LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

MASTER IN ACTUARIAL SCIENCE

Risk Models

16/01/2019

1st part of the exam

Time allowed: 2 hours

Instructions:

1. This paper contains 7 questions and comprises 3 pages including the title page.
2. Enter all requested details on the cover sheet.
3. You have 10 minutes of reading time. You must not start writing your answers until instructed to do so.
4. Number the pages of the paper where you are going to write your answers.
5. Attempt all 7 questions.
6. Begin your answer to each of the 7 questions on a new page.
7. Marks are shown in brackets. Total marks: **140**.
8. Show calculations where appropriate.
9. An approved calculator may be used.
10. The distributed formulary and the Formulae and Tables for Actuarial Examinations (the 2002 edition) may be used. Note that the parametrization used for the different distributions is that of the distributed formulary.

1. **[10]** A random sample of 62 students of a large university originates a positive correlation of 0.53 between the marks in Statistics and in Economics. At the 1% level of significance, can we confirm a positive correlation between the marks for these two courses in the population?

2. A random sample of 70 claim amounts (after sorting) is

36.6	36.6	40.1	41.8	42.9	44.4	44.7	46	46.3	46.7
51.6	52.9	54.9	55.5	55.8	56.5	56.6	57.3	61.3	61.6
63.1	64.1	68.6	69	69.5	70	72.6	75.9	78.8	80.1
81.7	82.8	83.6	84.5	85.8	89.1	90.1	92.9	93.1	97.1
98.5	101.8	103.6	104.3	104.6	105.4	107.9	109.1	109.6	110.5
111.3	116	118.6	118.8	121.2	123.6	124.5	124.9	126.7	134.2
134.4	134.6	137.8	140.8	145.1	153	155.8	159.1	167.7	176.7

   a. **[10]** Obtain a non-parametric estimate for the probability that a claim amount is between 50 (excluded) and 90 (included) and compute a 90% confidence interval for this probability.

   b. **[15]** Obtain a non-parametric estimate for the probability that a claim amount is less than or equal to 90 given that it is greater than 50. Explain why we cannot obtain the variance of the estimator without introducing an additional assumption.

3. **[15]** The time in years from the onset of a disease to death has been observed for a random sample of 10 lives originating the following values (3; 4; 4; 4; 5; 5; 5; 6; 6; 8). Using the kernel

$$k_y(x)=\begin{cases}(x-y+2)/4 & y-2\le x\le y\\ (y+2-x)/4 & y\le x\le y+2\\ 0 & \text{otherwise}\end{cases}$$

   estimate the density function at 5.5.

4. Suppose X is a discrete population with probability function given by $f_X(x\,|\,\theta)=\theta(1-\theta)^{x-1}$, $x=1,2,\cdots$, $0<\theta<1$, from which we observed a random sample with size n. We also know that $E(X)=1/\theta$ and $\text{var}(X)=(1-\theta)/\theta$.

   a. **[10]** Obtain the maximum likelihood estimator for θ.

   b. **[10]** Assuming that $n=100$ and that $\sum_{i=1}^{n}x_i=250$ get a ML estimate for θ and calculate a 95% approximate confidence interval for θ (If you were unable to obtain the ML estimator explain how to get the required confidence interval).

c. **[15]** Now assume that a Bayesian approach is used and that the prior is given by $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1/2}$, $0<\theta<1$. Show that the posterior distribution for θ is a beta distribution with parameters n and $\sum_{i=1}^{n} x_i - n + 0.5$.

d. **[15]** Consider the sample referred to in item b). Give a Bayes estimate for θ assuming a quadratic loss function. Using Bayes Central Limit Theorem obtain an approximate 95% HPD interval for θ.

5. **[10]** Assume that the density function for a continuous population X is given by $f(x \mid \theta)$, $x>0$, $\theta>0$ and that the corresponding distribution function is $F(x \mid \theta)$. A random sample of size 25 was observed. However due to sampling limitations you only got the values of the first 3 observations (2, 7 and 12). For the remaining observations you only know the results presented in the following table

Interval	Nº of observations
$0<x\leq 10$	11
$10<x\leq 25$	6
$x>25$	5

Write the log likelihood function that should be maximized to obtain a maximum likelihood estimate of θ as a function of F and f. No additional computations are required.

