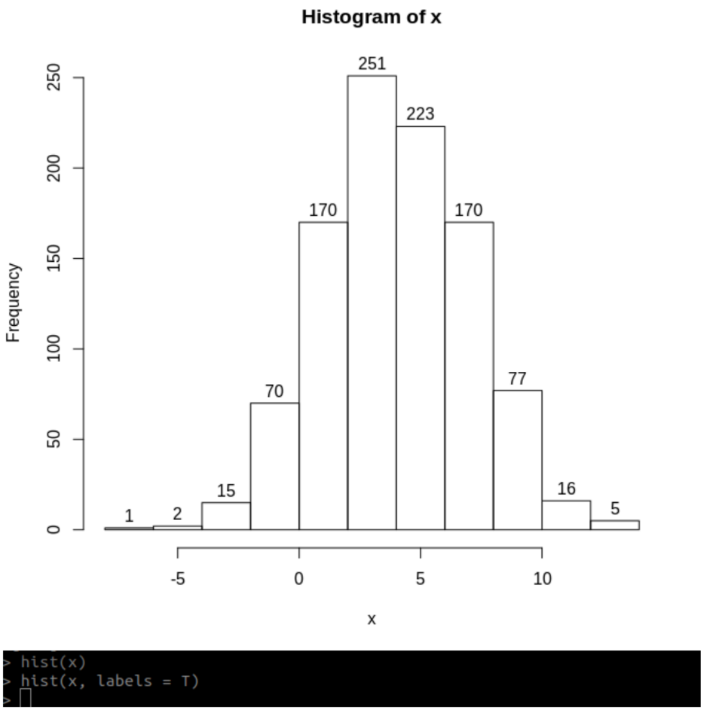
Calculo la suma de los números menores ó iguales a 6 de la muestra x
> sum(x <= 6)
[1] 732

Saco la diferencia y lo divido entre la muestra 1000 y tenemos una probabilidad de:
732 - 368 = 364
364/1000 = 0,364

Al realizar la comparación obtenida del calculo de los random de la variable Gaussian-distributed nos damos cuenta que la variación de la probabilidad es mínima.

Este histograma tiene una distribución normal "Experimental", —> tomando todos los x generados. (el histograma representa la curva de la muestra).

> hist(x, labels=T, freq=F)

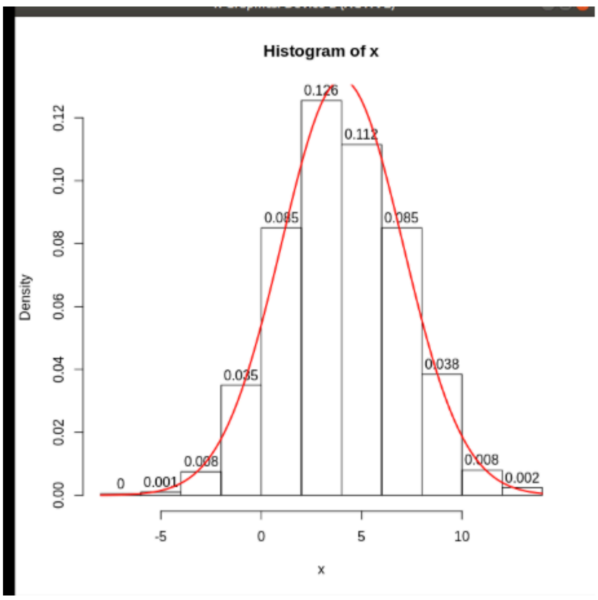


Check if the generated random numbers have a Gaussian distribution.

> media = mean(x) # calculo la media de x

> desv = sd(x) # calculo la desviación típica de x

> curve(dnorm(x,media, desv), add = T, col = "red", lwd = 2) # superpongo la función de la densidad, previamente calculada la media y la desviación típica.



