MASTERS IN ACTUARIAL SCIENCE

Risk Models

18/01/2021

2nd part of the exam

Time allowed: 1 hour

Instructions:

1. This paper contains **3** questions and comprises **2** 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 questions.
6. Begin your answer to each of the questions on a new page.
7. Marks are shown in brackets. Total marks: 60.
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. The Human Development Index (HDI) was created by the United Nations to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone. The index is computed from a set of indicators covering Income, Health, Education among other topics.

   A subset of indicators from 2018 dataset has been collected and is presented in file HDI.csv. The meaning of the variables is

   - LEF – Life expectancy at birth for females
   - LEM – Life expectancy at birth for males
   - GDPpc – GDP per capita ($1000 per year) corrected by ppp
   - DTP – Lack of vaccination (%) of **DTP vaccine** which is a class of combination vaccines against three diseases in humans: diphteria, pertusis (whooping cough), and tetanus.
   - Measles – vaccine that prevents measles. % of lack of vaccination
   - MalNut - % of children under 5 years old suffering from malnutrition.
   - InfMort – Infant mortality rate (per 1000 live births)
   - Und5Mort – Under 5 years old mortality rate (per 1000 live births)
   - FMR – Female mortality rate (per 1000 females)
   - MMR – Male mortality rate (per 1000 males)
   - Malaria – Incidence of malaria per 1000 people at risk
   - Tuberculosis - Incidence of tuberculosis per 100000 people at risk
   - HIV – % of infected people (age 15-49)
   - HE – Health expenditures (% of the GDP)

      a. **[5]** Assuming that life expectancy (for both females and males) is normally distributed, test $H_0: \rho = 0$ against $H_1: \rho > 0$. Explain the meaning of the test.

      b. **[13]** Use a PCA analysis to summarize our set of variables.
         i. Should we scale the variables? Justify.
         ii. How many principal components should be retained (use Kaiser criterion)?
         iii. Compute the loadings for the retained components. Can you interpret them?

2. File PropertyDammage.csv presents a set of claims related to property damage insurance. Each value is a proportion of the insured capital. Assume that this proportion can be well modelled using a beta distribution.

      a. [**10]** Obtain a maximum likelihood estimate for both parameters.

      b. [**12]** Obtain a 95% confidence interval for the first parameter and obtain a 95% confidence interval for the expected value of the proportion, using the delta method.

3. [**20**] As it is well known, in a normal population the mean is equal to the median. To estimate the population's mean, a junior actuary is not sure if it is more adequate to use the sample average or the sample median, when the sample size is small. So, she decides to use simulation to shed some light on the problem. She will consider samples with size 11 from a normal population with mean 100 and standard deviation 10 and she will compute the mean squared error for both estimators. Develop the simulation procedure and conclude.