6. **[15]** You are given the sorted random sample (0.76 1.95 2.36 9.84 44.72) from a population whose distribution function is assumed to be an inverse Pareto with parameters $\theta=3$ and $\tau=2$. Test, using the Kolmogorov-Smirnov test, if this assumption is acceptable ($\alpha=0.05$).

7. **[15]** Let X be a population Pareto distributed (with parameters $\alpha=3$ and $\theta=100$). Explain how to use simulation to obtain an approximation to the sampling distribution of the sample range (sample maximum minus the sample minimum) when the sample size is $n=10$.

Solutions

Exercise 1

$H_0 : \rho = 0 \quad H_1 : \rho > 0$

Test Statistic: $T = \frac{R\sqrt{60}}{\sqrt{1-R^2}} \sim t_{(60)}$

$T_{obs} = \frac{0.53 \times \sqrt{60}}{\sqrt{1-0.53^2}} = 2.7079$ p-value<0.5% or $t_{0.01} = 2.390$

Reject the null and then we confirm a positive correlation between the marks obtained in Statistics and in Economics in the population.

Exercise 2

a)

Let p be the probability that a claim amount is between 50 and 90 and let $N_{(50,90]}$ be the number of claim amounts in a random sample with size $n = 70$ whose value is between 50 (excluded) and 90 (included). $N_{(50,90]} \sim b(70, p)$ where $p = P(50 < X \leq 90)$.

The non-parametric estimator is $\hat{p} = \frac{N_{(50,90]}}{70}$ leading to the estimate $\hat{p} = \frac{n_{(50,90]}}{70} = \frac{26}{70} = 0.3714$.

The estimated variance of the estimator is given by $\hat{\text{var}}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{26 \times 44}{70^3} = 0.0033$.

Then the CI is given by $0.3714 \pm 1.645\sqrt{0.0033}$ → (0.2764, 0.4662)

b)

Now let us define p as $p = P(X \leq 90 \mid X > 50) = \frac{P(50 < X \leq 90)}{P(X > 50)}$. Let us use $N_{(50,90]}$ as defined in item a) and define $N_{(50,\infty)}$ as the number of claim amounts in a random sample with size $n = 70$ whose value is greater than 50. $N_{(50,\infty)} \sim b(70, P(X > 50))$. The estimator is given by $\hat{p} = \frac{\hat{P}(50 < X \leq 90)}{\hat{P}(X > 50)} = \frac{N_{(50,90]}/n}{N_{(50,\infty)}/n} = \frac{N_{(50,90]}}{N_{(50,\infty)}}$ and then the estimate is

$\hat{p} = \frac{26}{60} = 0.4333$.

The variance (and even the expected value) of this estimator is infinite (and then doesn't exist) because $N_{(50,\infty)}$ can be 0 with a positive probability.

Exercise 3

Sample (3; 4; 4; 4; 5; 5; 5; 6; 6; 8) →

j	1	2	3	4	5
y_j	3	4	5	6	8
$p(y_j)$	0.1	0.3	0.3	0.2	0.1

$$\hat{f}(5.5)=\sum_{j=1}^{5}p(y_j)\,k_{y_j}(5.5)=0.1\times 0+0.3\times\frac{4+2-5.5}{4}+0.3\times\frac{5+2-5.5}{4}+0.3\times\frac{5.5-6+2}{4}+0.1\times 0$$
$$=0.3\times 0.125+0.3\times 0.375+0.3\times 0.375=0.225$$

Exercise 4

a)

$$\ell(\theta)=\sum_{i=1}^{n}\ln f_X(x_i\mid\theta)=\sum_{i=1}^{n}\ln\left(\theta(1-\theta)^{x_i-1}\right)=\sum_{i=1}^{n}\left(\ln\theta+(x_i-1)\ln(1-\theta)\right)$$
$$=n\ln\theta+\left(\sum_{i=1}^{n}x_i-n\right)\ln(1-\theta)$$

$$\ell'(\theta)=\frac{n}{\theta}-\frac{\sum_{i=1}^{n}x_i-n}{1-\theta}$$

$$\ell'(\theta)=0\Leftrightarrow\frac{n}{\theta}=\frac{\sum_{i=1}^{n}x_i-n}{1-\theta}\Leftrightarrow\frac{1-\theta}{\theta}=\frac{\sum_{i=1}^{n}x_i-n}{n}\Leftrightarrow\frac{1}{\theta}-1=\overline{x}-1\Leftrightarrow\theta=\frac{1}{\overline{x}}$$

As $\sum_{i=1}^{n}x_i\geq n$ ($x_i\geq 1$, $i=1,2,\cdots,n$) we can guarantee that $\ell''(\theta)=-\frac{n}{\theta^2}-\frac{\sum_{i=1}^{n}x_i-n}{(1-\theta)^2}<0$

and the ML estimator for θ is $\hat{\theta}=\overline{X}^{-1}$

b)

The estimate is then $\hat{\theta}=\frac{100}{250}=0.4$

$\hat{\text{var}}(\hat{\theta})\approx I_{X_1,X_2,\cdots X_n}(\hat{\theta})^{-1}$ or $\hat{\text{var}}(\hat{\theta})\approx-\ell''(\hat{\theta})^{-1}$ can be used.

$$I_{X_1X_2,\cdots,X_n}(\theta)=-E\left(\ell''(\theta)\right)=-E\left(-\frac{n}{\theta^2}-\frac{\sum_{i=1}^{n}X_i-n}{(1-\theta)^2}\right)=\frac{n}{\theta^2}+\frac{(n/\theta)-n}{(1-\theta)^2}$$
$$=\frac{n}{\theta^2}+\frac{n(1-\theta)/\theta}{(1-\theta)^2}=\frac{n}{\theta^2}+\frac{n}{(1-\theta)\theta}=n\left(\frac{(1-\theta)+\theta}{(1-\theta)\theta^2}\right)=\frac{n}{(1-\theta)\theta^2}$$

$$I_{X_1X_2,\cdots,X_n}(\hat{\theta})=\frac{n}{(1-\hat{\theta})\hat{\theta}^2}=\frac{100}{0.6\times 0.16}=\frac{100}{0.096}$$ and then $\hat{\text{var}}(\hat{\theta})\approx\frac{0.096}{100}=0.00096$

$$\hat{\text{var}}(\hat{\theta})\approx\left(-\ell''(\hat{\theta})\right)^{-1}=\left(\frac{n}{\theta^2}+\frac{\sum_{i=1}^{n}x_i-n}{(1-\theta)^2}\right)^{-1}=\left(\frac{100}{0.16}+\frac{150}{0.36}\right)^{-1}=\left(\frac{36+24}{0.0576}\right)^{-1}=\frac{0.0576}{60}=0.00096$$

The 95% approximate confidence interval for θ is then given by $0.4\pm 1.96\sqrt{0.00096}$, i.e. (0.3393, 0.4607)

c)

$$\pi(\theta)\propto\theta^{-1}(1-\theta)^{-1/2}$$

$$L(\theta)=\prod_{i=1}^{n}f_X(x_i\mid\theta)=\prod_{i=1}^{n}\theta(1-\theta)^{x_i-1}=\theta^n(1-\theta)^{\sum x_i-n}$$

$\pi_{\underline{X}}(\theta)\propto\theta^{-1}(1-\theta)^{-1/2}\,\theta^n(1-\theta)^{\sum x_i-n}=\theta^{n-1}(1-\theta)^{\sum x_i-n-1/2}$ which is the core of a beta distribution with parameters n and $\sum_{i=1}^{n}x_i-n+0.5$.

d)

Bayes estimate against a quadratic loss function → Expected value of the posterior

$$\theta^B = E(\theta \mid \underline{x}) = \frac{n}{n + \sum x_i - n + 0.5} = \frac{100}{250.5} = 0.399$$

$$\text{var}(\theta \mid \underline{x}) = \frac{n\left(\sum_{i=1}^{n} x_i - n + 0.5\right)}{\left(\sum_{i=1}^{n} x_i + 0.5\right)^2 \left(\sum_{i=1}^{n} x_i + 1.5\right)} = \frac{100 \times 150.5}{250.5^2 \times 251.5} = 0.0009536$$

The HPD interval is then given by $0.399 \pm 1.96\sqrt{0.0009536}$, i.e. (0.3387, 0.4597)

Exercise 5

The contribution to the log likelihood of each of the first 3 observation is given by the log of the density function while for the remaining observations it si given by the log of the probability of the corresponding intervals.

$$\ell(\theta) = \ln f(2 \mid \theta) + \ln f(7 \mid \theta) + \ln f(12 \mid \theta) +$$
$$+11 \ln F(10 \mid \theta) + 6 \ln\left(F(25 \mid \theta) - F(10 \mid \theta)\right) + 5 \ln\left(1 - F(25 \mid \theta)\right)$$

Exercise 6

$$H_0 : X \sim F(x) = \left(\frac{x}{x+3}\right)^2 \qquad H_1 : H_0 \text{ is false}$$

i	$x_{(i)}$	$F_X(x_{(i)})$	$F_n(x_{(i)}^-)$	$F_n(x_{(i)})$	Max diff
1	0.76	0.04086	0	0.2	0.15914
2	1.95	0.15519	0.2	0.4	0.24482
3	2.36	0.19386	0.4	0.6	**0.40614**
4	9.84	0.58730	0.6	0.8	0.21270
5	44.72	0.87823	0.8	1	0.12178

$$D_n = 0.40614$$

The asymptotic critical value at 5% is $D_{5,0.05} \approx 1.36/\sqrt{5} = 0.60821$

As $D_n < D_{5,0.05}$ we do not reject the null, i.e. we do not reject that the distribution is an inverse Pareto with parameters $\theta = 3$ and $\tau = 2$.

Exercise 7

- Define NR, the number of replicas and define, if necessary an array $\mathbf{r}$ with size NR.
- For each replica, $j = 1, 2, \cdots, NR$
  - Generate 10 Pareto distributed variables $(x_1, x_2, \cdots, x_n)$ To generate each x_i: generate u_i as a Uniform(0,1) random variable and use the inverse method, i.e. compute $x_i = \theta\left((1-u_i)^{-1/\alpha} - 1\right)$.
  - Compute $r_i = x_{(n)} - x_{(1)}$ and keep the value as element i of the array $\mathbf{r}$. You can do it after sorting the sample or you can use the functions max and min.
- The NR elements of array $\mathbf{r}$ are used to get an approximation of the required distribution (histogram, kernel density estimation, ….)

MASTER IN ACTUARIAL SCIENCE

Risk Models

16/01/2019

2nd part of the exam

Time allowed: 1 hour

Instructions:

1. This paper contains 3 questions and comprises 3 pages including the title page.
2. Enter all requested details on the cover sheet.
3. You have 10 minutes of reading time. You must not start writing your answers until instructed to do so.
4. During the reading time you can download and install all the R packages that you need to answer the questions. Once the reading time is over no more access to the internet is allowed.
5. You are requested to summarize your answers on the cover sheet. You can add, using another paper sheet, any comments you think necessary to understand your answers. At the end of the exam you should submit your R files to Aquila using your usual username and password.
6. Attempt all questions.
7. Marks are shown in brackets. Total marks: **60**.
8. The distributed formulary and the Formulae and Tables for Actuarial Examinations (the 2002 edition) may be used. Note that the parametrization used for the different distributions is that of the distributed formulary.