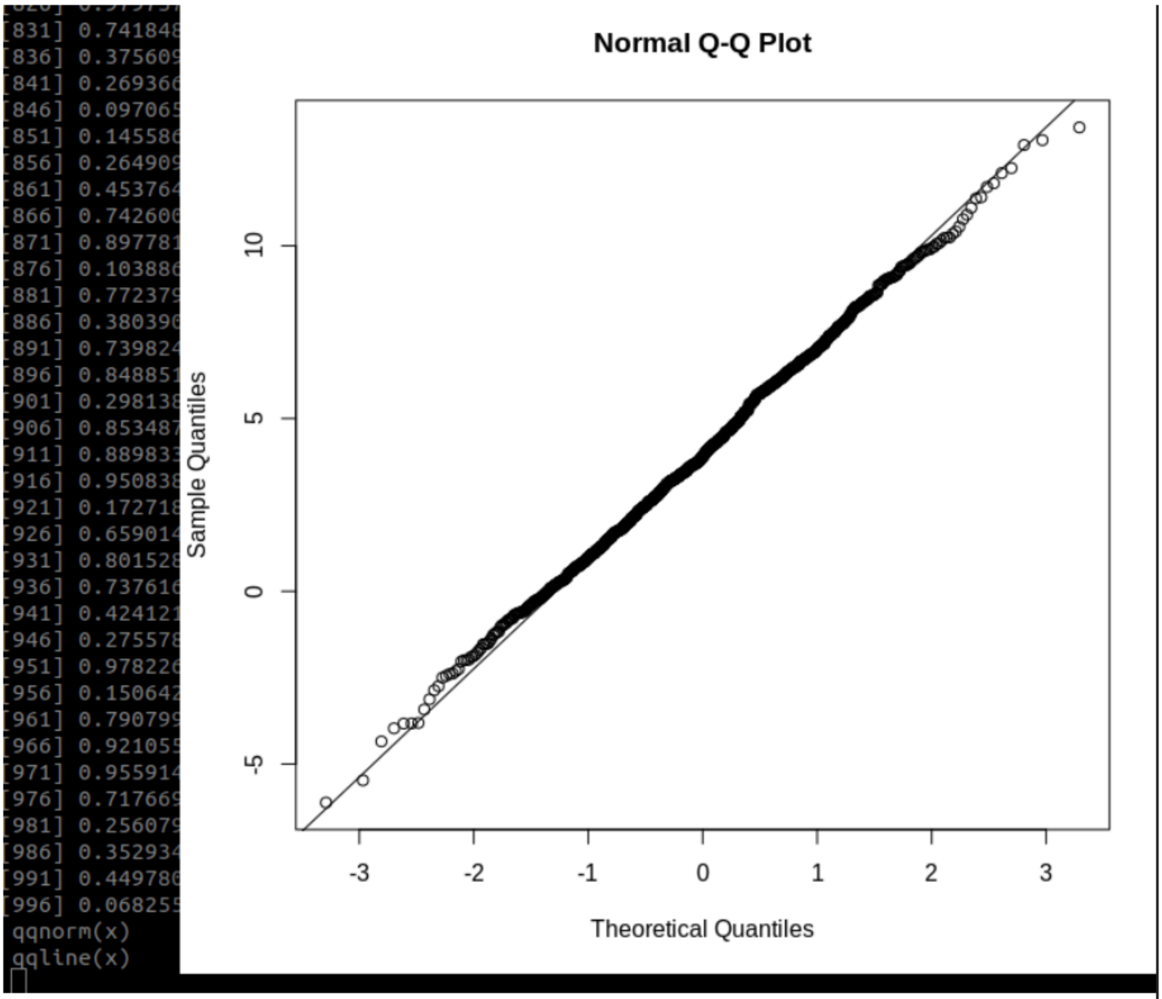
De esta manera defino de forma unívoca la distribución normal.

Curve me genera una gráfica de la función distribución de la distribución normal (Teórica con infinitos datos..), en función de x, con la media y desviación calculadas previamente.

Con el Q-QPlot —> Quantile-Quantile Plots - podemos ver gráficamente por medio de qqnorm y qqline.

Aquí podemos comprobar que las observaciones siguen una distribución normal siempre que los puntos de la gráfica de la distribución real "siguen" bastante bien la línea de la gráfica de la distribución teórica.