SOLUTION

```
> dta=read.csv("HDI.csv",header=T,sep=";")
> head(dta)
  Rank     Country  LEF  LEM  GDPpc DPT Measles MalNut InfMort Und5Mort FMR MMR
1    1      Norway 84.3 80.3 67.986   1       4      0     2.1      2.6  43  69
2    2 Switzerland 85.5 81.7 59.462   3       4      0     3.7      4.2  36  58
3    3     Ireland 83.7 80.4 55.752   2       8      0     3.0      3.5  47  81
4    4     Germany 83.6 78.8 47.060   2       3      0     3.1      3.7  50  92
5    6   Australia 85.3 81.3 44.130   2       5      2     3.0      3.5  45  76
6    6     Iceland 84.4 81.3 47.535   3       7      0     1.6      2.1  38  66
  Malaria Tuberculosis HIV   HE
1       0          5.1 0.1 10.5
2       0          7.2 0.0 12.2
3       0          7.3 0.2  7.4
4       0          7.5 0.2 11.1
5       0          6.8 0.1  9.3
6       0          4.5 0.0  8.3
> # a)
> attach(dta)
> cor.test(LEF,LEM,0,alternative="greater",method="pearson")

        Pearson's product-moment correlation

data:  LEF and LEM
t = 46.309, df = 155, p-value < 2.2e-16
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.9555312 1.0000000
sample estimates:
      cor
0.9657089

> # As the p-value is close to 0 we strongly reject the null and conclude that the correlation
> #coefficient between life expectancy for males and females in the same country is positive.
>
> # we are testing if, at the country level, there is a positive correlation
> # between life expectancy for males and females.
>
> # b)
> #Yes, as variables are measured in different units
>
> X=dta[,-c(1,2)]
> head(X)
   LEF  LEM  GDPpc DPT Measles MalNut InfMort Und5Mort FMR MMR Malaria
1 84.3 80.3 67.986   1       4      0     2.1      2.6  43  69       0
2 85.5 81.7 59.462   3       4      0     3.7      4.2  36  58       0
3 83.7 80.4 55.752   2       8      0     3.0      3.5  47  81       0
4 83.6 78.8 47.060   2       3      0     3.1      3.7  50  92       0
5 85.3 81.3 44.130   2       5      2     3.0      3.5  45  76       0
6 84.4 81.3 47.535   3       7      0     1.6      2.1  38  66       0
  Tuberculosis HIV   HE
1          5.1 0.1 10.5
2          7.2 0.0 12.2
3          7.3 0.2  7.4
4          7.5 0.2 11.1
5          6.8 0.1  9.3
6          4.5 0.0  8.3
> ncol(X)
[1] 14
> out=prcomp(X,scale=T); out
Standard deviations (1, .., p=14):
 [1] 2.96015526 1.21434716 0.98553593 0.93395162 0.78772736 0.62459682
 [7] 0.53195109 0.50072242 0.40384567 0.35442751 0.20570991 0.18725487
[13] 0.07072998 0.06221540

Rotation (n x k) = (14 x 14):
                PC1         PC2          PC3          PC4         PC5
LEF      -0.3276645  0.02237646 -0.063907407  0.140682940 -0.04930763
LEM      -0.3264599  0.09033071  0.004861431  0.094453183  0.04396700
GDPpc    -0.2365384 -0.01241465  0.060427348  0.259156944  0.79116931
DPT       0.2133620  0.46741895 -0.075815704  0.474317804 -0.02117912
Measles   0.2318912  0.41839751  0.040628900  0.461379356 -0.01780662
MalNut    0.2851724  0.07246708 -0.102647624 -0.098115848 -0.14962175
InfMort   0.3214525  0.05760733  0.122211221 -0.099424708  0.06140810
Und5Mort  0.3199641  0.08636461  0.170588568 -0.095493738  0.11861799
```

```
FMR           0.3195220 -0.17422142  0.108702432 -0.003077244  0.08629559
MMR           0.2987418 -0.25329095 -0.036991989  0.013722599 -0.07038202
Malaria       0.2341652  0.08824575  0.465259120 -0.259415153  0.40959111
Tuberculosis  0.2433059 -0.22315421 -0.350938789  0.297092101  0.04319328
HIV           0.1565976 -0.60424823 -0.141539904  0.342703656  0.15938675
HE           -0.1178380 -0.25378971  0.746610465  0.405514162 -0.34822128
                     PC6          PC7          PC8         PC9        PC10
LEF           0.06339215 -0.171066602  0.071134665 -0.09736861  0.02213419
LEM           0.20002052 -0.053278963  0.190367724  0.17137664 -0.01995040
GDPpc         0.11380003  0.376135418 -0.114676051 -0.22543341  0.11949845
DPT          -0.13310230 -0.138081865  0.067541287 -0.43842323 -0.49826658
Measles      -0.15912767 -0.002915343  0.091939270  0.42927622  0.56844942
MalNut        0.60584720  0.214421886  0.473639539 -0.38291364  0.29264915
InfMort       0.12341754  0.304426287 -0.036638277  0.26238325 -0.32336711
Und5Mort      0.06137920  0.260667950 -0.003857975  0.23205783 -0.29826894
FMR          -0.14294235  0.176363620 -0.095789193 -0.01187882  0.13728668
MMR          -0.36238300  0.029407195 -0.241257406 -0.45245254  0.28507930
Malaria       0.08890643 -0.670650958  0.107830265 -0.07079855  0.07946915