1. File decathlon.csv presents the results of the 2017 World Championship in Athletics – Men's decathlon in London. The decathlon is a combined event in athletics consisting of ten track and field events. Depending on the performance in each event the athletes win a given number of points. At the end, the one with the higher number of points is the winner. The events are:

    1) 100 m race (seconds)
    2) Long Jump (meters)
    3) Shot Put (meters)
    4) High Jump (meters)
    5) 400 m race (seconds)
    6) 110 m hurdles race (seconds)
    7) Discus throw (meters)
    8) Pole vault (meters)
    9) Javelin throw (meters)
    10) 1500 m race (seconds)

a. **[10]** Assuming that the dataset corresponds to a random sample of athletes (which is obviously not the case), test ($\alpha = 0.05$) if we can consider that the correlation between the time needed to complete the 100m race is positively correlated to the time needed to complete the 400m race.

b. **[5]** To carry on a PCA should we scale (or not) the data set? Explain. Answer to the next question according to your answer.

c. **[15]** Now, run a PCA.

    i. How many principal components should you use according to Kaiser's criterion?

    ii. Interpret the meaning of the second principal component.

    iii. Present the coordinates of the first athlete, Kévin Mayer, using the retained principal components.

2. Assume that X, the claim amounts for a given risk, is gamma distributed with unknown parameters α and θ. From this population we collect a random sample (see file gamma.csv).

    a. **[10]** Obtain the maximum likelihood estimates for α and θ and determine a 95% confidence interval for each parameter.

    b. **[5]** Obtain a maximum likelihood estimate for the probability that a claim is greater than 25.

3. **[15]** Return to the last exercise of the first part of the exam. The problem was how to use simulation to obtain an approximation to the sampling distribution of the sample range (sample maximum minus the sample minimum). Assume that you choose to use 1000 replicas and define the following approach:

   - Define the number of replicas as $NR = 1000$ and define an array $\mathbf{r}$ with size NR.
   - For each replica, $j = 1, 2, \cdots, NR$
     - Generate 10 Pareto distributed variables $(x_1, x_2, \cdots, x_n)$. To generate each x_i: generate u_i as a Uniform(0,1) random variable and use the inverse method, i.e. compute $x_i = \theta\left((1-u_i)^{-1/\alpha} - 1\right)$.
     - Compute $r_i = x_{(n)} - x_{(1)}$ and keep the value as element i of the array $\mathbf{r}$.

   Apply this approach using R and characterize the sampling distribution presenting an histogram of the sampling distribution. Also compute the mean and the standard deviation of the range. What is the probability that the sample range is higher than 1500?

Solutions of the second part

Exercise 1

a)

$H_0 : \rho = 0 \qquad H_1 : \rho > 0$

$p - value = 3.241 \times 10^{-5}$**,** we strongly reject the null. The time needed to complete the 100m race is positively correlated with the time needed to complete the 400m race.

b)

Yes, we should scale the data as we are dealing with a set of variables measured in different units and with different sizes (for instance, look at the Pole vault and the 1500m)

c)

According to Kaiser criterion (eigenvalues greater than 1) we should retain 4 PC.

Using the standardized loadings, we can see that the second PC is mostly linked to the variables "Shot Put", "Discus Throw" and "Javelin Throw".

The coordinates are (-1.623, 2.505, 1.122, -0.182)

Exercise 2

a)

$\hat{\alpha} = 5.529$, $\hat{\theta} = 3.577$ and $\hat{\text{var}}(\hat{\alpha}, \hat{\theta}) = \begin{bmatrix} 0.2897 & -0.1875 \\ -0.1875 & 0.1329 \end{bmatrix}$. Then, the 95% asymptotic confidence intervals for each parameter are:

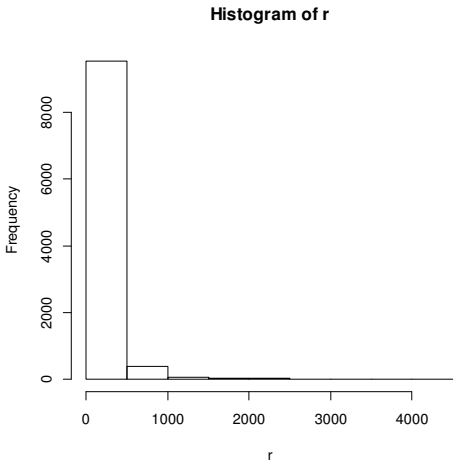
α: (4.4742, 6.5841) θ : (2.8627, 4.2919)

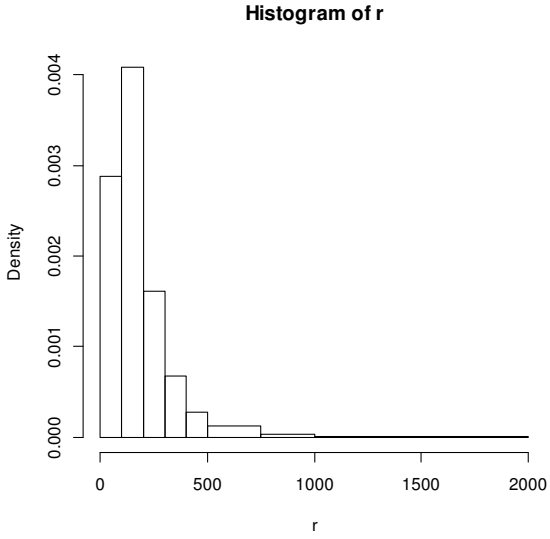
b)

$\hat{P}(X > 25) = 1 - F(25 \mid \alpha = 5.529, \theta = 3.577) = 0.238$ where F is a gamma distribution function

Exercise 3

```
> NR=10000; alpha=3; theta=100; n=10
>
> r=rep(NA,NR)
> for(i in 1:NR){
+   u=runif(n); x=theta*((1-u)^(-1/alpha)-1)
+   r[i]=max(x)-min(x)
+   }
>
> mean(r); sd(r); hist(r)
[1] 190.676
[1] 183.4617
> print("Prob(range>1500)");mean(r>1500)
[1] "Prob(range>1500)"
[1] 0.0026
>
```





$\mu \approx 190.6760$ $\sigma \approx 183.4617$ $P(R > 1500) \approx 0.0026$

R programs and outputs – Part 2

Exercise 1

```
> direct="C:/Users/joaoas/Desktop/trial/"
> file="decathlon.csv"
>
> dta=read.csv(paste(direct,file,sep=""),header=T,sep=";") # read data set
> # Check reading
> head(dta)
              Name        Country run100m LongJump ShotPut HighJump run400m
1      Kévin Mayer   France (FRA)   10.70     7.52   15.72     2.08   48.26
2    Rico Freimuth  Germany (GER)   10.53     7.48   14.85     1.99   48.41
3     Kai Kazmirek  Germany (GER)   10.91     7.64   13.78     2.11   47.19
4    Janek Õiglane  Estonia (EST)   11.08     7.33   15.13     2.05   49.58
5    Damian Warner   Canada (CAN)   10.50     7.44   13.45     2.02   47.47
6 Oleksiy Kasyanov  Ukraine (UKR)   10.77     7.28   14.99     2.02   48.64
  run110mHurdles DiscusThrow PoleVault JavelinThrow run1500m
1          13.75       47.14       5.1        66.10    276.7
2          13.68       51.17       4.8        62.34    281.6
3          14.66       45.06       5.1        62.45    278.1
4          14.56       42.11       5.1        71.73    279.2
5          13.63       40.67       4.7        56.63    268.4
6          14.05       48.79       4.7        50.82    273.9
> attach(dta)
>
> # a)
> cor.test(run100m,run400m,alternative="greater")

        Pearson's product-moment correlation

data:  run100m and run400m
t = 5.1666, df = 18, p-value = 3.241e-05
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.5569272 1.0000000
sample estimates:
      cor
0.7728245

>
> # c) PCA based on scaled data
>
x=cbind(run100m,LongJump,ShotPut,HighJump,run400m,run110mHurdles,DiscusThrow,PoleVault,J
avelinThrow,run1500m)
> out1=prcomp(x,center=T,scale=T)
> out1
Standard deviations (1, .., p=10):
 [1] 1.7665945 1.4611509 1.2324929 1.0469983 0.8681724 0.8071679 0.6114752
 [8] 0.4167608 0.3078143 0.2852252

Rotation (n x k) = (10 x 10):
                     PC1          PC2         PC3         PC4          PC5
run100m         0.50006092 -0.116517303  0.09925368 -0.27786509  0.005300623
LongJump       -0.47266700  0.144883059 -0.02727576  0.04954576  0.402795785
ShotPut         0.01053579  0.599461439 -0.16532487 -0.07492435  0.025738265
HighJump       -0.06998335  0.216845060  0.62673606 -0.20890343 -0.057725851
run400m         0.49517556  0.040994639  0.07923111 -0.19523288 -0.199002425
run110mHurdles  0.32619863 -0.248204254 -0.24431638  0.17893676  0.670312176
DiscusThrow     0.14659569  0.545490097 -0.17039971  0.41269998 -0.161746916
PoleVault      -0.04288229 -0.008906601  0.64363604  0.16846621  0.319306782
JavelinThrow    0.14661671  0.417353668 -0.05603983 -0.50676575  0.455509642
run1500m        0.35367539  0.154751578  0.23845838  0.58773633  0.095939716
                     PC6         PC7          PC8         PC9         PC10
run100m        -0.14529344  0.03877925 -0.5225405417 -0.18912876  0.56377485
LongJump       -0.31378631  0.26452821  0.2397936093 -0.37680821  0.47141192
ShotPut         0.30163827  0.48679399 -0.3733069335 -0.24883182 -0.27874214
HighJump       -0.56578960  0.20003055 -0.0863265488  0.27759416 -0.25232651
run400m         0.08353502  0.39950881  0.6958341743 -0.12193818  0.05903967
run110mHurdles -0.16159371  0.32941591 -0.0413396327  0.33495843 -0.20598142
DiscusThrow    -0.09611670 -0.05837631  0.0523048302  0.51499604  0.41812429
PoleVault       0.61950129  0.05317858  0.0001373733  0.15121574  0.20976347
JavelinThrow    0.03930975 -0.54520002  0.1833795283  0.02426563 -0.06223538
run1500m       -0.20412816 -0.28273564  0.0195458120 -0.51723117 -0.22348819
> summary(out1)
```

```
Importance of components:
                          PC1    PC2    PC3    PC4     PC5     PC6     PC7
Standard deviation     1.7666 1.4612 1.2325 1.0470 0.86817 0.80717 0.61148
Proportion of Variance 0.3121 0.2135 0.1519 0.1096 0.07537 0.06515 0.03739
Cumulative Proportion  0.3121 0.5256 0.6775 0.7871 0.86248 0.92763 0.96502
                           PC8     PC9    PC10
Standard deviation     0.41676 0.30781 0.28523
Proportion of Variance 0.01737 0.00947 0.00814
Cumulative Proportion  0.98239 0.99186 1.00000
>
> # How many components should be retained?
> # Kaiser criterion -> 4 (78.71% of the total variance explained)
>
> ## loadings
> #1st alternative
> W=out1$rotation[,1:4]
> sqrt.lamb=out1$sdev[1:4]
> cbind(W[,1]*sqrt.lamb[1],W[,2]*sqrt.lamb[2],W[,3]*sqrt.lamb[3],W[,4]*sqrt.lamb[4])
                     [,1]        [,2]        [,3]        [,4]
run100m         0.88340486 -0.17024936  0.12232945 -0.29092427
LongJump       -0.83501091  0.21169601 -0.03361718  0.05187433
ShotPut         0.01861247  0.87590363 -0.20376173 -0.07844567
HighJump       -0.12363220  0.31684336  0.77244774 -0.21872154
run400m         0.87477441  0.05989936  0.09765179 -0.20440849
run110mHurdles  0.57626071 -0.36266387 -0.30111820  0.18734648
DiscusThrow     0.25897513  0.79704336 -0.21001644  0.43209617
PoleVault      -0.07575561 -0.01301389  0.79327685  0.17638383
JavelinThrow    0.25901228  0.60981669 -0.06906869 -0.53058287
run1500m        0.62480100  0.22611541  0.29389826  0.61535892
> # 2nd alternative
> cor(x,out1$x[,1:4])
                       PC1         PC2         PC3         PC4
run100m         0.88340486 -0.17024936  0.12232945 -0.29092427
LongJump       -0.83501091  0.21169601 -0.03361718  0.05187433
ShotPut         0.01861247  0.87590363 -0.20376173 -0.07844567
HighJump       -0.12363220  0.31684336  0.77244774 -0.21872154
run400m         0.87477441  0.05989936  0.09765179 -0.20440849
run110mHurdles  0.57626071 -0.36266387 -0.30111820  0.18734648
DiscusThrow     0.25897513  0.79704336 -0.21001644  0.43209617
PoleVault      -0.07575561 -0.01301389  0.79327685  0.17638383
JavelinThrow    0.25901228  0.60981669 -0.06906869 -0.53058287
run1500m        0.62480100  0.22611541  0.29389826  0.61535892
>
> ## Plotting athletes using PC's
> PC1=out1$x[,1]; PC2=out1$x[,2]; PC3=out1$x[,3]; PC4=out1$x[,4];
> cbind(PC1[1],PC2[1],PC3[1],PC481])
Error: unexpected ']' in "cbind(PC1[1],PC2[1],PC3[1],PC481]"
>
> cbind(PC1[1],PC2[1],PC3[1],PC4[1])
          [,1]     [,2]     [,3]       [,4]
[1,] -1.623168 2.505216 1.122369 -0.1823041
>
```

Exercise 2

```
> #### gamma exercise
> # data generated using a gamma with parameters alpha=5 and theta=4
>
> direct="C:/Users/joaoas/Desktop/trial/"
> file="gamma.csv"
>
> dta=read.csv(paste(direct,file,sep=""),header=T) # read data set
> # Check reading
> head(dta)
         x
1 23.06331
2 41.00129
3 18.14969
4 10.54072
5 22.11920
6 22.01414
> attach(dta)
>
> # a)
> minusloglikgamma=function(param,x){
```

```
+   alpha=param[1]; theta=param[2]
+   return(-sum(dgamma(x,shape=alpha,scale=theta,log=T)))
+   }
>
> out=nlm(minusloglikgamma,c(1,10),hessian=T,x=x); out
Warning messages:
1: In dgamma(x, shape = alpha, scale = theta, log = T) : NaNs produced
2: In nlm(minusloglikgamma, c(1, 10), hessian = T, x = x) :
  NA/Inf replaced by maximum positive value
3: In dgamma(x, shape = alpha, scale = theta, log = T) : NaNs produced
4: In nlm(minusloglikgamma, c(1, 10), hessian = T, x = x) :
  NA/Inf replaced by maximum positive value
5: In dgamma(x, shape = alpha, scale = theta, log = T) : NaNs produced
6: In nlm(minusloglikgamma, c(1, 10), hessian = T, x = x) :
  NA/Inf replaced by maximum positive value
7: In dgamma(x, shape = alpha, scale = theta, log = T) : NaNs produced
8: In nlm(minusloglikgamma, c(1, 10), hessian = T, x = x) :
  NA/Inf replaced by maximum positive value
$`minimum`
[1] 697.1006

$estimate
[1] 5.529150 3.577327

$gradient
[1] -4.317885e-06 -5.720370e-06

$hessian
         [,1]     [,2]
[1,] 39.63455 55.90487
[2,] 55.90487 86.37696

$code
[1] 1

$iterations
[1] 60

>
> vcov=solve(out$hessian); vcov
           [,1]       [,2]
[1,]  0.2897028 -0.1875014
[2,] -0.1875014  0.1329318
>
> print("95% CI for alpha");
[1] "95% CI for alpha"
>
cbind(out$estimate[1]+qnorm(0.025)*sqrt(vcov[1,1]),out$estimate[1]+qnorm(0.975)*sqrt(vco
v[1,1]))
         [,1]     [,2]
[1,] 4.474218 6.584082
> print("95% CI for theta");
[1] "95% CI for theta"
>
cbind(out$estimate[2]+qnorm(0.025)*sqrt(vcov[2,2]),out$estimate[2]+qnorm(0.975)*sqrt(vco
v[2,2]))
         [,1]     [,2]
[1,] 2.862728 4.291926
>
>
> # b)
> print("ML estimate of P(X>25)");
[1] "ML estimate of P(X>25)"
> pgamma(25,shape=out$estimate[1],scale=out$estimate[2],lower=F)
[1] 0.2380413
>
```

Exercise 3

Nothing to add