> qqnorm(x) # calculo de qq normal para x - distribución real -
> qqline(x) # adiciona la qqline al plot - distribución teórica



D. Inference

5. Some studies suggest that there is a relationship between the cheese's flavour and their chemical composition, especially with the lactic acid content.
   The lactic acid content has been measured in ten cheeses, resulting the following values:
   0.86, 1.53, 1.57, 1.81, 0.99, 1.09, 1.29, 1.78, 1.29, 1.58

Assuming that the lactic acid content can be modelled as a Gaussian distributed random variable:

- Estimate a value for the mean μ and variance σ^2
- Compute a confidence interval for the mean μ ($\alpha = 0.05$)
- Compute a confidence interval for the variance σ^2 ($\alpha = 0.05$)
- Solve the following hypothesis test ($\alpha = 0.05$):

$$H_0 : \mu = 1$$
$$H_1 : \mu \neq 1$$

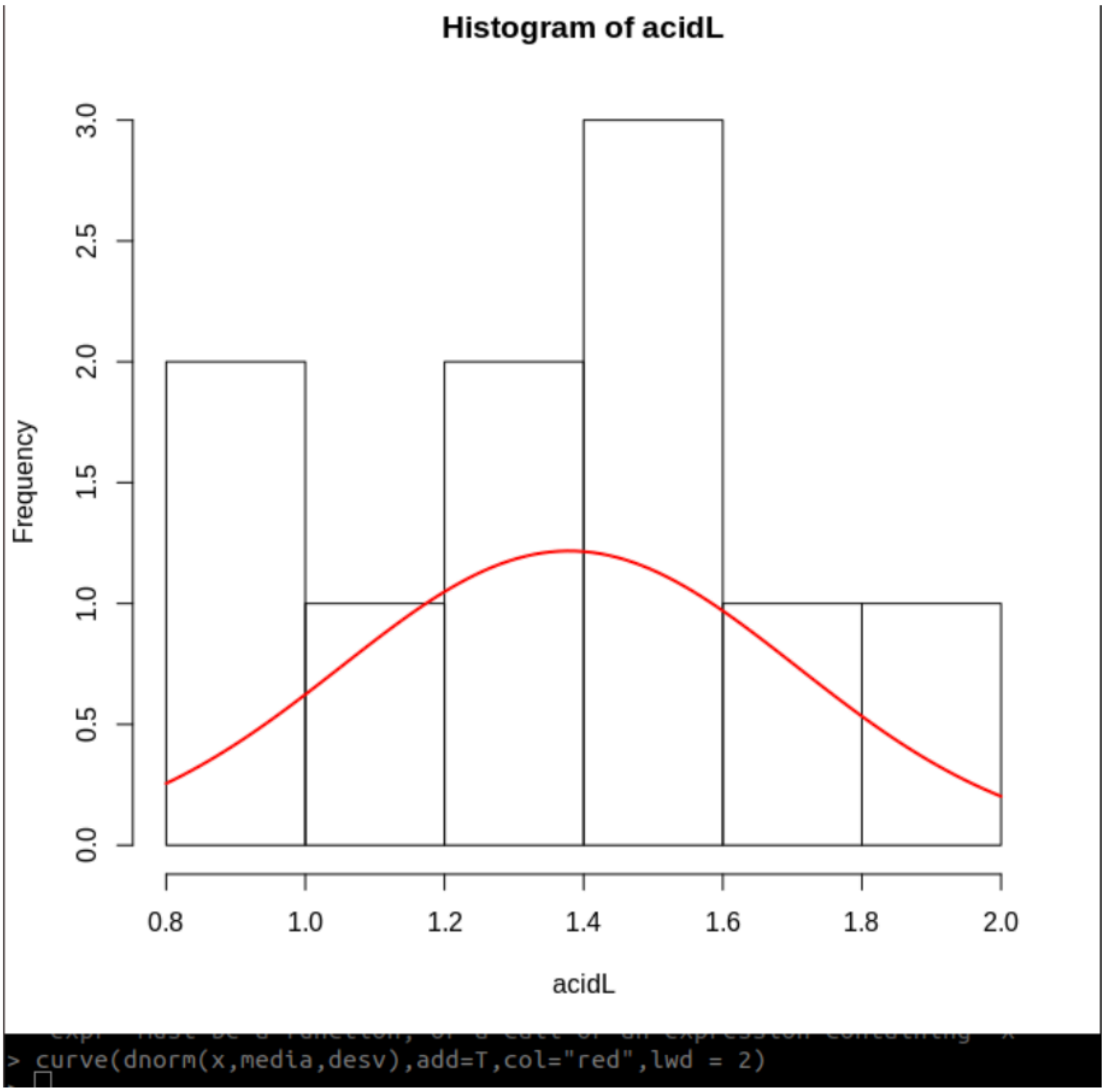
- Estimate a value for the mean and variance

```
> acidL # lactic acid

> acidL = c( 0.86, 1.53,1.57, 1.81, 0.99, 1.09, 1.29, 1.78, 1.29, 1.58 )

> acidL
 [1] 0.86 1.53 1.57 1.81 0.99 1.09 1.29 1.78 1.29 1.58

> summary (acidL)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.860   1.140   1.410   1.379   1.577   1.810 

> media = mean(acidL)
> media
[1] 1.379

> var(acidL)
[1] 0.1073656

> sqrt(var(acidL))
[1] 0.3276668
```

