1. Let X be a normal random variable with mean $\mu = -1.25$ and variance $\sigma^2 = 0.36$. Find

   a) $P[X > -1.9]$ `1-pnorm(-1.9,-1.25,0.60) 0.8606698`

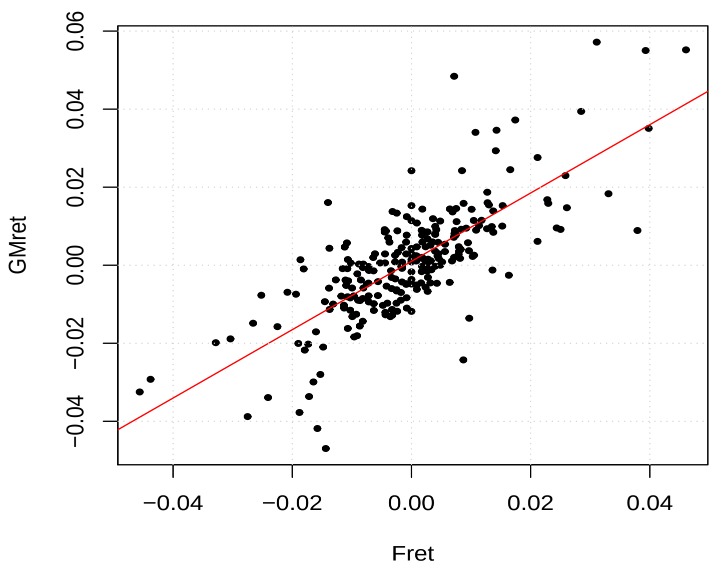
   b) $x_{0.25}$ (the 25^{th} percentile of X) `qnorm(0.25,-1.25,0.6) -1.654694`

2. Download stock prices from Ford and General Motors from August 25, 2016 to August 25, 2017.

   a) Find the correlation of the adjusted closing daily prices
   b) Find the correlation of the daily returns (based on adj.Closed Prices)
   c) Find kurtosis of GM daily returns
   d) Create a scatter plot of daily returns with a least squares line on it. What is the slope of that line?

```
setwd("C:/Users/USC Guest/Downloads2")
d0 = read.csv("Ford.csv",header=T)
n = nrow(d0)
head(d0)#         Date  Open  High   Low Close Adj.Close   Volume
# 1 2016-08-25 12.30 12.52 12.21 12.47  11.80837 44713700
# 2 2016-08-26 12.47 12.55 12.34 12.38  11.72315 22645900
# 3 2016-08-29 12.38 12.50 12.38 12.47  11.80837 22243200
# 4 2016-08-30 12.47 12.56 12.43 12.55  11.88413 26040400
# 5 2016-08-31 12.48 12.61 12.48 12.60  11.93148 26030600
# 6 2016-09-01 12.66 12.72 12.35 12.44  11.77997 40510400

d1 = read.csv("GMotors.csv",header=T)
head(d1)
#         Date  Open  High   Low Close Adj.Close   Volume
# 1 2016-08-25 31.71 31.75 31.36 31.54  30.18328 15930800
# 2 2016-08-26 31.67 31.82 31.42 31.53  30.17371 10436200
# 3 2016-08-29 31.78 31.96 31.57 31.81  30.44167  7708300
# 4 2016-08-30 31.81 32.00 31.23 31.67  30.30769 10748700
# 5 2016-08-31 31.67 31.94 31.65 31.92  30.54694  9625600
# 6 2016-09-01 31.86 32.55 31.34 31.80  30.43210 14898800

Fprice = d0$Adj.Close
Fret = Fprice[2:n]/Fprice[1:(n-1)] - 1
GMprice = d1$Adj.Close
GMret = GMprice[2:n]/GMprice[1:(n-1)] - 1
```

```
# a) correlation of Ford and GM prices
cor(Fprice,GMprice)
#[1] 0.3610031

# [1] 0.3610031

# b) correlation of Ford and GM returns
cor(Fret,GMret)
# [1] 0.7543391

# c) kurtosis of GM returns
library(moments)
kurtosis(GMret)
# [1] 5.917187

# d) scatterplot
plot(GMret~Fret,pch=19,cex=0.6)
m1 = lm(GMret~Fret)
abline(m1,col="red")
grid()

coef(m1)
# (Intercept)         Fret
#0.0009955661 0.8757107246

# slope is 0.8757
```



3. The dataframe VIT2005 in the PASWR2 package contains descriptive information and the appraised total price (in euros) for apartments in Vitoria, Spain.

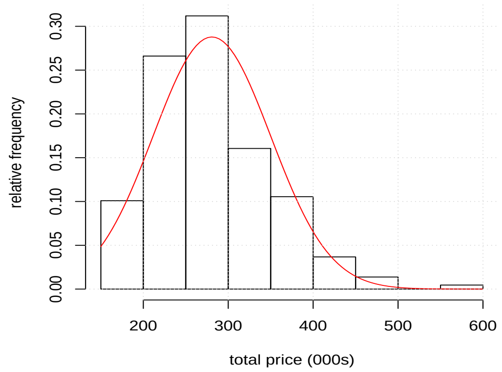
   a) How many one-garage apartments have a `totalprice` greater than 400,000 euros?

   b) Build a histogram of relative frequencies of `totalprice`. Add a normal density curve to the plot.

   c) Find the covariance and correlation matrix between all numerical variables (use `str()` to identify which variables are numerical).

   d) Make a scatterplot of `totalprice` and `area`. Report the row number, `totalprice` and `area` of the largest outlier.

```
library(PASWR2)
d0=VIT2005

# a) How many one-garage apartments have a totalprice greater than 400,000 euros?

d1 = subset(d0,subset = garage==1 & totalprice > 400000)
#      totalprice   area zone category age floor rooms  out conservation toilets garage elevator
# 2        409000 100.65  Z31       3B   5     7     5  E50           1A       2      1        1
# 13       560000 155.90  Z21       2B   7     4     6 E100           1A       2      1        1
# 34       403000 118.86  Z41       3A  25     5     6 E100           1A       2      1        1
# 48       426200 141.48  Z42       3A  13     5     5 E100           1A       2      1        1
# 133      457000 142.18  Z42       3A  14     4     6 E100           2A       2      1        1
# 186      433500 113.51  Z41       2B   8     4     5 E100           1A       2      1        1

# b) Build a histogram of relative frequencies of totalprice.
# Add a normal density curve to the plot

tprice = d0$totalprice/1000
h1=hist(tprice)
h1$counts = h1$counts/sum(h1$counts)
width1 = h1$breaks[2]-h1$breaks[1]
mu = mean(tprice)
stdev = sd(tprice)
plot(h1,ylab="relative frequency",xlab="total price (000s)",main="")
curve(dnorm(x,mu,stdev)*width1,col="red",add=T)
grid()
```



```
# c) Find the covariance and correlation matrix between all numerical variables
# (use str() to identify which variables are numerical).

str(d0)
'data.frame':   218 obs. of  15 variables:
 $ totalprice    : num  228000 409000 200000 180000 443600 ...
 $ area          : num  75.3 100.7 88.9 62.6 146.1 ...
 $ zone          : Factor w/ 23 levels "Z11","Z21","Z31",..: 14 3 19 23 3 1 16 17 1 1 ...
 $ category      : Factor w/ 7 levels "2A","2B","3A",..: 6 4 3 5 3 6 3 4 6 4 ...
 $ age           : int  33 5 14 41 22 35 14 36 37 11 ...
 $ floor         : int  3 7 8 3 6 4 6 3 4 5 ...
 $ rooms         : int  5 5 5 4 7 5 4 4 4 4 ...
 $ out           : Factor w/ 4 levels "E100","E25","E50",..: 1 3 3 3 1 3 3 1 2 3 ...
 $ conservation  : Factor w/ 4 levels "1A","2A","2B",..: 3 1 1 2 1 1 1 4 2 1 ...
 $ toilets       : int  1 2 2 1 2 1 1 1 1 1 ...
 $ garage        : int  0 1 0 0 0 0 0 0 0 0 ...
 $ elevator      : int  1 1 1 0 1 0 1 0 0 0 ...
 $ streetcategory: Factor w/ 4 levels "S2","S3","S4",..: 2 4 1 2 3 3 2 2 3 3 ...
 $ heating       : Factor w/ 4 levels "1A","3A","3B",..: 2 4 2 1 4 2 4 2 2 2 ...
 $ storage       : int  0 1 0 0 1 1 1 1 1 1 ...

# cols 1,2,5,6,7,10,11,12,15 are numerical
d2 = d0[,c(1,2,5,6,7,10,11,12,15)]
options(digits=4)
cov(d2)
#            totalprice       area        age      floor      rooms    toilets     garage   elev
# totalprice 4802276444  1.163e+06 -2.756e+05 4306.92948  2.310e+04  2.383e+04  1.636e+04  1.424
# area          1162639  4.299e+02 -1.582e+01    3.60091  8.392e+00  6.485e+00  3.286e+00  3.292
# age          -275570 -1.582e+01  2.130e+02   -2.52217 -7.660e-01 -2.049e+00 -1.409e+00 -2.403
# floor           4307  3.601e+00 -2.522e+00    4.52408  1.755e-01  1.110e-01 -2.921e-02  1.722
# rooms          23104  8.392e+00 -7.660e-01    0.17554  4.023e-01  1.335e-01  5.889e-02  9.005
# toilets        23829  6.485e+00 -2.049e+00    0.11098  1.335e-01  2.501e-01  1.023e-01  9.026
# garage         16365  3.286e+00 -1.409e+00   -0.02921  5.889e-02  1.023e-01  2.033e-01  4.930
# elevator       14244  3.292e+00 -2.403e+00    0.17224  9.005e-02  9.026e-02  4.930e-02  1.618
# storage         7537  1.274e+00 -1.582e+00   -0.10371  8.456e-03  4.448e-02  2.862e-02  3.467

cor(d2)
#            totalprice     area      age    floor    rooms toilets   garage elevator  storage
# totalprice    1.00000  0.80921 -0.27245  0.02922  0.52563  0.6876  0.52374   0.5109  0.26736
# area          0.80921  1.00000 -0.05226  0.08166  0.63817  0.6254  0.35154   0.3946  0.15109
# age          -0.27245 -0.05226  1.00000 -0.08124 -0.08275 -0.2808 -0.21403  -0.4092 -0.26641
# floor         0.02922  0.08166 -0.08124  1.00000  0.13011  0.1043 -0.03046   0.2013 -0.11986
# rooms         0.52563  0.63817 -0.08275  0.13011  1.00000  0.4209  0.20593   0.3529  0.03277
# toilets       0.68757  0.62543 -0.28075  0.10433  0.42089  1.0000  0.45373   0.4486  0.21862
# garage        0.52374  0.35154 -0.21403 -0.03046  0.20593  0.4537  1.00000   0.2718  0.15605
# elevator      0.51094  0.39464 -0.40924  0.20129  0.35291  0.4486  0.27177   1.0000  0.21184
# storage       0.26736  0.15109 -0.26641 -0.11986  0.03277  0.2186  0.15605   0.2118  1.00000
```

```
# d) Make a scatterplot of totalprice and area.
# Report the row number, totalprice and area of the largest outlier.

plot(totalprice~area,d0)
m1 = lm(totalprice~area,d0)
abline(m1,col="red")
identify(d0$area,d0$totalprice,rownames(d0))

# rows 44,31 are outliers

d0[44,c(1,2)]
#    totalprice   area
# 44     178000 108.44

d0[31,c(1,2)]
#    totalprice   area
# 31     407000 187.91
```