Tuberculosis  0.53305157 -0.311413773 -0.521595339  0.14619059 -0.02296705
HIV          -0.16708933 -0.121042693  0.585990852  0.16143944 -0.16330313
HE            0.20545971  0.085867562 -0.111301528 -0.08494877 -0.02500190
                      PC11        PC12         PC13         PC14
LEF          -0.7158478968  0.05333925  0.237623000  0.495749072
LEM          -0.2549541681  0.54362731 -0.249137693 -0.585601935
GDPpc         0.0392359482 -0.11105624 -0.017718437  0.016337834
DPT           0.0854145729  0.12178547 -0.028630130 -0.006783384
Measles      -0.0510687584 -0.11010487 -0.012226696  0.015967451
MalNut        0.0008342219 -0.01764176  0.026196222  0.015772903
InfMort      -0.3376751023 -0.08659295 -0.632721146  0.245434556
Und5Mort     -0.2904659301 -0.07265823  0.658146176 -0.317053703
FMR           0.1062343116  0.79347709  0.127894369  0.333662189
MMR          -0.4416498821 -0.07979323 -0.160085964 -0.369679228
Malaria      -0.0451232521 -0.01093888 -0.054170785  0.011391038
Tuberculosis  0.0034123838  0.01614504  0.031263618 -0.037201124
HIV           0.0110708728 -0.09248433 -0.010160269 -0.003548061
HE            0.0648902375 -0.06588387 -0.002077516 -0.014349922
> summary(out)
Importance of components:
                          PC1    PC2     PC3    PC4     PC5     PC6     PC7
Standard deviation     2.9602 1.2143 0.98554 0.9340 0.78773 0.62460 0.53195
Proportion of Variance 0.6259 0.1053 0.06938 0.0623 0.04432 0.02787 0.02021
Cumulative Proportion  0.6259 0.7312 0.80060 0.8629 0.90723 0.93510 0.95531
                           PC8     PC9    PC10    PC11   PC12    PC13    PC14
Standard deviation     0.50072 0.40385 0.35443 0.20571 0.1873 0.07073 0.06222
Proportion of Variance 0.01791 0.01165 0.00897 0.00302 0.0025 0.00036 0.00028
Cumulative Proportion  0.97322 0.98487 0.99384 0.99686 0.9994 0.99972 1.00000
> #using Kaiser criterion we retain 2 PC
> round(cor(X,out$x[,c(1,2)]),3)
                PC1    PC2
LEF          -0.970  0.027
LEM          -0.966  0.110
GDPpc        -0.700 -0.015
DPT           0.632  0.568
Measles       0.686  0.508
MalNut        0.844  0.088
InfMort       0.952  0.070
Und5Mort      0.947  0.105
FMR           0.946 -0.212
MMR           0.884 -0.308
Malaria       0.693  0.107
Tuberculosis  0.720 -0.271
HIV           0.464 -0.734
HE           -0.349 -0.308
>
> # PC1 is linked to almost all variable with 2 exceptions HIV and HE
> # PC2 is linked to HIV and also to DPT and Measles
>
>
>
> ### exer 2
> dta=read.table("PropertyDammage.csv",header=F)
> head(dta)
         V1
1 0.9926329
2 0.1024562
3 0.5725513
4 0.2428617
```

```
5 0.7247885
6 0.4310237
>
> # a)
> minusloglik=function(param){
+    alpha=param[1]; beta=param[2]
+    return(-sum(dbeta(dta$V1,alpha,beta,log=T)))
+  }
>
> out=nlm(minusloglik,c(1,1),hessian=T); out
Warning messages:
1: In dbeta(dta$V1, alpha, beta, log = T) : NaNs produced
2: In nlm(minusloglik, c(1, 1), hessian = T) :
  NA/Inf replaced by maximum positive value
3: In dbeta(dta$V1, alpha, beta, log = T) : NaNs produced
4: In nlm(minusloglik, c(1, 1), hessian = T) :
  NA/Inf replaced by maximum positive value
5: In dbeta(dta$V1, alpha, beta, log = T) : NaNs produced
6: In nlm(minusloglik, c(1, 1), hessian = T) :
  NA/Inf replaced by maximum positive value
$minimum
[1] -35.08118

$estimate
[1] 1.559240 0.795248

$gradient
[1]  3.202828e-06 -1.069942e-05

$hessian
           [,1]      [,2]
[1,]   72.14313 -105.4473
[2,] -105.44730  358.6013

$code
[1] 1

$iterations
[1] 11

>
> # b)
> a.h=out$estimate[1]; b.h=out$estimate[2]
> V=solve(out$hessian)
> cbind(a.h-1.96*sqrt(V[1,1]),a.h+1.96*sqrt(V[1,1]))
         [,1]     [,2]
[1,] 1.253647 1.864834
>
> (mu.h=a.h/(a.h+b.h))
[1] 0.6622417
>
> der1=b.h/(a.h+b.h)^2
> der2=-a.h/(a.h+b.h)^2
> der=matrix(c(der1,der2),nrow=2,ncol=1)
> var.g=t(der)%*%V%*%der; var.g
             [,1]
[1,] 0.0003103145
>
> cbind(mu.h-1.96*sqrt(var.g),mu.h+1.96*sqrt(var.g))
          [,1]      [,2]
[1,] 0.6277148 0.6967686
>
>
> ### exer 3
>
> m=100; sdev=10; n=11   # Problem setup
>
> NR=100000;
> sample.mean=rep(NA,NR); sample.median=rep(NA,NR)
> for(i in 1:NR){
+   x=rnorm(n,m,sdev)
+   sample.mean[i]=mean(x)
+   sample.median[i]=median(x)
+ }
>
> (mse.sample.mean=mean((sample.mean-m)^2))
[1] 9.095478
```

```
> (mse.sample.median=mean((sample.median-m)^2))
[1] 13.73475
>
> er.median=abs(sample.median-m)
> er.mean=abs(sample.mean-m)
>
> (prop=mean(er.median<er.mean))
[1] 0.38945
>
```