> desv = sd(acidL)
> desv
[1] **0.3276668**

> n = length(acidL)
> n
[1] **10**

> hist(acidL) —> curva de la muestra
> curve(dnorm (x, media, desv), add=T, col="red", lwd = 2) —> curva de la distribución normal



-Compute a confidence interval for the mean μ (σ = 0,05)

$$IC(\mu) = \bar{x} \pm \hat{S} \cdot t_{n-1} \cdot 1/\sqrt{n}$$

> # Intervalo de confianza de la media con un 90% es:

media + desv * (tstudent (0.05) , (n-1)) * (1 / sqrt (n))
media - desv * (tstudent (0.05) , (n-1)) * (1 / sqrt (n))

> media-desv*qt(0.05,n-1)/sqrt(n)
[1] 1.568942

> media+desv*qt(0.05,n-1)/sqrt(n)
[1] 1.189058

> 1.568942-1.189058
[1] **0.379884**

Con un **90%** de confianza puedo afirmar que la media del acidL está entre el **1.568942-1.189058**, con una precisión aproximada de: **0.379884.**

Sé que entre 1.568942-1.189058 está la media del acidL # lactic acid, con una precisión aproximada de 0.379884.

-Compute a confidence interval for the variance sigma (variance = 0,05)

$$IC(\sigma^2) = \left(\frac{(n-1)\hat{S}^2}{\chi^2_b} , \frac{(n-1)\hat{S}^2}{\chi^2_a} \right)$$

> # Intervalo de confianza para la varianza

(n-1) * desv ^ 2 / qchisq (1-0.05 , n-1)
(n-1) * desv ^ 2 / qchisq (0.05 , n-1)

> (n-1) * desv ^ 2 / qchisq (1 - 0.05, n-1)
[1] **44.76498**

> (n-1) * desv ^ 2 / qchisq (0.05, n-1)
[1] **60.25382**

Con un **90%** de confianza puedo afirmar que la la varianza sigma del acidL está entre **44.76498 y 60.25382**

Con un **90%** de confianza puedo afirmar que (sigma), esta entre estos valores

> Xa = qchisq(0.05 , n-1)
> Xa
[1] 210.6873

> Xb = qchisq(1 - 0.05 , n-1)
> Xb
[1] 283.5858

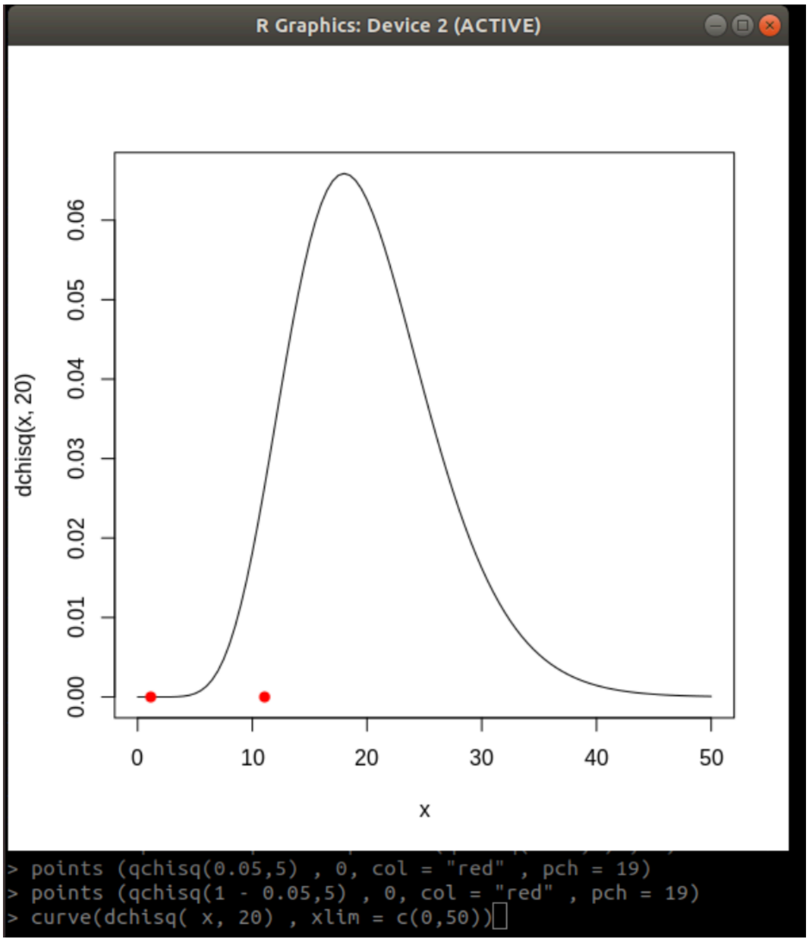
> (n-1) * desv ^ 2 / Xb
[1] 44.76498

> (n-1) * desv ^ 2 / Xa
[1] 60.25382

curve(dchisq (x, 20) , xlim = c(0, 50))

points (qchisq(0.05,5) , 0, col = "red" , pch = 19)

points (qchisq(1 - 0.05,5) , 0, col = "red" , pch = 19)



- Solve the following hypothesis test ($\alpha = 0.05$):

$$H_0 \;:\; \mu = 1$$
$$H_1 \;:\; \mu \neq 1$$

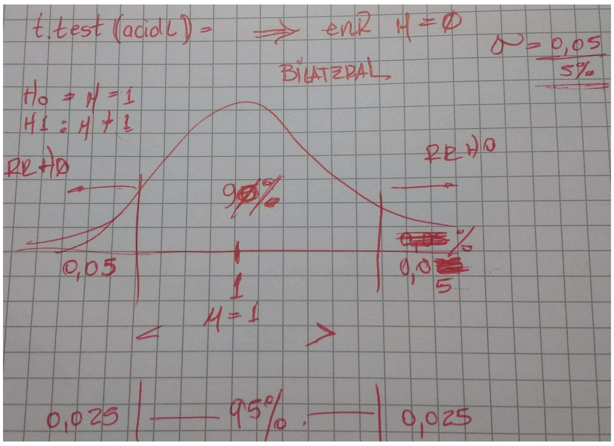
Queremos comprobar que la **μ** = 1 ó es diferente de 1 =>

Seguramente para cualquier muestra que seleccione.., la **μ** va a estar a la derecha ó a la izquierda de este valor. - Por lo tanto es un Contraste Bilateral, ya que tiene una región de rechazo. RRH0 a cada lado, estas serán las regiones en las que puedo decir que **μ** no es igual a 1.

Es decir, si **H0 : μ = 1**, y suponiendo que esto es cierto, entonces siempre va a dar una 1 < **μ** < 1; por tanto **H1: μ ≠ 1.**

Si **α** = 0.05, entonces se establece un límite de 90% para el cual **μ = 1**

Si quiero saber si **μ ≠ 1**



> t.test(acidL) **μ = 0**

One Sample t-test

data: acidL
t = 13.309, df = 9, p-value = 3.174e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1.144601 1.613399
sample estimates:
mean of x
 1.379

> t.test(acidL, **mu = 1**, conf.level = 0.95, alternative = "two.sided") => **μ = 1**

One Sample t-test

data: acidL
t = 3.6577, df = 9, **p-value = 0.005254 < 0.05 => p-value < α => rechazo H0**
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 1.144601 1.613399
sample estimates:
mean of x
 1.379

DATA ANALYTICS: DESCRIPTIVE, INFERENCE, AND PREDICTIVE TECHNIQUES

Student declaration

I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.

Student signature: Date: