Extreme value theory in medical sciences: modeling total high cholesterol levels in Portugal [1]

P. de Zea Bermudez [2]
DEIO, Faculty of Sciences, University of Lisbon, Portugal and CEAUL
pcbermudez@fc.ul.pt

Zilda Mendes
CEFAR, INFOSAÚDE
zilda.mendes@anf.pt

Abstract
The World Health Organization (WHO) estimated that over 50% of the mortality and disability caused by the ischemic heart disease and stroke could be avoided by implementing simple measures at individual and national levels. Programs targeted to promote the control of the main risk factors for these pathologies, such as hypertension, hypercholesterolemia, smoking and obesity, should be design and implemented.

In 2005, the Department of Pharmaceutical Care Services of the Portuguese National Association of Pharmacies, known as ANF, developed a survey with the purpose of assessing the cardiovascular risk of the Portuguese population that attended the pharmacies. As part of these campaign, several parameters were measured - the total cholesterol, the blood glucose and the triglycerides levels, the blood pressure and the body mass index. Along with this information, the gender, age and geographical location of the pharmacy that the individual visited was recorded.

To our best knowledge, the studies performed so far in our country describe the mean behavior of the individuals. However, this approach is completely unable to address the subjects which have very high levels of certain parameters, such as the total cholesterol. These individuals are located in the tail of the distribution and are the ones most at risk of developing a cardiovascular disease. A precise knowledge of this subpopulation is crucial for developing new strategies/campaigns that focus on reducing the cardiovascular risk of the Portuguese population. An appropriate way to address this problem is to use the extreme value theory (EVT). The EVT has extensively been applied to model data of many scientific areas, such as hydrology, environment, meteorology and forest fires, but seldom used in Medical Sciences. In this paper, we will use the Peaks over threshold (POT) method to model the sample of excesses above a sufficiently high value of total cholesterol. As usual, in any EVT application, we will estimate the levels of total cholesterol, which are to be observed with a small probability (extreme quantiles), per geographical region. Even though the results obtained cannot be generalized to the whole Portuguese population, because only the users of the pharmacies are analyzed, they indicate the cardiovascular trend of the population and may motivate further analysis.

Keywords: Cardiovascular diseases, total cholesterol levels, generalized Pareto distribution, peaks over threshold, extreme quantile estimation

[1] The research is (partially) supported by FCT/OE and by the Sabbatical Grant SFRH/BSAB/1138/2011, FCT, Portugal.

[2] Corresponding author: DEIO, FCUL. Edifício C6, Piso 4. Cidade Universitária, Campo Grande. 1749-016 Lisboa (Portugal).

1 Introduction

The WHO (2002), estimated that over 50% of the mortality and disability caused by ischemic heart disease and stroke could be avoid by controlling individual risk factors, such as hypertension, hypercholesterolemia, smoking and obesity. Also, according to estimates published by WHO, about 9 million deaths per year and more than 75 million life years lost per year result from hypertension and hypercholesterolemia. The impact of risk attributed to hypercholesterolemia is 4.4 million deaths per year and 40.4 million disability-adjusted life years (DALYs), which represents 2.8% of the world's total. Several studies (see the National Cholesterol Education Program (2011) and the Report of the Expert Panel on Population Strategies for Blood Cholesterol Reduction (1991)) revealed a clear relation between the cholesterol levels and the occurrence of new cases of coronary heart disease.

The lack of information, at a national level, about the prevalence and incidence of hypercholesterolemia, motivated a systematic two-part review published by Costa et al. (2003). The incidence of hypercholesterolemia in Portugal (found in only one paper) is estimated in 559 new cases per 100 000 inhabitants, increasing with age, up to 54 years for men and 64 years for women. In prevalence studies, the most frequently used criteria for hypercholesterolemia is >200mg/dL, being the mean level higher in most studies. In spite of the heterogeneity of the results analyzed in that paper, the mean prevalence above the 200mg/dL criteria was 56.7% and for >190mg/dL was 63.8%.

Two years after, in 2005, the Department of Pharmaceutical Care Services of the Portuguese National Association of Pharmacies, developed a national survey with the purpose of assessing the cardiovascular risk of the Portuguese population that attended the pharmacies. As a part of this campaign, several parameters were measured in the pharmacy - total cholesterol, triglycerides and blood glucose level, blood pressure and body mass index. After the pharmacists' evaluation, the study protocol suggested that patients at cardiovascular risk should be referred to the physician (CheckSaúde 2008).

In the present study, all the parameters collected are described. However, our attention focuses on the analysis of the largest total cholesterol levels observed. To our best knowledge, the studies performed so far in our country describe the mean behavior of the individuals. That approach is completely unable to characterize the subjects which have very high levels of total cholesterol. In fact, a model (such as the normal distribution) can fit very well the central part of the distribution but be completely inadequate to represent the tails. In terms of public health, it is most important to have a better understanding of the extent and characteristics of the individuals who are most at risk of developing a cardiovascular disease. An adequate way to deal with this problem is to use the EVT. We will apply the widely known POT approach in order to model the largest values of total cholesterol. Estimated extreme quantiles will also be presented. These estimates provide information of how high is a total cholesterol value likely to be observed with a very small probability. The data set was analyzed by district and Portugal was classified in terms of tail weight. We will show that the heaviest tails (representing the highest values of total cholesterol) are observed in the northern part of the country and are most likely to be connected, not only to the genetics of the population, but also to their gastronomical habits. It should be mentioned that the conclusions cannot be generalized to the whole population, because only the users of the pharmacies were analyzed. However, it may indicate the cardiovascular trend of the population and motivate further analysis.

The paper is organized as follows. The data is described in Section 2. Some basic concepts of the EVT and the way the data sets are usually analyzed in this framework are presented in Section 3. The data modeling is presented in Section 4. Finally, the conclusions and the limitations of this study are presented in Section 5, as well as the future work.

2 The total cholesterol levels

The data was collected between the 14th and the 28th of November 2005, in all pharmacies associated to the Portuguese National Association of Pharmacies which were willing to participate in this campaign. Those pharmacies were located in all the 18 districts of Continental Portugal, Açores and Madeira. The map of Portugal is presented in Figure 1. The variables recorded, as well as the coding used, are presented in Table 1.

As a result of this campaign, the information of $n = 40\ 065$ individuals was obtained. The number of respondents analyzed in each district is presented in Table 2. The distribution by district obtained is different from the Portuguese inhabitants distribution by district ($p < 0.05$), since this campaign was directly targeted to people that voluntarily attended the pharmacies with the purpose of participating in the campaign. The profiles of the individuals regarding several parameters are presented in Table 3. The descriptives statistics show that 63% of the sampled individuals were female and 41% were, at least, 65 years old, which is consistent with other studies carried out in the Portuguese pharmacies (Martins et al. (2002)). It can also be observed, for the majority of measures presented, that a high percentage of people are above the reference values (see "Interesting remarks" in Table 3).

Figure 1: Map of Portugal (source: google images)

Variable	Coding	Variable	Coding
Gender	Male / Female	Age	Years
District	(see Figure 1)	Smoking habits	Yes / No
Body Mass Index (BMI)	kg/m^2	Fasting Blood Glucose level (FBG)	mg/dL
Triglyceride level	mg/dL	Blood Glucose level at Random Time (BGRT)	mg/dL
Systolic Blood Pressure (SBP)	mm/Hg	Diastolic Blood Pressure (DBP)	mm/Hg
Physician visit	Yes / No	Total cholesterol level	mg/dL

Table 1: Variables recorded and corresponding units (or coding) used

District				
Aveiro	Beja	Braga	Bragança	Castelo Branco
3339 8.4%	586 1.5%	3057 7.7%	1005 2.5%	634 1.6%
District				
Coimbra	Évora	Faro	Guarda	R. A. Madeira
1493 3.8%	749 1.9%	841 2.1%	427 1.1%	1434 3.6%
District				
R. A. Açores	Leiria	Lisboa	Portalegre	Porto
413 1.0%	1718 4.3%	9593 24.3%	366 0.9%	6519 16.5%
District				
Santarém	Setúbal	Viana do Castelo	Vila Real	Viseu
1722 4.4%	3156 8.0%	376 1.0%	1027 2.6%	1094 2.8%

Table 2: Number and corresponding percentage of respondents classified by district

Variable	**Minimum**	**Mean**	**Median**	**Sd**	**Maximum**	**Interesting remarks**	# NA
Age	6.0	59.0	61.0	15.3	105.0	41.4% exceed 64 years of age	1903
BMI	13.1	27.9	27.5	4.6	61.1	73.3% have BMI $\geq$ 25 kg/m^2	8074
FBG	20.0	102.0	95.0	32.3	600.0	29.0% $<$ 70 or $\geq$110 mg/dL	16681
BGR	27.0	113.2	102.0	45.0	508.5	15.2% $\geq$ 140 mg/dL	25356
Trig. level	70.0	194.4	168.0	102.0	598.0	59.5% $\geq$ 150 mg/dL	26016
BP	30.0/64.0	81.0/139.1	80.0/138.0	12.4/21.7	169.0/259.0	86.0% beyond the limits 80 DBP / 120 SBP	3380
Chol. level	55.0	198.9	194.0	33.0	590.0	56.0% $\geq$ 190 mg/dL	8083
Gender	63.2 % Females and 36.8% Males						82
S. habits	10.1 % Smokers and 89.9% Non-smokers						1817

Table 3: Some descriptive statistics

(Trig - Triglyceride; Chol - Cholesterol; S - Smoking; Sd - standard deviation; NA - Missing observations)

The box-plots of the total cholesterol levels classified by gender, smoking habits, age, BMI, FBG, BGRT, triglycrides and Blood Pressure (BP) are presented in Figures 2, 3 and 4. For gender and smoking habits, cholesterol levels are similar in both boxes. However, more extreme cholesterol values are observed for females and for non-smokers. As expected, cholesterol levels seem to increase with age. Cholesterol levels, when classified by BMI and blood glucose, are similar in both group. Nevertheless, higher variability is observed for more obese subjects and for those with blood glucose levels (fasting or at random) within the reference limits. In what concerns to triglycerides levels and blood pressure, both categories under analysis have similar medians, even though higher variability is observed for people that have triglycerides and blood pressure levels, away from the reference values.

Table 5 presents the probability of having high cholesterol levels (over 190 mg/dL) for each category under analysis in each variable. This analysis reinforces the descriptive image provided by the box-plots previously presented, where higher levels of cholesterol were observed in females, older people and individuals with higher values of BMI. The table should be read as follows. Amongst the male subjects, 52.8% have total cholesterol levels $\geq$ 190 mg/dL, while in the female subgroup the percentage is 57.9. In terms of smoking habits, 58.2% of the smokers have large values of total cholesterol, while in the subgroup of nonsmokers, the percentage is 55.9%. The age and the BMI should be interpreted the same way. The 0.95 sample quantile, distributed by district, is presented in Figure 5. It clearly shows that

the 5% inhabitants with the highest values of total cholesterol are located in the regions by the Atlantic Ocean, specially in the northern part of the country and also in the Algarve.

Variable	**Level**	**Classification**
FBG	≥ 70 and < 110	Normal
	< 70 or ≥ 110	Abnormal
BGRT	< 140	Normal
	≥ 140	Reduced tolerance to glucose or Diabetes
Triglyceride level	< 150	Normal
	≥ 150	At risk
BP	Ideal	Diastolic < 80 and Sistolic < 120
	Not ideal	Diastolic ≥ 80 and/or Sistolic ≥ 120
Total cholesterol level	< 190	Normal

Table 4: Cut-off points for several parameters

Variable		**P(Cholesterol $\geq$ 190 mg/dL)**	**Variable**		**P(Cholesterol $\geq$ 190 mg/dL)**
Age	<25	0.314	BMI	<18.5	0.472
	25-34	0.421		[18.5; 24.9]	0.522
	35-44	0.529		[25.0; 29.9]	0.573
	45-54	0.598		$\geq$30	0.585
	55-64	0.600	Smoking Habits	Yes	0.582
	65-74	0.566		No	0.559
	$\geq$75	0.536	Gender	Male	0.528
				Female	0.579

Table 5: Total cholesterol levels classified by age, BMI, smoking habits and gender

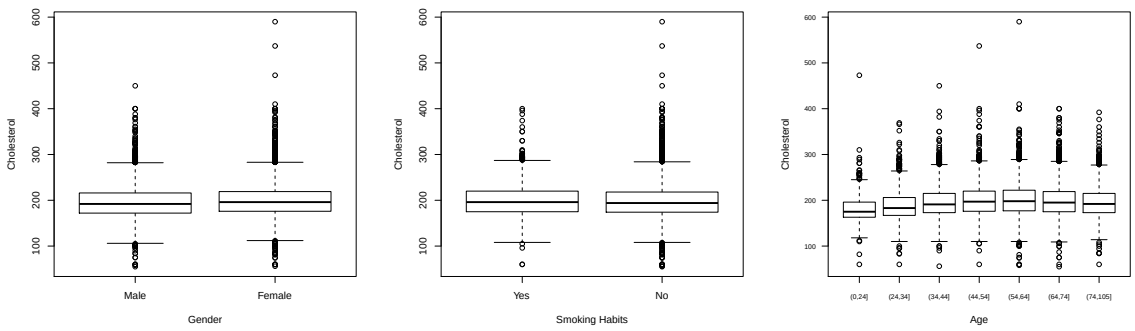

Figure 2: Total cholesterol levels distributed by gender, smoking habits and age

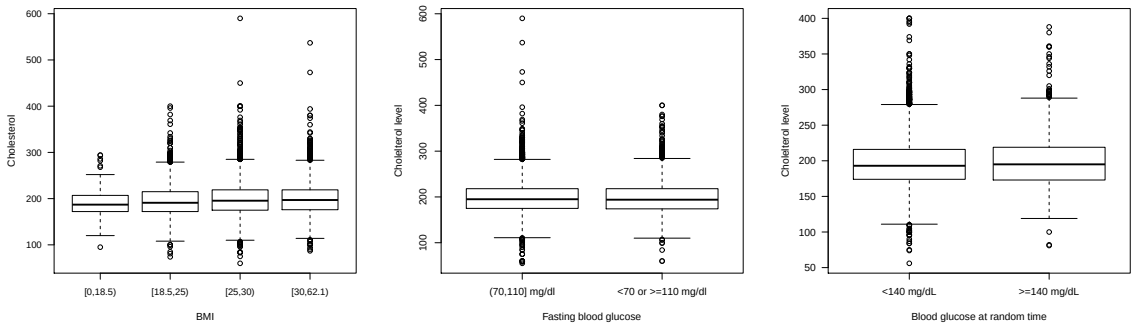

Figure 3: Total cholesterol levels distributed by body mass index and blood glucose

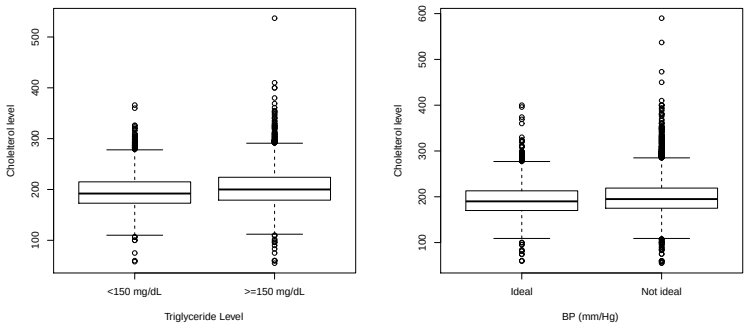

Figure 4: Total cholesterol levels distributed by triglycerides and blood pressure

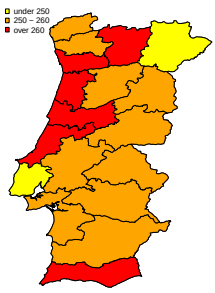

Figure 5: Empirical quantile ($p_{0.95}$) for the total cholesterol level distributed by districts of the mainland of Portugal

3 Some basic concepts of EVT

This study aimed to characterize the population that is most at risk of developing a cardiovascular disease in Portugal. In this framework, interest does not focus on analyzing the mean total cholesterol level, but on the largest values observed. The EVT provides the natural tools to deal with such problems. The spatial distribution of the subjects with very high cholesterol levels is also a very important and interesting issue to address. The possible regional differences are certainly connected to the culture of each region, namely in terms of their gastronomical preferences, but also on genetical and economical factors. Traditionally, the EVT has been applied to various areas such as hydrology, environment, forest fires and finance. However, at least to the best of our knowledge, it has seldom been used in the medical area. The large total cholesterol values are here modeled by means of the widely known POT methodology (see Coles (2001) and Embrechts et al. (1997)). This method consists of fitting the excesses (or the exceedances) above a sufficiently high threshold by the generalized Pareto distribution, GPD (Pickands (1975)). The cumulative distribution function (cdf) of the GDP (k, σ), $-\infty < k < \infty$ e $\sigma > 0$, is given as

$$F(x \mid k, \sigma) = \begin{cases} 1 - \left(1 + \frac{kx}{\sigma}\right)^{-1/k}, & k \neq 0 \\ 1 - \exp(-\frac{x}{\sigma}), & k = 0 \end{cases} \tag{1}$$

where k and σ are the shape and scale parameters, respectively. For $k \geq 0$, $x > 0$, while $0 < x < -\sigma/k$ if $k < 0$. In this parametrization, heavy-tailed distributions are obtained for $k > 0$ (e.g. the Pareto distribution), the exponential distribution for $k = 0$ and light-tailed distributions with finite endpoint (such as the uniform) result when $k < 0$.
One easy and common way to assess the tail heaviness is by plotting the exponential QQ-plot. Let $(x_1, x_2, ..., x_n)$ be an observed sample of size n and $(x_{1:n}, x_{2:n}, ..., x_{n:n})$ the corresponding ascending ordered sample. If the tail of the underlying distribution fits the exponential distribution, then the plot of the pairs $(x_{i:n}, -\log(1 - p_{i:n}))$, $i = 1, 2, ..., n$, should be linear. Otherwise, the data is heavier-tailed (lighter-tailed) than the exponential distribution. The plotting positions generally used are $p_{i:n} = \frac{i}{n+1}$, for $i = 1, 2, ..., n$.

In real data applications, the choice of an adequate threshold above which it is appropriate to consider that the data follows a GPD is frequently a very difficult issue. Several techniques are simultaneously used, such as, the plot of the empirical mean excess function (MEF). The MEF is defined as $e(u) = E(X - u \mid X > u)$ and is estimated by

$$e_n(u) = \frac{\sum_{i=1}^{n} (x_i - u) 1_{\{x_i > u\}}}{\sum_{i=1}^{n} 1_{\{x_i > u\}}}, \tag{2}$$

where $\mathbf{1}$ is the indicator function and u is the threshold. The empirical MEF is plotted as a function of several possible values of u (or as a function of the number of upper order statistics). For the GPD defined in (1) we have

$$e(u) = \frac{\sigma}{1-k} + u\frac{k}{1-k}, k < 1 \text{ and } \sigma + ku > 0. \tag{3}$$

If the data appropriately fits a GPD, then the plot of the empirical MEF, as a function of different values of u, should (approximately) be a straight line above a certain high level u. The slope and intercept of the straight line are $k/(1-k)$ and $\sigma/(1-k)$, respectively. If the underlying distribution is heavy-tailed (light-tailed) then the line should increase (decrease) for all higher thresholds beyond a certain u. If the tail is exponential, then, above some high u, the line should be horizontal.

At least theoretically, once an appropriate threshold has been chosen, and the corresponding excess data fits the GPD for some shape parameter k, then the fit of the GPD remains if the threshold is raised;

only the scale parameter changes. This is know as the stability property of the GPD and was used in this cholesterol data analysis.

It is also common to plot the estimates of the shape parameter k, as a function of various threshold, using for instance the maximum likelihood estimation (see Grimshaw (1983)), the probability weighted moments or the estimates calculated by the moments estimator proposed by Dekkers et al. (1989). A review of methods to estimate the parameters of the GPD is done in the two papers by de Zea Bermudez and Kotz (2010). The moments estimator is defined as

$$\hat{k} = M_n^{(1)} + 1 - \frac{1}{2}\left[1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right]^{-1} \tag{4}$$

where

$$M_n^{(t)} = \frac{1}{r}\sum_{i=1}^{r}\left(\log X_{n-i+1:n} - \log X_{n-r:n}\right)^t, \tag{5}$$

where $t = 1$ and 2, r is the number of upper order statistics and $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ is the ordered sample of size n.
Irrespectively of the estimation method used, the estimates of the shape parameter are graphically represented. Then, the user searches for a stable part in the graph, in the sense that there is some balance between bias and variance. It is widely known that too many upper order statistics increase the bias of the estimator, while two few raise its variance. The threshold (or the corresponding upper order statistics) should be chosen in such a way that there is a "trade-off" between bias and variance. In application, this is a very troublesome task to accomplish. There have been many contributions throughout the years in the statistical literature for assisting the choice of the threshold or, equivalently, the number of upper order statistics. However, a "benchmark", which is simple and very useful, is to consider a threshold equal to the 0.90 sample quantile, as claimed by DuMouchel (1983).

4 Results

4.1 Modeling the total cholesterol levels

The exponential QQ-plots of the observations exceeding the reference total cholesterol level of 190 mg/dL (see Table 4) and the estimated MEFs for the (entire) samples are presented in Figures 6 and 7, respectively. Only the first, alphabetically ordered, nine regions of Portugal are represented.

The plots in Figure 6 clearly show that the exponential distribution may be adequate to model the largest observations of some of the regions, such as Aveiro and Faro. It can clearly be seen that the observed data practically coincides with what is expected to be observed in an exponential model. Others reveal departures from the exponential distribution towards a lighter-tailed distribution, like Castelo Branco and Évora. The large values of cholesterol observed in Braga will certainly be represented by a heavier-tailed distribution.

The empirical MEFs are in general very difficult to analyze. However, it is evident that a threshold around 250 mg/dL seems to be appropriate for Aveiro and Faro. The fact that a horizontal tendency can be visualized from that value onwards reflects the exponential behavior of the tail mentioned above. The clear decline of the empirical MEFs of Castelo Branco, Évora and Guarda support the light-tailed distribution referred while analyzing the QQ-plots. However, in these cases, a visual choice of a possible

threshold is virtually impossible.

The ML estimates and the Moments estimates for the shape parameter k of the GPD for the same nine regions are given in Figures 8 and 9, respectively. The stable parts of the plots for Aveiro and Faro, using both estimators, correspond to shape parameters k very close to zero, also supporting the exponential tail. The estimates of k for Castelo Branco, Évora and Guarda are clearly negative and around -0.3/-0.2. Besides the graphical analysis, the choice of a proper threshold for each data set was complemented by means of two goodness of fit test - the Anderson-Darling and the Cramer-von Mises, which were adapted by Choulakian and Stephens (2001) for testing the null hypothesis H_0: the data comes from a population with cdf given in (1) for some k and σ. The results of the application of these tests for choosing u are presented in Table 6 for the 20 districts. All the GPD models presented are not rejected at the significance level of 5% (see Choulakian and Stephens (2001) for critical values).

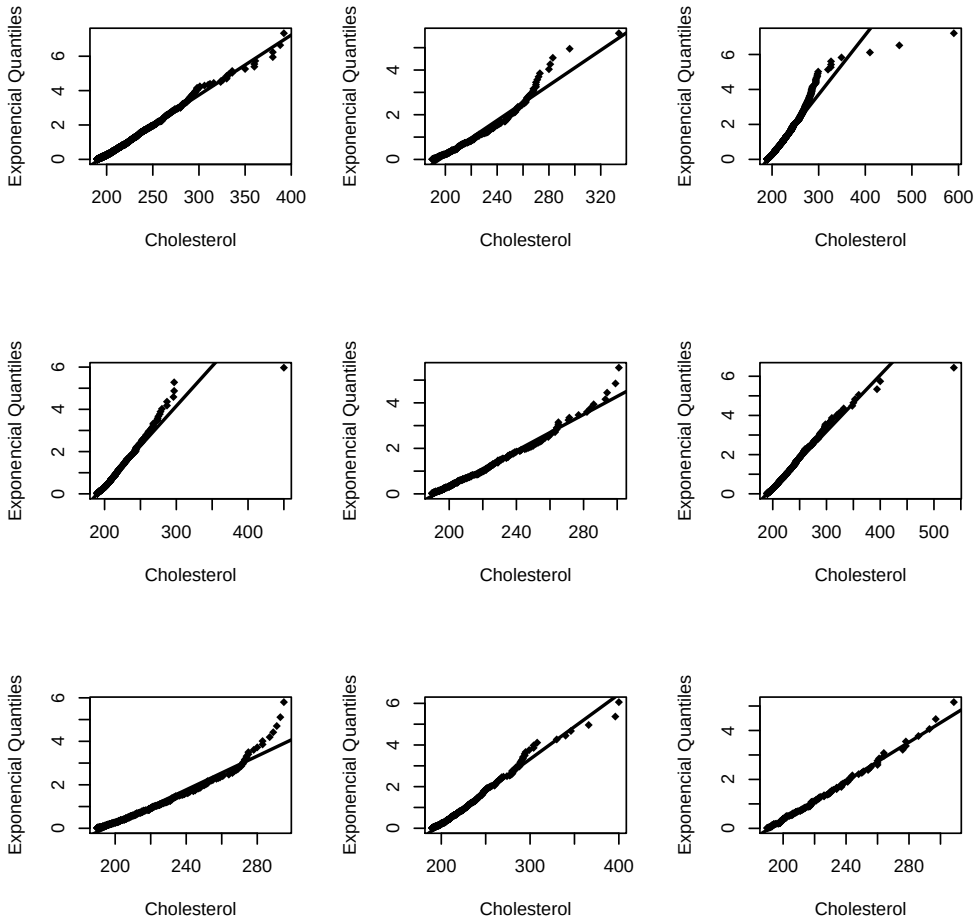

Figure 6: Exponential QQ-plots for the data exceeding the reference value of 190 mg/dL - Aveiro, Beja, Braga, (top row, left to right) Bragança, Castelo Branco, Coimbra (central row, left to right) Évora, Faro e Guarda (bottom row, left to right)

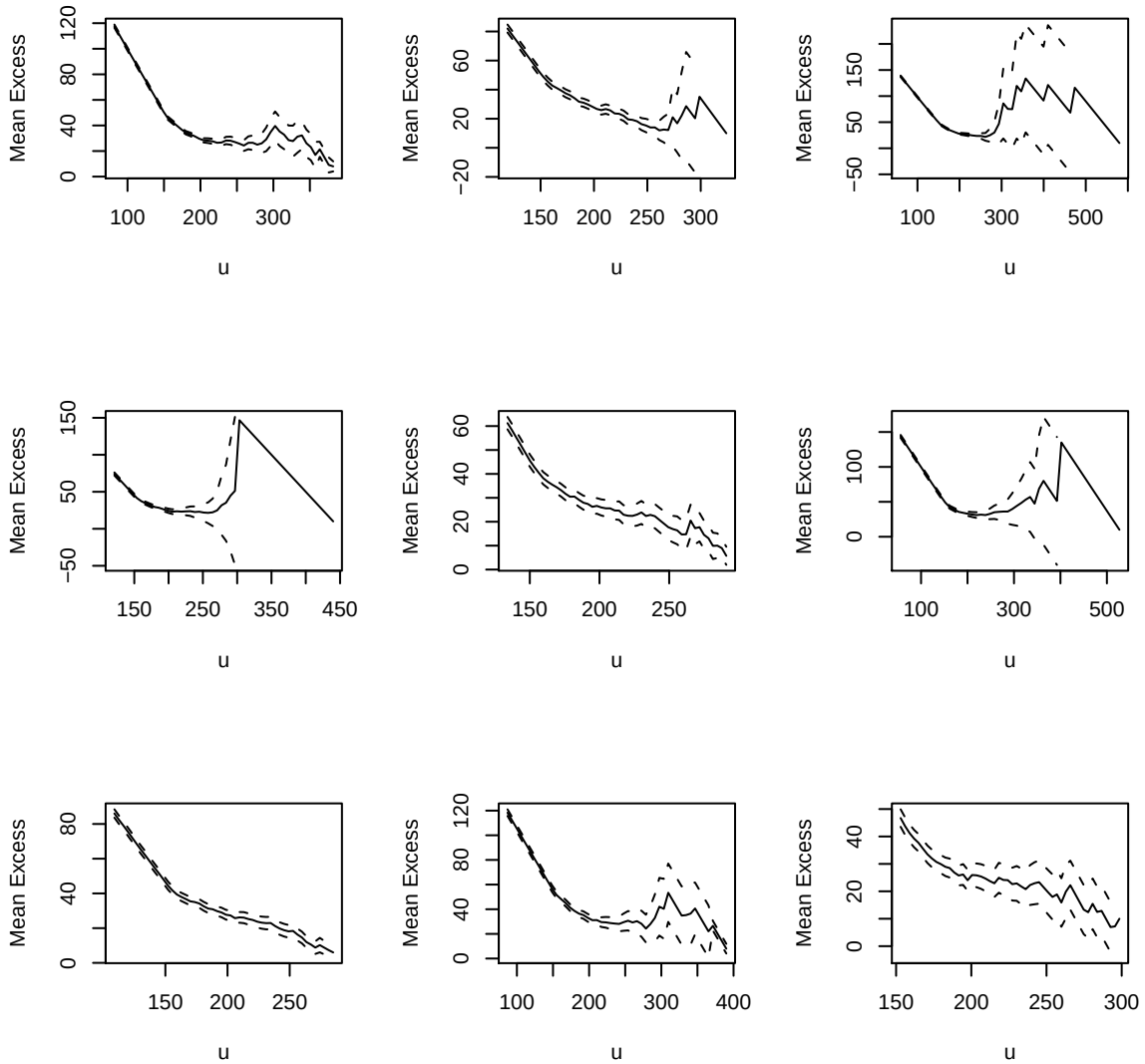

Figure 7: Estimated mean excess functions for the data of 9 districts - Aveiro, Beja, Braga, (top row, left to right) Bragança, Castelo Branco, Coimbra (central row, left to right) Évora, Faro e Guarda (bottom row, left to right)

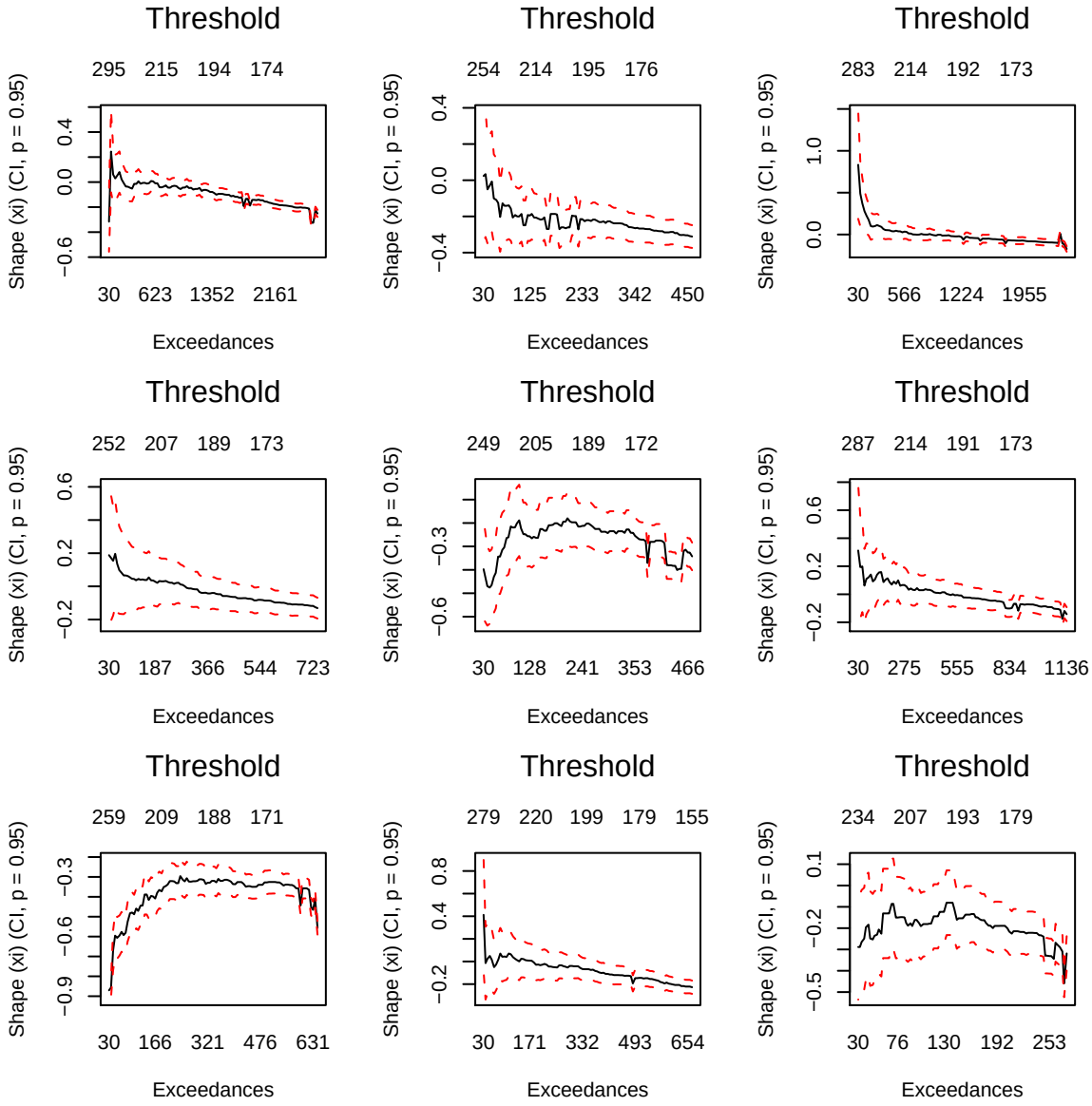

Figure 8: ML estimates for the data of 9 districts - Aveiro, Beja, Braga, (top row, left to right) Bragança, Castelo Branco, Coimbra (central row, left to right) Évora, Faro e Guarda (bottom row, left to right)

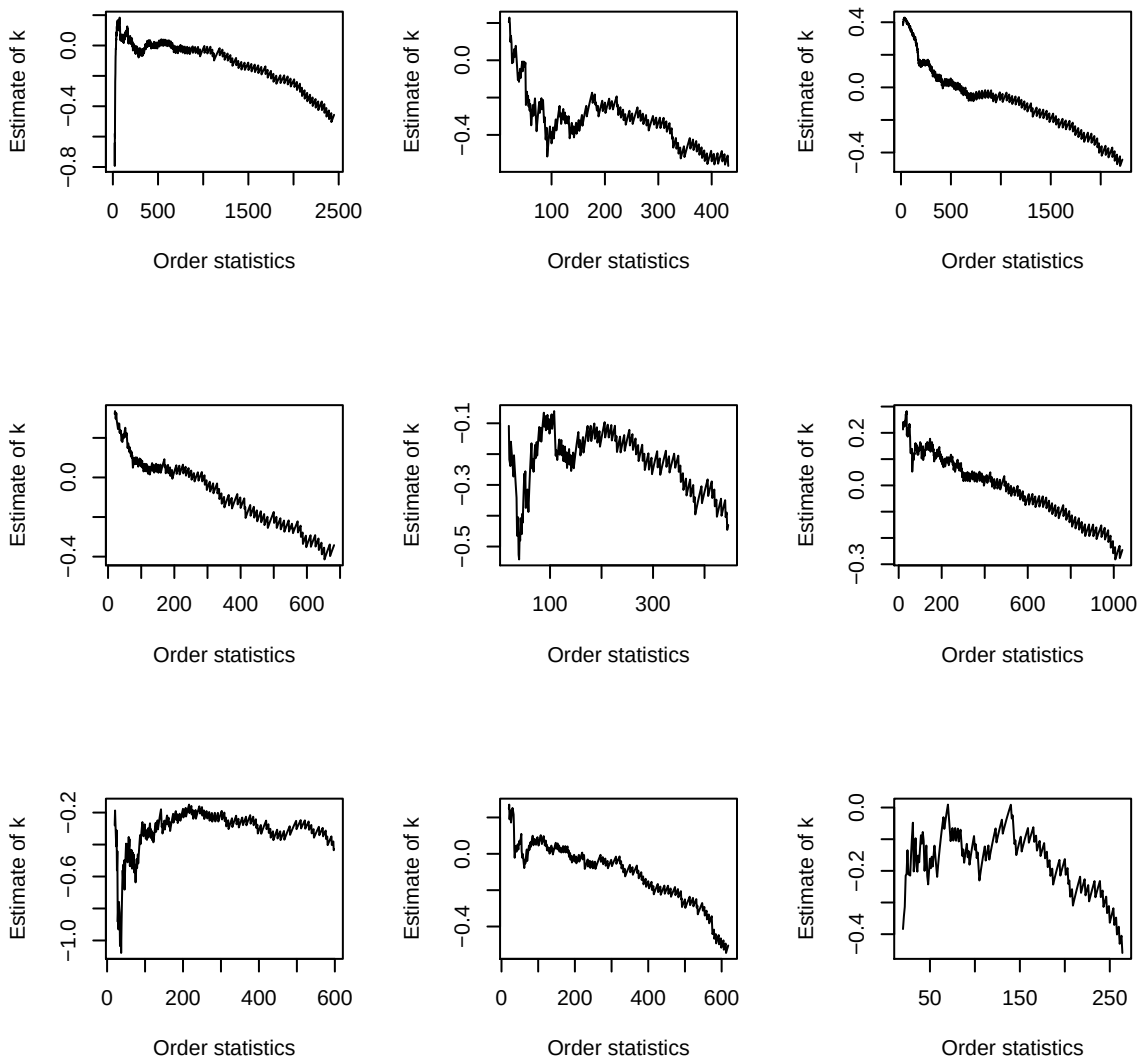

Figure 9: Moments estimates for the data of 9 districts - Aveiro, Beja, Braga, (top row, left to right) Bragança, Castelo Branco, Coimbra (central row, left to right) Évora, Faro e Guarda (bottom row, left to right)

District	u	$\hat{k}$	$\hat{\sigma}$	# Exceed.	Anderson Darling test	Cramer von Mises test
Aveiro	250	0.01	26.3	213	0.854	0.105
Beja	250	-0.04	16.7	37	0.916	0.139
Braga	270	0.42	13.2	71	0.966	0.147
Bragança	240	0.16	18.0	57	0.450	0.049
Castelo Branco	240	-0.52	35.0	39	0.567	0.093
Coimbra	270	0.18	28.1	53	0.447	0.061
Évora	250	-0.65	30.4	44	0.470	0.063
Faro	260	0.05	27.9	51	0.643	0.123
Guarda	230	-0.25	29.1	38	0.242	0.039
Leiria	250	-0.34	24.9	101	0.709	0.071
R. A. dos Açores	230	-0.50	29.4	33	0.741	0.134
R. A. da Madeira	250	-0.28	26.2	76	1.034	0.185
Lisboa	260	0.00	22.2	385	2.688	0.377
Portalegre	240	-0.30	27.1	25	0.417	0.065
Porto	270	0.08	13.7	132	0.865	0.108
Santarém	260	-0.35	18.5	55	0.667	0.116
Setúbal	270	-0.03	16.0	94	1.511	0.250
Viana do Castelo	230	-0.38	27.0	54	2.799	0.401
Vila Real	240	0.14	20.9	74	0.540	0.063
Viseu	250	0.07	15.4	41	0.680	0.119

Table 6: Test statistics of the Anderson-Darling and the Cramer-Von Mises goodness-of-fit tests for the total cholesterol levels of the 20 districts of Portugal

4.2 Fitted models

The GPD models fitted to the data of the 20 districts are presented in Table 7. The parameters of the models were fitted by maximum likelihood (ML), using the R package evir. An extensive grid of thresholds was considered. The lowest possible value was 190 mg/dL. In fact, the cholesterol levels below 190 mg/dL are perfectly normal (see Table 4). The models were chosen considering several aspects. For some districts, such as Aveiro, $u = 190$ was considered far too low. The upper 10% indication of DuMouchel (1983) was also taken into account. Moreover, we tried to choose a threshold that gave rise to a sample large enough in order to enable the use of the asymptotic theory. In fact, the confidence intervals (CIs) of the parameters of the models are based on the limit normal distribution.

As mentioned before, the upper bound of the GPD is finite and equal to $-\sigma/k$ when $k < 0$. For obvious reasons it is omitted from Table 7 when $k > 0$. The table shows that most of the regions are modeled either by some light-tailed GPD ($k < 0$) or by an exponential distribution (k close to zero). The only definite exceptions are Braga and Bragança. The fact that several regions are modeled by GPD(k, σ) with negative k, is most reasonable due to the obvious fact that the total cholesterol levels cannot increase indefinitely. Exponential distribution is also possible because it has a tail that dies out pretty fast. What is slightly disturbing is to fit a moderately heavy-tailed distribution, such as one fitted to the Braga data, to the total cholesterol level, which has an infinite upper bound. The table shows that some of the asymptotic confidence interval of the shape parameter of the GPD presented in Table 7 definitely include zero. Almost symmetrical (or slightly asymmetrical) CIs for the shape parameter k, that definitely include zero, are obtained for Aveiro, Beja, Faro, Lisboa, Setúbal and Viseu. Clearly zero is not included in the CIs of Braga, Castelo Branco, Évora, Leiria, R. A. Madeira and Santarém. It is

barely included in Bragança, Guarda, R. A. Açores, Porto and Viana do Castelo. Although the CI for k corresponding to Portalegre includes zero, the great majority of the interval covers the negative side of the real line. The plots of the Moments estimator and of the ML estimator for the Portalegre data set show that it is more reasonable to consider that the underlying distribution is light-tailed (see Figure 10). Finally, Coimbra and Vila Real might also be reduced to an exponential model.

The exponential models fitted to Aveiro, Beja, Coimbra, Faro, Lisboa, Setúbal, Vila Real and Viseu are presented in Table 8. The table shows some very interesting facts. All the 95 % asymptotic CIs are narrower than the corresponding asymptotic CI given in Table 7, which is a very positive aspect. Moreover, the fact that the log likelihoods of the two models are practically equal supports that the exponential models can be considered instead of the two-parameter GPD models for these 8 regions at any reasonable significance level.

Finally, the results of the Lilliefors test also enable us to conclude, at a 5% significance level, that the upper tail of the total cholesterol level is adequately represented by an exponential distribution (1% was used for Setúbal). For Lisbon, due to the very large sample size, the test statistics slightly surpasses the critical value for $\alpha = 0.01$, although all the previous evidence pointed towards an exponential model. As such, the GPD models can be reduced to its simpler exponential form.

The profile loglikelihoods of some districts are presented in Figure 11. These graphs were obtained using the R package ismev. According to Coles (2001), the profile loglikelihood generally produces more precise CIs for the parameters. That is somewhat visible in the plots mentioned above. The CIs for k are obtained as the values of the dotted lines that intercept the x-axis.

The fits of the exponential models for Aveiro and Beja are graphically presented in Figure 12 (top row). The bottom row the figure contains the PP-plots and QQ-plots of the GPD models fitted to Évora and Santarém. The plots show that the models fit very well to the data.
In terms of tail weight, the mainland of Portugal can then be classified as presented in Figure 13. The districts can then be grouped in three:

- Castelo Branco, Évora, Guarda, Leiria, R. A. Açores, R. A. Madeira, Portalegre, Santarém and Viana do Castelo - light-tailed;
- Aveiro, Beja, Coimbra, Faro, Lisboa, Setúbal, Vila Real and Viseu - exponential tail;
- Braga and Bragança - heavy-tailed.

It must be referred that Porto was not classified in terms of tail weight. The estimate of k is very close to zero, even though practically the whole CI for k is on the positive side, reflecting a tendency to a heavy-tailed distribution.

District	u	$\hat{k}$	r	CI (95%) for k	$\hat{\sigma}$	CI (95%) for σ	Max	Upper bound	log(L)
Aveiro	250	0.01	213	(-0.13; 0.14)	26.3	(21.3; 31.3)	392	− − −	-911.0
Beja	250	-0.04	37	(-0.29; 0.20)	16.7	(10.0; 23.5)	334	(*)	-139.6
Braga	270	0.42	71	(0.13; 0.70)	13.2	(8.5; 17.8)	590	− − −	-283.6
Bragança	240	0.16	57	(-0.09; 0.41)	18.0	(11.5; 24.4)	450	− − −	-230.8
Castelo Branco	240	-0.52	39	(-0.86; -0.18)	35.0	(19.9; 50.1)	301	307	-157.3
Coimbra	270	0.18	53	(-0.11; 0.47)	28.1	(17.0; 39.2)	537	− − −	-239.4
Évora	250	-0.65	44	(-0.95; -0.36)	30.4	(19.2; 41.7)	295	297	-165.5
Faro	260	0.05	51	(-0.24; 0.34)	27.9	(16.8; 39.0)	400	− − −	-223.4
Guarda	230	-0.25	38	(-0.59; 0.09)	29.0	(15.7; 42.4)	309	347	-156.5
Leiria	250	-0.34	101	(-0.47; -0.21)	24.9	(19.3; 30.5)	316	324	-391.9
R. A. Açores	230	-0.50	33	(-1.01; 0.00)	29.4	(12.2; 46.6)	281	288	-128.0
R. A. Madeira	250	-0.28	76	(-0.41; -0.14)	26.2	(19.5; 32.8)	330	345	-302.8
Lisboa	260	0.00	385	(-0.08; 0.09)	22.2	(19.2; 25.2)	400	− − −	-1580.3
Portalegre	240	-0.30	25	(-0.79; 0.18)	27.1	(10.6; 43.7)	302	329	-99.9
Porto	270	0.08	132	(-0.07; 0.23)	13.7	(10.6; 16.8)	400	− − −	-487.9
Santarém	260	-0.35	55	(-0.60; -0.10)	18.5	(12.0; 25.1)	304	313	-196.4
Setúbal	270	-0.03	94	(-0.17; 0.12)	16.0	(12.1; 20.0)	380	(*)	-352.5
Viana do Castelo	230	-0.38	54	(-0.80; 0.04)	27.0	(13.9; 40.1)	286	301	-211.4
Vila Real	240	0.14	74	(-0.11; 0.40)	20.9	(13.7; 28.0)	400	− − −	-309.5
Viseu	250	0.07	41	(-0.19; 0.32)	15.4	(9.2; 21.5)	361	− − −	-155.7

Table 7: GPD models fitted to the 20 districts of Portugal
(*) - Values larger than 600 mg/dL

District	**u**	$\hat{\sigma}$	**CI 95% for σ**	**Lilliefors test**	**Sample size**	**Critical values**
Aveiro	250	26.5	(22.9; 30.1)	0.066	213	0.073
Beja	250	16.0	(10.9; 21.2)	0.140	37	0.174
Coimbra	270	34.4	(25.1; 43.7)	0.122	53	0.146
Faro	260	29.4	(21.4; 37.5)	0.108	51	0.148
Lisboa	260	22.3	(20.1; 24.5)	0.073	385	(a)
Setúbal	270	15.6	(12.5; 18.8)	0.113	94	0.129 (b)
Vila Real	240	24.4	(18.8; 29.9)	0.082	74	0.123
Viseu	250	16.5	(11.4; 21.5)	0.102	41	0.166

Table 8: Exponential models fitted to some districts of Portugal, Lilliefors test statistics and corresponding critical values for $\alpha = 0.05$
(a) - critical value is slightly surpassed ($\alpha = 0.01$); (b) - $\alpha = 0.01$.

4.3 Predicting high cholesterol levels

One of the most interesting final product of an EVT data analysis is the prediction of extreme quantiles and tail probabilities. In the first situation the aim focuses on determining a value z such that $P(X > z) = p$, where p is a fixed, very small probability, frequently much smaller than $1/n$, where n is the sample size. The estimation of a tail probability consists on estimating $P(X > z)$, where z is a very large (and fixed) quantity. The level that is surpassed once every m observations is given as

$$z_m = \begin{cases} u + \frac{\sigma}{k}\left[(m\tau_u)^k - 1\right], & k \neq 0 \\ u + \sigma \log(m\tau_u), & k = 0 \end{cases} \tag{6}$$

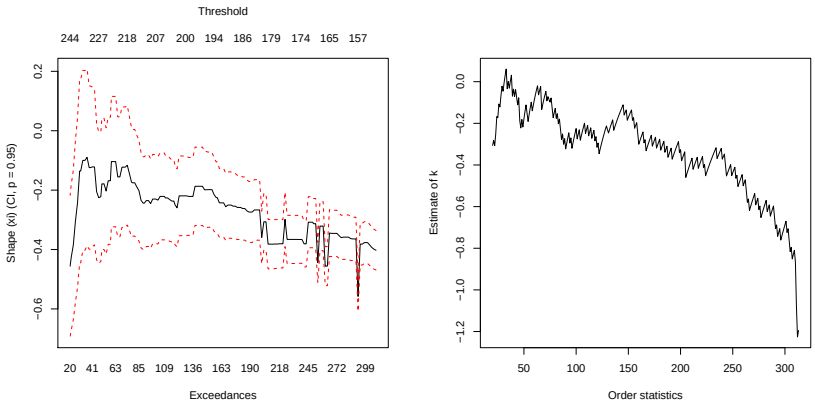

Figure 10: Plots of the estimates of the shape parameter of the GPD for Portalegre - ML estimates (left) and Moments estimates (right)

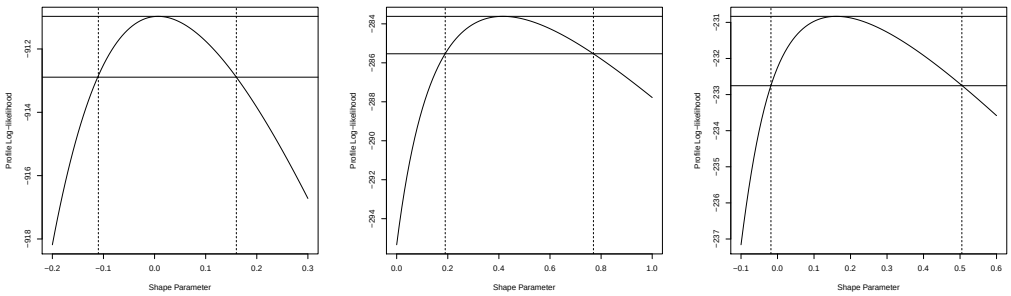

Figure 11: Profile loglikelihood plots for Aveiro $u = 250$ (left), Braga $u = 270$ (center) and Bragança $u = 240$ (right)

where u is the threshold and $\tau_u = P(X > u)$. This is usually estimated by $\frac{r}{n}$, being r the number of upper order statistics. In the above expression, m is equal to $1/p$, and should be large enough so that z_m definitely exceeds u.

Some empirical high quantiles ($p = 0.99$ and $p = 0.995$) and the corresponding estimated GPD quantiles, as well as the 95% CIs, are given in Table 9. The values clearly show the close agreement between the empirical quantiles and the quantiles estimated by the fitted GPD models. We draw the attention of the reader to the estimates of the total cholesterol levels, and corresponding (narrow) CIs, obtained for the districts of Castelo Branco, Leiria, Lisboa and Santarém ($p = 0.99$). These three districts are closely followed by Braga and R. A. Madeira.

The quantiles obtained using the exponential models are presented in Table 10. The CI were computed adapting (to the exponential case) the results that are in Coles (2001) for the general case of the GPD ($k \neq 0$). They are based on the delta method. In general, although the exponential models are better for the regions indicated in Table 8, specially because they are less parsimonious, the GPD models are, in some cases, more accurate in terms of quantile estimation.

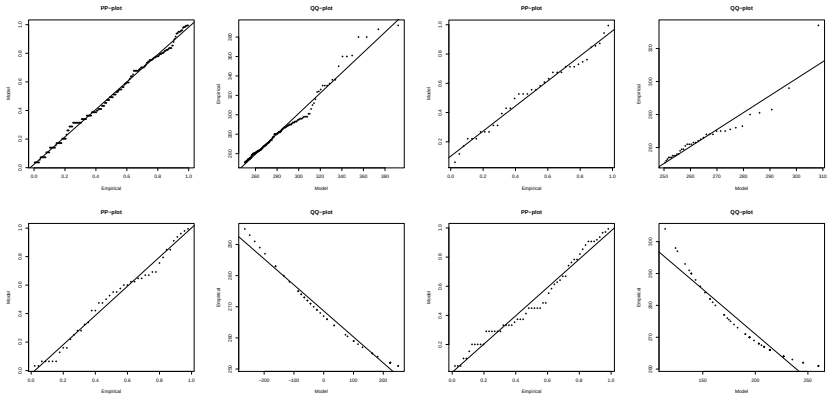

Figure 12: Top row, from left to right - Exponential model fit for Aveiro $u = 250$ (1st and 2nd) and for Beja $u = 250$ (3rd and 4th). Bottom row, from left to right - GPD model fit for Évora $u = 250$ (1st and 2nd) and for Santaréem $u = 260$ (3rd and 4th)

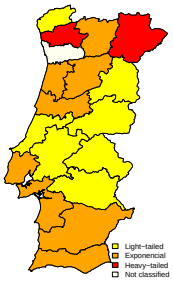

Figure 13: Mainland of Portugal classified according to tail-weight

For explanation purposes, let us consider the quantiles estimated for Castelo Branco. For $p = 0.99$, the empirical quantiles is 284.1 mg/dL and the estimated one is 284.3 mg/dL (the prediction is only 0.2 mg/dL above the observed value). This figure is interpreted as follows: in the district of Castelo Branco 1% of the population that attends the pharmacies has, according to the fitted model, a total cholesterol level above 284.3 mg/dL. Moveover, 0.5% of such population surpasses the estimated value of 291.2 mg/dL. Clearly the analyzed population has a total cholesterol level way over the guideline of 190 mg/dL. Let us now consider the results obtained for Braga (which was modeled by a heavy-tailed distribution). According to the GPD fitted model, 1% of the studied population has total cholesterol level above 287.5 mg/dL and 0.5% is above 304.0 mg/dL.

We do not precisely know the characteristics of the population in terms of the disease background. However, it is clear that all the threshold considered are much higher than the guideline of 190 mg/dL, which reflects a very high risk of developing a cardiovascular event.

A superficial analysis of Figure 14 seems to show that, for this kind of population, there does not

seem to exist a very strong relation between the largest value of total cholesterol and the other variables such as the triglycerides and the BMI (see the Appendix for similar graphs of other variables, such as gender and age). The values of u considered are the ones indicated in Table 7, that is, u =250, 250, 270, 240, 240, 270, 250, 260 and 230 for Aveiro, Beja, Braga, Bragança, Castelo Branco, Coimbra, Évora, Faro and Guarda, respectively. However, it should be mentioned that the graphical representations between the high cholesterol levels and, for instance, the corresponding triglycerides are based on very reduced sample sizes in some districts. The fact that some of the other variables have a considerable number of NAs (see for instance the number of missing values existing for the triglyceride levels, which are indicated in Table 3). Therefore, the statement that there is not a strong relation between certain variable and the high cholesterol levels should be considered with extreme care.

District	u	Empirical $q_{0.99}$	Model $q_{0.99}$	IC 95 % for $q_{0.99}$	Empirical $q_{0.995}$	Model $q_{0.995}$	IC 95 % for $q_{0.995}$
Aveiro	250	296.8	304.5	(297.8 ; 312.7)	327.6	323.1	(313.9 ; 336.0)
Beja	250	274.5	282.7	(274.3 ; 296.2)	282.2	293.1	(282.7 ; 315.3)
Braga	270	285.0	287.5	(283.4 ; 294.6)	294.7	304.0	(297.5 ; 318.8)
Bragança	240	277.4	282.8	(272.3 ; 300.3)	288.9	301.1	(285.4 ; 334.9)
Castelo Branco	240	284.1	284.3	(280.2 ; 292.6)	293.5	291.2	(287.5 ; 302.8)
Coimbra	270	313.6	319.3	(306.3 ; 338.3)	348.5	346.6	(326.6 ; 384.6)
Évora	250	281.1	283.0	(284.2 ; 288.1)	288.4	288.0	(287.5 ; 293.1)
Faro	260	304.6	318.9	(304.7 ; 341.9)	343.4	340.8	(321.7 ; 383.9)
Guarda	230	286.6	285.0	(274.0 ; 306.6)	295.2	294.8	(282.7 ; 331.3)
Leiria	250	286.0	286.0	(284.9 ; 291.3)	290.0	293.9	(292.4 ; 300.0)
R. A. Açores	230	276.0	272.1	(266.8 ; 285.5)	278.1	276.9	(273.5 ; 299.6)
R. A. Madeira	250	286.0	290.6	(287.4 ; 297.9)	295.0	300.0	(295.5 ; 309.1)
Lisboa	260	293.0	295.9	(293.0 ; 298.0)	302.0	311.5	(308.4 ; 317.4)
Portalegre	240	283.6	281.0	(270.1 ; 298.6)	290.7	290.2	(279.4 ; 322.8)
Porto	270	283.0	283.1	(283.5 ; 283.5)	292.0	293.6	(289.8 ; 298.4)
Santarém	260	280.6	280.7	(278.5 ; 285.5)	290.0	287.7	(285.0 ; 293.3)
Setúbal	270	291.0	290.2	(287.0 ; 295.0)	296.0	300.8	(295.9 ; 307.8)
Viana do Castelo	230	278.9	276.8	(270.1 ; 294.1)	282.8	282.4	(276.2 ; 312.3)
Vila Real	240	294.2	295.3	(283.2 ; 316.2)	321.2	316.3	(298.6 ; 356.1)
Viseu	250	270.0	275.2	(268.5 ; 285.5)	280.7	287.3	(277.3 ; 305.6)

Table 9: Some empirical and estimated quantiles using the GPD fitted models to the 20 districts

District	Empirical $q_{0.99}$	Model $q_{0.99}$	IC 95 % for $q_{0.99}$	Empirical $q_{0.995}$	Model $q_{0.995}$	IC 95 % for $q_{0.995}$
Aveiro	296.8	299.1	(291.7 ; 306.5)	327.6	317.5	(307.8 ; 327.1)
Beja	274.5	279.5	(268.8 ; 290.3)	282.2	290.6	(276.6 ; 304.7)
Coimbra	313.6	322.4	(305.6 ; 339.1)	348.5	346.4	(323.9 ; 368.9)
Faro	304.6	313.0	(296.5 ; 329.6)	343.4	333.4	(311.8 ; 355.1)
Lisboa	293.0	291.0	(287.2 ; 294.8)	302.0	306.4	(301.3 ; 311.6)
Porto	283.0	280.5	(277.4 ; 283.6)	292.0	290.8	(286.5 ; 295.2)
Setúbal	291.0	287.1	(282.4 ; 291.7)	296.0	297.9	(291.5 ; 304.4)
Vila Real	294.2	294.7	(281.2 ; 308.2)	321.2	311.7	(294.5 ; 328.9)
Viseu	270.0	271.8	(263.5 ; 280.0)	280.7	283.2	(271.9 ; 294.5)

Table 10: Empirical and estimated quantiles using the exponential models

4.4 Characterization of the "extremal" individuals

The distribution of the highest cholesterol levels according to the district is presented in Figure 15 [3]. The largest values are observed in Coimbra, Bragança and Braga. The high values observed in Bragança and Braga clearly influence the tail weight and that is the reason why the underlying distributions fitted to the data of these two districts is heavier-tailed than the others.

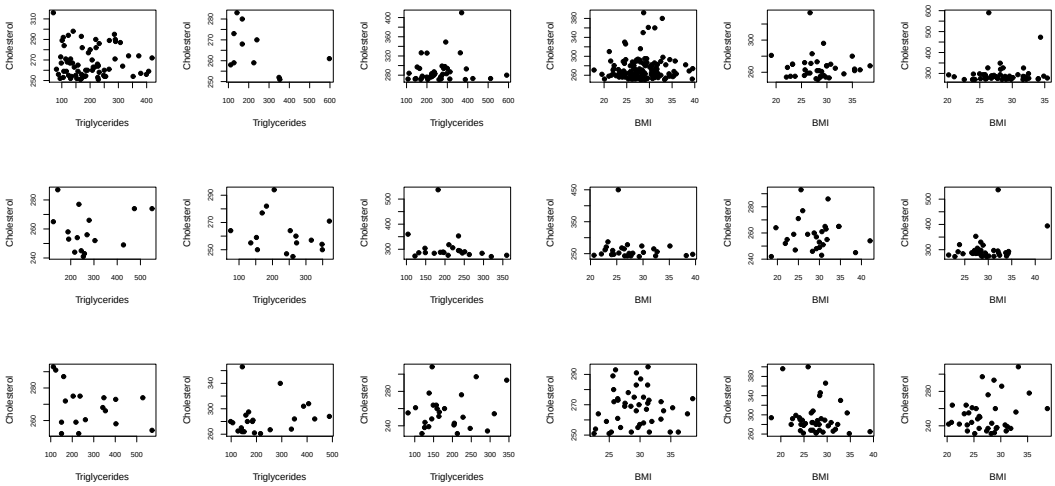

Figure 14: High total cholesterol levels distributed by triglyceride and district / BMI and district - Aveiro, Beja, Braga (top row, left to right), Bragança, Castelo Branco, Coimbra (central row, left to right) and Évora, Faro, Guarda (bottom row, left to right)

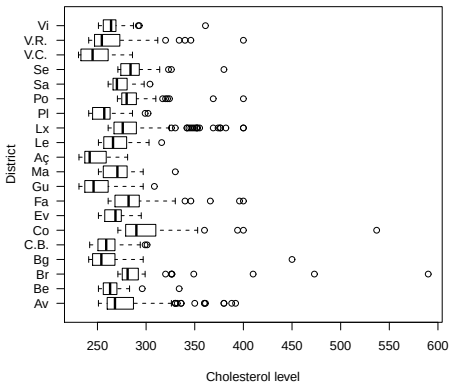

Figure 15: Distribution of high cholesterol levels by district

[3] Abbreviations: Aveiro (Av), Beja (Be), Braga (Br), Bragança (Bg), Castelo Branco (C. B.), Coimbra (Co), Évora (Ev), Faro (Fa), Guarda (Gu), R. A. Madeira (Ma), R. A. Açores (A), Leiria (Le), Lisboa (Lx), Portalegre (Pl), Porto (Po), Santarém (Sa) , Setúbal (Se), Viana do Castelo (V. C.), Vila Real (V. R.) and Viseu (Vi).

5 Conclusions, comments and future work

The study performed shows that

1. the GPD models fit exceeding well to the highest values of cholesterol observed in all the regions. They also provide very precise estimates of the high values of total cholesterol (extreme quantile estimation). We should keep in mind that the extreme quantile estimation is the core aspect of any EVT application.

2. the heaviest tails, in terms of total cholesterol levels, are observed in the north of Portugal - Braga and Bragança. These tails represent the individuals with higher risk of developing a cardiovascular disease. All the other regions have lighter tails (less cardiovascular risk in terms of total cholesterol levels). The grouping of districts performed may be related to the heterogeneity of the population analyzed and/or to the different gastronomic habits that exist throughout the country. In order to support the regional differences found, other indirect indicators to high levels of cholesterol revealed regional asymmetries in the consumption of statins in Defined Daily Doses per 1.000 inhabitants (see Fontes and Mendes (2006)).

3. the analyzed population has obvious problems in controlling the cholesterol level. As mentioned throughout the study, a considerable percentage of the population has a cholesterol level significantly superior than the 190 mg/dL recommended by the guidelines. It should be recalled that all the thresholds presented in Table 7 are way above the cut-point 190 mg/dL.

The authors hope that this study might have contributed to clarify the situation of Portugal in terms of the individuals with the highest cholesterol levels. In spite of the promising results, we must keep in mind that

1. the results presented are based on a large sample of subjects, that were freely willing to participate in the survey. As such, the sample possibly represents the kind of individuals that regularly go to the pharmacy, but it might not represent the general Portuguese population. The issue is that the health condition of the individuals are not really known.

2. despite the above comments about the representativeness of the sample analyzed, several studies indicate high prevalence of hypercholesterolemia in Portugal, which supports that this is a serious health problem in our country. The studies performed in Portugal use different methodologies and the data were collect in different settings and as such are difficult to compare. However, in the majority of this prevalence studies, the mean cholesterol levels are above 200 mg/dL (Simões et al. (2000), Becel (2001) and Estudo ALTO-MAR (2001)).

Finally, the authors would like to stress the fact that

1. initiatives like this one, carried out at a national level in the pharmacies, can contribute to promote a better knowledge of the cardiovascular risk factors of the Portuguese population.

2. the analysis of a possible spatial component might be the next step to pursue in the analysis of this kind of data. At this moment, the authors already have the coordinates of the pharmacies considered in this study and as such a spatial analysis should follow.

Acknowledgements

The first author would like to thank the support of the Department of Statistics of the University Carlos III, Madrid (Spain), where she stayed during her last four months of Sabbatical Leave and where this paper was concluded. Both authors would like to thank the Department of Pharmaceutical Care Services of the Portuguese National Association of Pharmacies for providing the data set that enabled this work.

References

CheckSaúde, Guia Prático, Risco Cardiovascular, 2 Edição (2008). Associação Nacional das Farmácias, Lisboa, Portugal.

Choulakian, V. and Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43, 478-484.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag: London.

Costa, J., Borges, M., Oliveira, E. (2003). Incidência e Prevalência da Hipercolesterolémia em Portugal: Uma Revisão Sistemática da Literatura. Part I. *Rev. Port. Cardiol.*, **22** (4), 569-577.

Costa, J., Borges, M., Oliveira, E. (2003). Incidência e Prevalência da Hipercolesterolémia em Portugal: Uma Revisão Sistemática da Literatura. Part II. *Rev. Port. Cardiol.*, **22**(5), 683-702.

de Zea Bermudez, P. and Kotz, S. (2010,a). Parameter estimation of the generalized Pareto distribution - PartI. *Journal of Statistical Planning and Inference*, **140**, 1353-1373.

de Zea Bermudez, P. and Kotz, S. (2010,b). Parameter estimation of the generalized Pareto distribution - PartII. *Journal of Statistical Planning and Inference*, **140**, 1374-1388.

Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. Ann. Statist. 17, 1833-1855.

DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *Ann. Statist.*, **11**, 1019-1031.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin.

Estudo ALTO-MAR (2001). Caracterização da População Portuguesa seguida em consulta de clínica geral com Alto e Muito Alto Risco Cadiovascular. Fundação Portuguesa de Cardiologia.

Estudo BECEL (2001). Perfil Lipídico da População Portuguesa. Fundação Portuguesa de Cardiologia. Instituto de Alimentação Becel.

Fontes E. and Mendes Z. (2006). Utilization of Statins in Portugal. *Pharmacoepidemiology and Drug Safety*, 15: S182.

Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 35, 185-191.

Martins, A. P. , Miranda, A. C., Mendes, Z. Ferreira, A. P., Soares, M. A. and Nogueira, A. (2002). Self-medication on Portuguese Urban Population: a prevalence study. *Pharmacoepidemiology and Drug Safety*, 11, 409-414.

National Cholesterol Education Program (2011) - http://www.Nhlbi.nih.gov/about/ncep/.

Pickands, J.(1975). Statistical Inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.

Report of the Expert Panel on Population Strategies for Blood Cholesterol Reduction. A statement from the National Education Program. National Heart, Lung and Blood Institute. National Institutes of Health. Circulation 1991; 83:2154-232.

Simões, J. A., Gama, M. E. and Contente, C. B. (2000). Prevalência de factores de risco cardiovascular numa população rural com idade entre 25 e 44 anos. *Revista Portuguesa de Cardiologia*, 19, 656-667.

Third Joint Task Force of European Society of Cardiology and other societies on Cardiovascular Disease Prevention in Clinical Pratice (2003). European guidelines on cardiovascular disease prevention in clinical pratice. *European Heart Journal*, **24**, 1601-1610.

The Task Force for the management of Arterial Hypertension of the European Society of Hypertension and of the European Society of Cardiology (2007). Guidelines for the Management of Arterial Hypertension. Journal of Hypertension, 25, 1105-1187.

The World Health Report (WHO) 2002. Reducing Risk and Promoting Healthy Life (http://www.who.in/whr/2002/en).

Appendix

District	$q_{0.90}$	$q_{0.95}$	$q_{0.975}$	$q_{0.99}$	$q_{0.995}$	$q_{0.999}$	Sample maximum
Aveiro	242	262	281	297	328	380	392
Beja	247	259	268	275	282	316	334
Braga	241	260	272	285	295	382	590
Bragança	231	248	262	277	289	334	450
Castelo Branco	235	254	265	284	294	300	301
Coimbra	244	265	288	314	348	399	537
Évora	239	257	271	281	288	294	295
Faro	250	273	289	305	343	397	400
Guarda	237	254	264	287	295	305	309
R. A. Madeira	243	262	277	286	295	297	330
R. A. Açores	236	249	262	276	278	280	281
Leiria	243	257	274	286	290	301	316
Lisboa	245	261	276	293	302	354	400
Portalegre	238	250	261	284	291	301	302
Porto	240	257	270	283	292	315	400
Santarém	240	258	267	281	290	298	304
Setúbal	248	263	276	291	296	317	380
Viana do Castelo	237	257	272	279	283	285	286
Vila Real	240	253	270	294	321	358	400
Viseu	233	250	263	270	281	303	361

Table 11: Quantiles of the total cholesterol levels classified by district

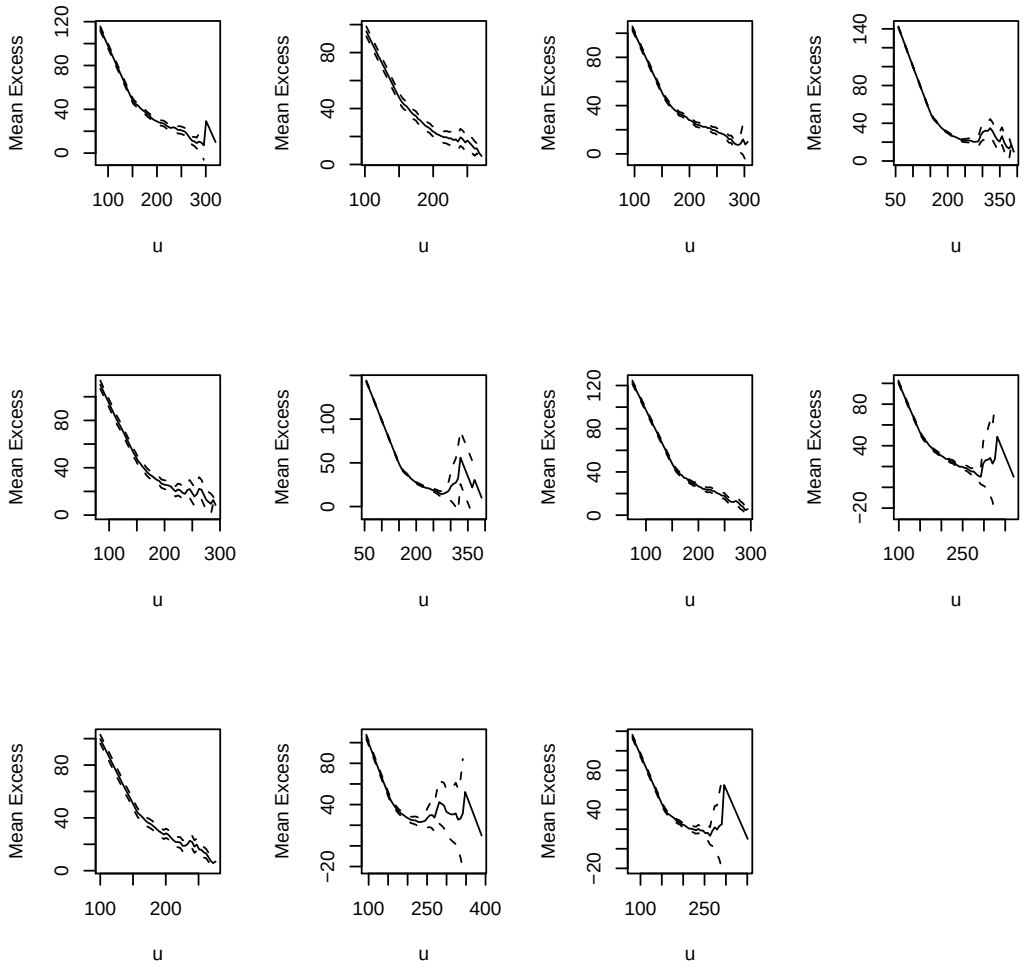

Figure 16: Empirical MEF - R. A. Madeira, R. A. Açores, Leiria, Lisboa (top row, left to right), Portalegre, Porto, Santarém, Setúbal (central row, left to right), Viana do Castelo, Vila Real and Viseu (bottom row, left to right)

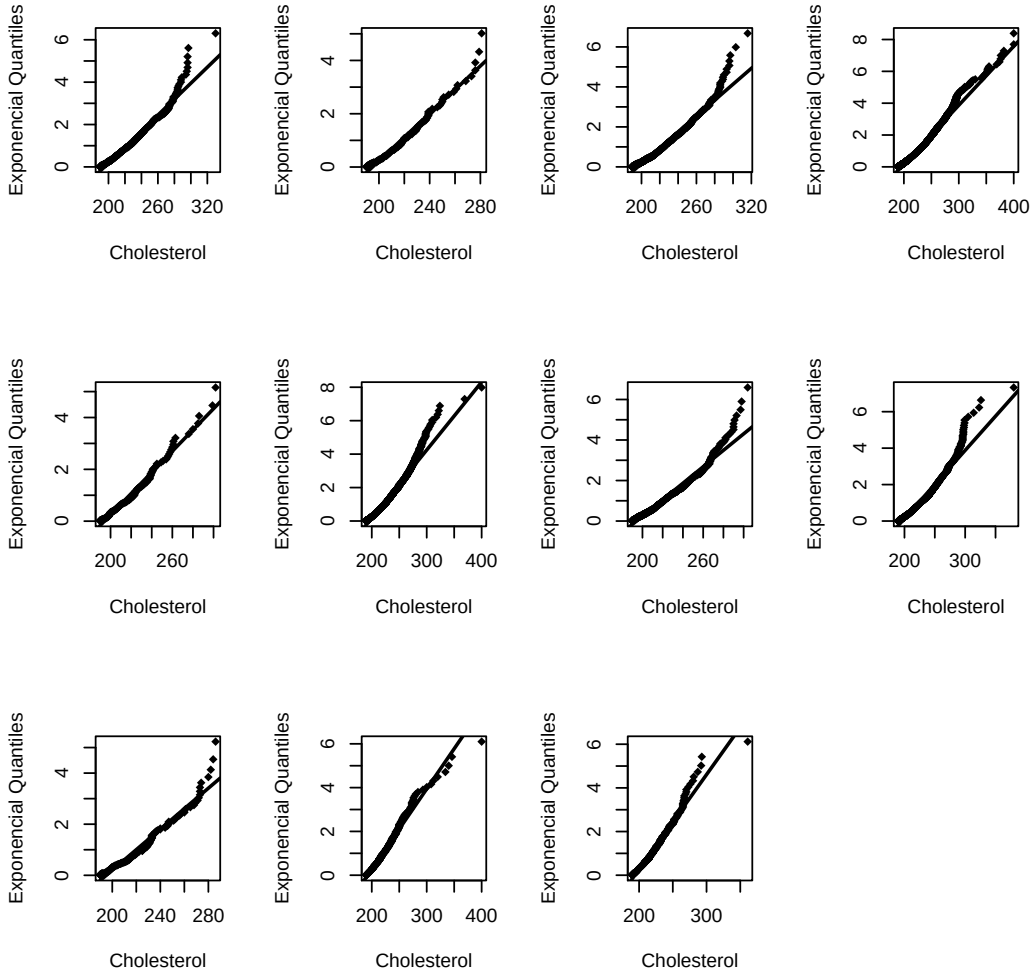

Figure 17: Exponential QQ-plot for the data exceeding the reference value of 190 mm/dl - R. A. Madeira, R. A. Açores, Leiria, Lisboa (top row, left to right), Portalegre, Porto, Santarém, Setúbal (central row, left to right) and Viana do Castelo, Vila Real and Viseu (bottom row, left to right)

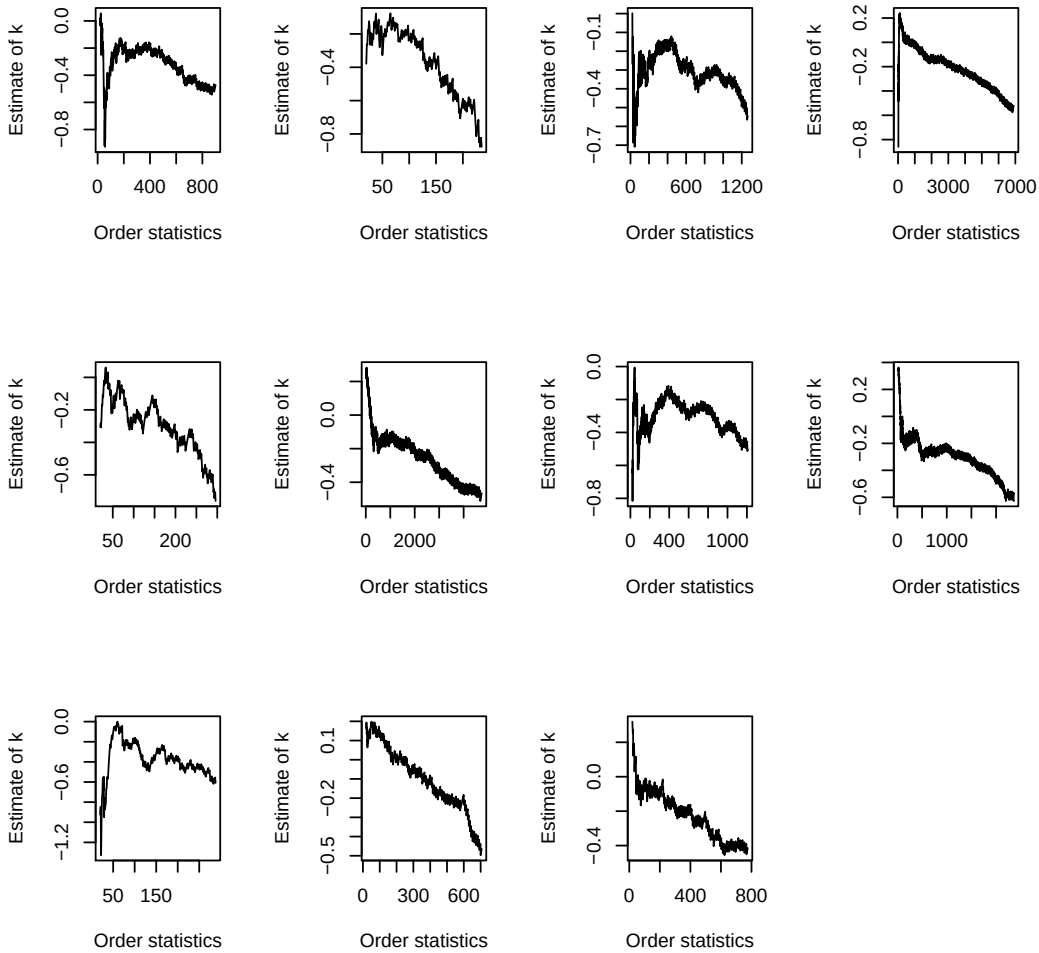

Figure 18: Moments estimates for k - R. A. Madeira, R. A. Açores, Leiria, Lisboa (top row, left to right), Portalegre, Porto, Santarém, Setúbal (central row, left to right), Viana do Castelo, Vila Real and Viseu (bottom row, left to right)

District	u	k	σ	Number Exceed.	Anderson Darling	Cramer Von Mises
Aveiro	230	-0.02	28.255	424	1.0722	0.0990
Aveiro	240	-0.06	30.176	285	1.4765	0.1945
Aveiro	250	0.01	26.318	213	0.8537	0.1052
Aveiro	260	0.06	24.244	148	0.5820	0.0755
Aveiro	270	0.04	25.859	97	0.8876	0.1623
Aveiro	280	0.18	21.291	69	0.7041	0.1249
Aveiro	290	0.05	27.887	41	1.2481	0.2231
Aveiro	300	-0.81	75.921	23	0.7222	0.1223
Beja	220	-0.20	28.713	118	0.8879	0.1262
Beja	230	-0.16	24.143	86	1.1358	0.1681
Beja	240	-0.12	20.665	59	1.1329	0.1764
Beja	250	-0.04	16.733	37	0.9160	0.1389
Beja	260	0.15	10.702	23	0.4470	0.0638
Braga	240	0.10	21.382	258	2.3940	0.4027
Braga	250	0.13	20.870	170	4.4096	0.6803
Braga	260	0.24	16.334	117	1.9486	0.2859
Braga	270	0.42	13.183	71	0.9660	0.1470
Braga	280	0.68	11.339	38	0.5941	0.0879
Braga	290	0.94	13.373	19	0.5997	0.1056
Bragança	230	0.07	22.295	80	0.8233	0.1227
Bragança	240	0.16	17.969	57	0.4496	0.0489
Bragança	250	0.21	17.524	35	0.4289	0.0569
Bragança	260	0.31	15.455	22	0.4055	0.0579
Castelo Branco	220	-0.22	28.230	92	0.6150	0.0998
Castelo Branco	230	-0.41	34.352	57	0.2868	0.0379
Castelo Branco	240	-0.52	35.007	39	0.5668	0.0929
Castelo Branco	250	-0.45	27.310	29	0.5169	0.0997
Castelo Branco	260	-0.73	31.028	18	1.3284	0.2218
Coimbra	240	0.13	27.619	138	0.3625	0.0409
Coimbra	250	0.13	29.295	96	0.3093	0.0538
Coimbra	260	0.06	35.879	63	1.2190	0.1824
Coimbra	270	0.18	28.122	53	0.4472	0.0608
Coimbra	280	0.23	27.784	38	0.6060	0.0688
Coimbra	290	0.27	29.068	26	0.2397	0.0302
Coimbra	300	0.15	40.087	17	0.2234	0.0248
Évora	200	-0.31	36.868	251	0.6988	0.1055
Évora	210	-0.37	36.454	182	0.6923	0.1078
Évora	220	-0.44	36.751	129	0.5384	0.0895
Évora	230	-0.48	34.125	94	0.8573	0.1041
Évora	240	-0.56	32.484	66	0.3794	0.0666
Évora	250	-0.65	30.440	44	0.4695	0.0631
Évora	260	-0.70	25.154	30	0.5526	0.1061
Évora	270	-0.59	16.048	19	0.8483	0.1560
Faro	210	-0.05	33.538	256	0.9242	0.1272
Faro	220	-0.03	31.537	192	1.0217	0.1207
Faro	230	0.00	29.794	140	0.6833	0.0876
Faro	240	0.04	27.938	101	0.5583	0.0741
Faro	250	-0.04	33.227	65	0.7540	0.1256
Faro	260	0.05	27.894	51	0.6433	0.1226
Faro	270	0.02	29.992	35	1.6671	0.2730
Faro	280	0.24	21.766	26	0.7588	0.1229
Faro	290	0.49	18.944	16	0.4771	0.0745
Guarda	190	-0.15	30.935	169	0.5200	0.0600
Guarda	200	-0.18	31.228	116	0.2810	0.0376
Guarda	210	-0.15	28.047	85	0.3050	0.0421
Guarda	220	-0.21	29.223	56	0.1917	0.0257
Guarda	230	-0.25	29.045	38	0.2420	0.0388
Guarda	240	-0.28	27.663	26	0.4124	0.0545
Guarda	250	-0.43	30.308	17	0.3446	0.0682
R. A. da Madeira	190	-0.32	46.666	533	3.7650	0.6986
R. A. da Madeira	220	-0.25	32.174	229	0.8142	0.0972
R. A. da Madeira	240	-0.34	33.888	110	1.3629	0.1832
R. A. da Madeira	250	-0.28	26.151	76	1.0344	0.1854
R. A. da Madeira	260	-0.24	21.772	53	1.6698	0.2934
R. A. da Madeira	270	-0.11	14.207	38	0.7068	0.0886

Table 12: Test statistics of the Anderson-Darling and the Cramer-Von Mises goodness-of-fit tests

District	u	k	σ	Number Exceed.	Anderson Darling	Cramer Von Mises
R. A. Açores	190	-0.33	36.268	148	0.5130	0.0754
R. A. Açores	200	-0.29	30.905	114	0.4307	0.0479
R. A. Açores	210	-0.29	28.107	80	0.5268	0.0595
R. A. Açores	220	-0.41	30.241	50	0.6571	0.0941
R. A. Açores	230	-0.50	29.405	33	0.7411	0.1338
Leiria	190	-0.37	47.936	776	6.4448	1.2010
Leiria	240	-0.31	26.997	156	0.4774	0.0552
Leiria	250	-0.34	24.924	101	0.7088	0.0705
Leiria	260	-0.35	22.076	63	0.6908	0.1086
Leiria	270	-0.38	20.145	42	0.7799	0.1265
Leiria	280	-0.18	11.350	23	0.6759	0.1028
Lisboa	190	-0.16	38.048	4290	23.2285	3.5209
Lisboa	240	-0.03	23.469	944	3.1032	0.2862
Lisboa	250	-0.01	22.395	614	2.6254	0.2491
Lisboa	260	0.00	22.191	385	2.6879	0.3770
Lisboa	270	0.09	18.838	259	1.9368	0.2660
Lisboa	280	0.14	18.672	151	2.6737	0.4416
Lisboa	290	0.40	14.449	90	1.4163	0.2173
Lisboa	300	-0.34	44.765	42	0.6768	0.1056
Portalegre	190	-0.26	36.394	167	1.0213	0.1160
Portalegre	220	-0.16	24.700	67	0.5350	0.0599
Portalegre	230	-0.15	23.100	44	0.5022	0.0783
Portalegre	240	-0.30	27.141	25	0.4166	0.0646
Portalegre	250	-0.37	25.934	17	0.6222	0.1255
Porto	190	-0.22	47.746	2859	81.0967	14.4548
Porto	240	-0.09	22.908	523	4.0614	0.5608
Porto	250	-0.12	24.939	337	4.2843	0.5908
Porto	260	-0.02	17.348	220	2.5970	0.3464
Porto	270	0.08	13.733	132	0.8647	0.1080
Porto	280	0.14	13.330	63	0.8574	0.1077
Porto	290	0.26	12.911	29	0.3848	0.0477
Santarém	190	-0.29	38.302	705	0.9101	0.1230
Santarém	240	-0.39	27.236	130	0.6831	0.0818
Santarém	250	-0.60	33.228	82	2.7593	0.5712
Santarém	260	-0.35	18.547	55	0.6673	0.1158
Santarém	270	-0.62	22.105	24	0.1787	0.0223
Santarém	280	-0.63	16.131	14	0.2615	0.0399
Setúbal	190	-0.20	39.758	1495	7.3144	1.0768
Setúbal	240	-0.16	27.525	357	2.9874	0.4338
Setúbal	250	-0.15	25.572	231	2.0087	0.2449
Setúbal	260	-0.14	23.788	146	1.9784	0.2510
Setúbal	270	-0.03	16.040	94	1.5109	0.2502
Setúbal	280	0.04	12.909	56	1.9043	0.3297
Setúbal	290	0.29	7.945	28	1.3832	0.2610
Viana do Castelo	190	-0.39	42.763	180	0.7614	0.1020
Viana do Castelo	200	-0.44	41.801	134	1.1428	0.1958
Viana do Castelo	210	-0.35	32.952	110	0.8729	0.1446
Viana do Castelo	220	-0.42	32.425	76	1.5598	0.2918
Viana do Castelo	230	-0.38	27.000	54	2.7991	0.4007
Vila Real	190	-0.09	32.858	435	3.1012	0.4427
Vila Real	200	-0.04	28.462	338	1.2589	0.1718
Vila Real	210	-0.01	25.898	245	0.9896	0.1262
Vila Real	220	0.01	25.173	166	1.6286	0.2041
Vila Real	230	0.10	21.318	117	0.5189	0.0633
Vila Real	240	0.14	20.880	74	0.5397	0.0633
Vila Real	250	0.12	24.362	43	0.3366	0.0479
Vila Real	260	0.06	29.015	27	0.6069	0.1087
Viseu	190	-0.14	31.172	439	2.6973	0.4055
Viseu	200	-0.12	27.937	327	1.7472	0.2516
Viseu	210	-0.09	24.750	233	1.6538	0.1808
Viseu	220	-0.06	22.194	157	0.6040	0.0630
Viseu	230	-0.04	20.668	100	0.5928	0.0711
Viseu	240	-0.03	20.053	61	1.0548	0.1559
Viseu	250	0.07	15.364	41	0.6803	0.1190
Viseu	260	0.24	10.730	25	0.5934	0.0969

Table 13: Test statistics of the Anderson-Darling and the Cramer-von Mises goodness-of-fit tests (cont.)

District	u	$\hat{k}$	$\hat{\sigma}$	Number exced.	n	% upper Order Stat.	CI for k(95%)	CI for σ (95%)
Aveiro	230	-0.02	28.25	424	2721	15.6	-0.11 ; 0.07	24.54 ; 31.97
Aveiro	240	-0.06	30.18	285	2721	10.5	-0.16 ; 0.05	25.51 ; 34.84
Aveiro	250	0.01	26.32	213	2721	7.8	-0.13 ; 0.14	21.33 ; 31.31
Aveiro	260	0.06	24.24	148	2721	5.4	-0.12 ; 0.24	18.38 ; 30.11
Aveiro	270	0.04	25.86	97	2721	3.6	-0.19 ; 0.26	18.11 ; 33.61
Aveiro	280	0.18	21.29	69	2721	2.5	-0.15 ; 0.51	12.69 ; 29.89
Aveiro	290	0.05	27.89	41	2721	1.5	-0.49 ; 0.59	10.54 ; 45.24
Aveiro	300	-0.81	75.92	23	2721	0.8	-1.38 ; -0.24	29.12 ; 122.72
Beja	220	-0.20	28.71	118	479	24.6	-0.30 ; -0.09	22.76 ; 34.67
Beja	230	-0.16	24.14	86	479	18.0	-0.28 ; -0.03	18.28 ; 30.01
Beja	240	-0.12	20.66	59	479	12.3	-0.28 ; 0.05	14.39 ; 26.94
Beja	250	-0.04	16.73	37	479	7.7	-0.29 ; 0.20	9.99 ; 23.47
Beja	260	0.15	10.70	23	479	4.8	-0.26 ; 0.57	4.48 ; 16.92
Beja	270	0.31	11.50	8	479	1.7	-0.72 ; 1.35	-2.60 ; 25.60
Braga	240	0.10	21.38	258	2463	10.5	0.00 ; 0.20	18.04 ; 24.73
Braga	250	0.13	20.87	170	2463	6.9	0.01 ; 0.24	16.94 ; 24.80
Braga	260	0.24	16.33	117	2463	4.8	0.07 ; 0.42	12.29 ; 20.38
Braga	270	0.42	13.18	71	2463	2.9	0.13 ; 0.70	8.54 ; 17.82
Braga	280	0.68	11.34	38	2463	1.5	0.19 ; 1.17	5.16 ; 17.52
Braga	290	0.94	13.37	19	2463	0.8	0.06 ; 1.82	1.42 ; 25.33
Bragança	230	0.07	22.29	80	758	10.6	-0.10 ; 0.23	16.14 ; 28.45
Bragança	240	0.16	17.97	57	758	7.5	-0.09 ; 0.41	11.55 ; 24.39
Bragança	250	0.21	17.52	35	758	4.6	-0.12 ; 0.53	9.52 ; 25.53
Bragança	260	0.31	15.45	22	758	2.9	-0.14 ; 0.76	6.15 ; 24.76
Castelo Branco	220	-0.22	28.23	92	495	18.6	-0.45 ; 0.01	19.68 ; 36.78
Castelo Branco	230	-0.41	34.35	57	495	11.5	-0.68 ; -0.15	22.18 ; 46.53
Castelo Branco	240	-0.52	35.01	39	495	7.9	-0.86 ; -0.18	19.87;50.15
Castelo Branco	250	-0.45	27.31	29	495	5.9	-0.93 ; 0.03	11.52 ; 43.10
Castelo Branco	260	-0.73	31.03	18	495	3.6	-1.55 ; 0.10	3.70 ; 58.35
Coimbra	240	0.13	27.62	138	1156	11.9	-0.05 ; 0.31	20.87 ; 34.37
Coimbra	250	0.13	29.29	96	1156	8.3	-0.08 ; 0.34	20.84 ; 37.75
Coimbra	260	0.06	35.88	63	1156	5.4	-0.15 ; 0.27	24.24 ; 47.52
Coimbra	270	0.18	28.12	53	1156	4.6	-0.11 ; 0.47	17.04 ; 39.20
Coimbra	280	0.23	27.78	38	1156	3.3	-0.14 ; 0.59	14.40 ; 41.17
Coimbra	290	0.27	29.07	26	1156	2.2	-0.23 ; 0.77	10.91 ; 47.23
Coimbra	300	0.15	40.09	17	1156	1.5	-0.38 ; 0.69	11.55 ; 68.62
Évora	200	-0.31	36.87	251	664	37.8	-0.44 ; -0.19	30.55 ; 43.19
Évora	210	-0.37	36.45	182	664	27.4	-0.51 ; -0.22	29.22 ; 43.69
Évora	220	-0.44	36.75	129	664	19.4	-0.61 ; -0.28	28.46 ; 45.04
Évora	230	-0.48	34.13	94	664	14.2	-0.68 ; -0.29	25.11 ; 43.14
Évora	240	-0.56	32.48	66	664	9.9	-0.79 ; -0.33	22.57 ; 42.40
Évora	250	-0.65	30.44	44	664	6.6	-0.95 ; -0.36	19.20 ; 41.68
Évora	260	-0.70	25.15	30	664	4.5	-1.09 ; -0.31	13.39 ; 36.91
Évora	270	-0.59	16.05	19	664	2.9	-1.27 ; 0.10	3.55 ; 28.55
Faro	210	-0.05	33.54	256	687	37.3	-0.15 ; 0.05	28.20 ; 38.88
Faro	220	-0.03	31.54	192	687	27.9	-0.15 ; 0.10	25.61 ; 37.47
Faro	230	0.00	29.79	140	687	20.4	-0.16 ; 0.16	23.01 ; 36.58
Faro	240	0.04	27.94	101	687	14.7	-0.16 ; 0.24	20.07 ; 35.81
Faro	250	-0.04	33.23	65	687	9.5	-0.26 ; 0.19	22.27 ; 44.19
Faro	260	0.05	27.89	51	687	7.4	-0.24 ; 0.34	16.78 ; 39.01
Faro	270	0.02	29.99	35	687	5.1	-0.31 ; 0.36	15.79 ; 44.19
Faro	280	0.24	21.77	26	687	3.8	-0.27 ; 0.76	7.95 ; 35.59
Faro	290	0.49	18.94	16	687	2.3	-0.57 ; 1.55	-2.66 ; 40.55
Guarda	190	-0.15	30.93	169	293	57.7	-0.29 ; -0.01	24.58 ; 37.29
Guarda	200	-0.18	31.23	116	293	39.6	-0.35 ; -0.01	23.51 ; 38.94
Guarda	210	-0.15	28.05	85	293	29.0	-0.37 ; 0.07	19.55 ; 36.55
Guarda	220	-0.21	29.22	56	293	19.1	-0.48 ; 0.06	18.34 ; 40.11
Guarda	203	-0.19	31.35	104	293	35.5	-0.37 ; -0.01	23.23 ; 39.48
Guarda	240	-0.28	27.66	26	293	8.9	-0.73 ; 0.18	11.42 ; 43.90
Guarda	250	-0.43	30.31	17	293	5.8	-0.98 ; 0.12	9.11 ; 51.51
R. A. da Madeira	190	-0.32	46.67	533	1001	53.2	-0.36 ; -0.28	41.90 ; 51.43
R. A. da Madeira	220	-0.25	32.17	229	1001	22.9	-0.34 ; -0.17	27.29 ; 37.05
R. A. da Madeira	240	-0.34	33.89	110	1001	11.0	-0.46 ; -0.23	25.81 ; 41.96
R. A. da Madeira	250	-0.28	26.15	76	1001	7.6	-0.41 ; -0.14	19.47 ; 32.83
R. A. da Madeira	260	-0.24	21.77	53	1001	5.3	-0.41 ; -0.08	15.14 ; 28.40
R. A. da Madeira	270	-0.11	14.21	38	1001	3.8	-0.34 ; 0.12	8.63 ; 19.78

Table 14: All the models fitted to the data

District	u	$\hat{k}$	$\hat{\sigma}$	Number exced.	n	% upper Order Stat.	CI for k(95%)	CI for σ (95%)
R. A. Açores	190	-0.33	36.27	148	261	56.7	-0.47 ; -0.19	28.78 ; 43.76
R. A. Açores	200	-0.29	30.91	114	261	43.7	-0.46 ; -0.11	23.26 ; 38.55
R. A. Açores	210	-0.29	28.11	80	261	30.7	-0.51 ; -0.06	19.40 ; 36.82
R. A. Açores	220	-0.41	30.24	50	261	19.2	-0.73 ; -0.09	17.83 ; 42.65
R. A. Açores	230	-0.50	29.40	33	261	12.6	-1.01 ; 0.00	12.18 ; 46.63
Leiria	190	-0.37	47.94	776	1404	55.3	-0.41 ; -0.33	43.89 ; 51.99
Leiria	240	-0.31	27.00	156	1404	11.1	-0.42 ; -0.21	22.02 ; 31.97
Leiria	250	-0.34	24.92	101	1404	7.2	-0.47 ; -0.21	19.31 ; 30.54
Leiria	260	-0.35	22.08	63	1404	4.5	-0.51 ; -0.18	15.94 ; 28.21
Leiria	270	-0.38	20.15	42	1404	3.0	-0.63 ; -0.14	11.72 ; 28.58
Leiria	280	-0.18	11.35	23	1404	1.6	-0.52 ; 0.15	5.40 ; 17.30
Lisboa	190	-0.16	38.05	4290	7671	55.9	-0.17 ; -0.14	36.79 ; 39.30
Lisboa	240	-0.03	23.47	944	7671	12.3	-0.08 ; 0.03	21.50 ; 25.44
Lisboa	250	-0.01	22.39	614	7671	8.0	-0.08 ; 0.06	20.03 ; 24.76
Lisboa	260	0.00	22.19	385	7671	5.0	-0.08 ; 0.09	19.21 ; 25.17
Lisboa	270	0.09	18.84	259	7671	3.4	-0.03 ; 0.22	15.55 ; 22.13
Lisboa	280	0.14	18.67	151	7671	2.0	-0.04 ; 0.32	14.18 ; 23.16
Lisboa	290	0.40	14.45	90	7671	1.2	0.05 ; 0.74	8.77 ; 20.13
Lisboa	300	-0.34	44.77	42	7671	0.5	-0.69 ; 0.02	24.43 ; 65.10
Portalegre	190	-0.26	36.39	167	329	50.8	-0.37 ; -0.14	29.58 ; 43.21
Portalegre	220	-0.16	24.70	67	329	20.4	-0.40 ; 0.08	16.37 ; 33.03
Portalegre	230	-0.15	23.10	44	329	13.4	-0.48 ; 0.17	13.02 ; 33.18
Portalegre	240	-0.30	27.14	25	329	7.6	-0.79 ; 0.18	10.56 ; 43.72
Portalegre	250	-0.37	25.93	17	329	5.2	-1.07 ; 0.33	4.51 ; 47.36
Porto	190	-0.22	47.75	2859	5262	54.3	-0.24 ; -0.21	45.31 ; 50.18
Porto	240	-0.09	22.91	523	5262	9.9	-0.14 ; -0.04	20.67 ; 25.14
Porto	250	-0.12	24.94	337	5262	6.4	-0.17 ; -0.07	21.50 ; 28.38
Porto	260	-0.02	17.35	220	5262	4.2	-0.11 ; 0.07	14.57 ; 20.13
Porto	270	0.08	13.73	132	5262	2.5	-0.07 ; 0.23	10.62 ; 16.85
Porto	280	0.14	13.33	63	5262	1.2	-0.10 ; 0.38	8.80 ; 17.86
Porto	290	0.26	12.91	29	5262	0.6	-0.16 ; 0.69	5.79 ; 20.03
Santarém	190	-0.29	38.30	705	1342	52.5	-0.35 ; -0.23	34.75 ; 41.86
Santarém	240	-0.39	27.24	130	1342	9.7	-0.52 ; -0.25	21.57 ; 32.90
Santarém	250	-0.60	33.23	82	1342	6.1	-0.81 ; -0.39	22.96 ; 43.50
Santarém	260	-0.35	18.55	55	1342	4.1	-0.60 ; -0.10	11.98 ; 25.11
Santarém	270	-0.62	22.11	24	1342	1.8	-1.00 ; -0.24	11.15 ; 33.06
Santarém	280	-0.63	16.13	14	1342	1.0	-1.22 ; -0.03	4.50 ; 27.76
Setúbal	190	-0.20	39.76	1495	2620	57.1	-0.22 ; -0.18	37.65 ; 41.87
Setúbal	240	-0.16	27.53	357	2620	13.6	-0.21 ; -0.11	23.93 ; 31.12
Setúbal	250	-0.15	25.57	231	2620	8.8	-0.22 ; -0.09	21.37 ; 29.78
Setúbal	260	-0.14	23.79	146	2620	5.6	-0.23 ; -0.05	18.73 ; 28.84
Setúbal	270	-0.03	16.04	94	2620	3.6	-0.17 ; 0.12	12.08 ; 20.00
Setúbal	280	0.04	12.91	56	2620	2.1	-0.15 ; 0.23	8.71 ; 17.10
Setúbal	290	0.29	7.95	28	2620	1.1	-0.10 ; 0.69	3.74 ; 12.16
Viana do Castelo	190	-0.39	42.76	180	320	56.3	-0.52 ; -0.26	34.70 ; 50.82
Viana do Castelo	200	-0.44	41.80	134	320	41.9	-0.59 ; -0.28	32.72 ; 50.88
Viana do Castelo	210	-0.35	32.95	110	320	34.4	-0.55 ; -0.15	24.21 ; 41.69
Viana do Castelo	220	-0.42	32.43	76	320	23.8	-0.68 ; -0.16	21.75 ; 43.11
Viana do Castelo	230	-0.38	27.00	54	320	16.9	-0.80 ; 0.04	13.92 ; 40.08
Vila Real	190	-0.09	32.86	435	784	55.5	-0.15 ; -0.03	29.18 ; 36.53
Vila Real	200	-0.04	28.46	338	784	43.1	-0.13 ; 0.04	24.62 ; 32.30
Vila Real	210	-0.01	25.90	245	784	31.3	-0.11 ; 0.10	21.65 ; 30.15
Vila Real	220	0.01	25.17	166	784	21.2	-0.12 ; 0.14	20.13 ; 30.21
Vila Real	230	0.10	21.32	117	784	14.9	-0.09 ; 0.29	15.77 ; 26.86
Vila Real	240	0.14	20.88	74	784	9.4	-0.11 ; 0.40	13.74 ; 28.02
Vila Real	250	0.12	24.36	43	784	5.5	-0.23 ; 0.47	13.11 ; 35.61
Vila Real	260	0.06	29.01	27	784	3.4	-0.39 ; 0.50	12.05 ; 45.98
Viseu	190	-0.14	31.17	439	861	51.0	-0.19 ; -0.10	27.93 ; 34.41
Viseu	200	-0.12	27.94	327	861	38.0	-0.18 ; -0.06	24.46 ; 31.41
Viseu	210	-0.09	24.75	233	861	27.1	-0.18 ; -0.01	20.99 ; 28.51
Viseu	220	-0.06	22.19	157	861	18.2	-0.17 ; 0.05	17.92 ; 26.47
Viseu	230	-0.04	20.67	100	861	11.6	-0.19 ; 0.10	15.63 ; 25.71
Viseu	240	-0.03	20.05	61	861	7.1	-0.21 ; 0.14	13.92 ; 26.18
Viseu	250	0.07	15.36	41	861	4.8	-0.19 ; 0.32	9.21 ; 21.51
Viseu	260	0.24	10.73	25	861	2.9	-0.18 ; 0.66	4.64 ; 16.82

Table 15: All the models fitted to the data (cont.)

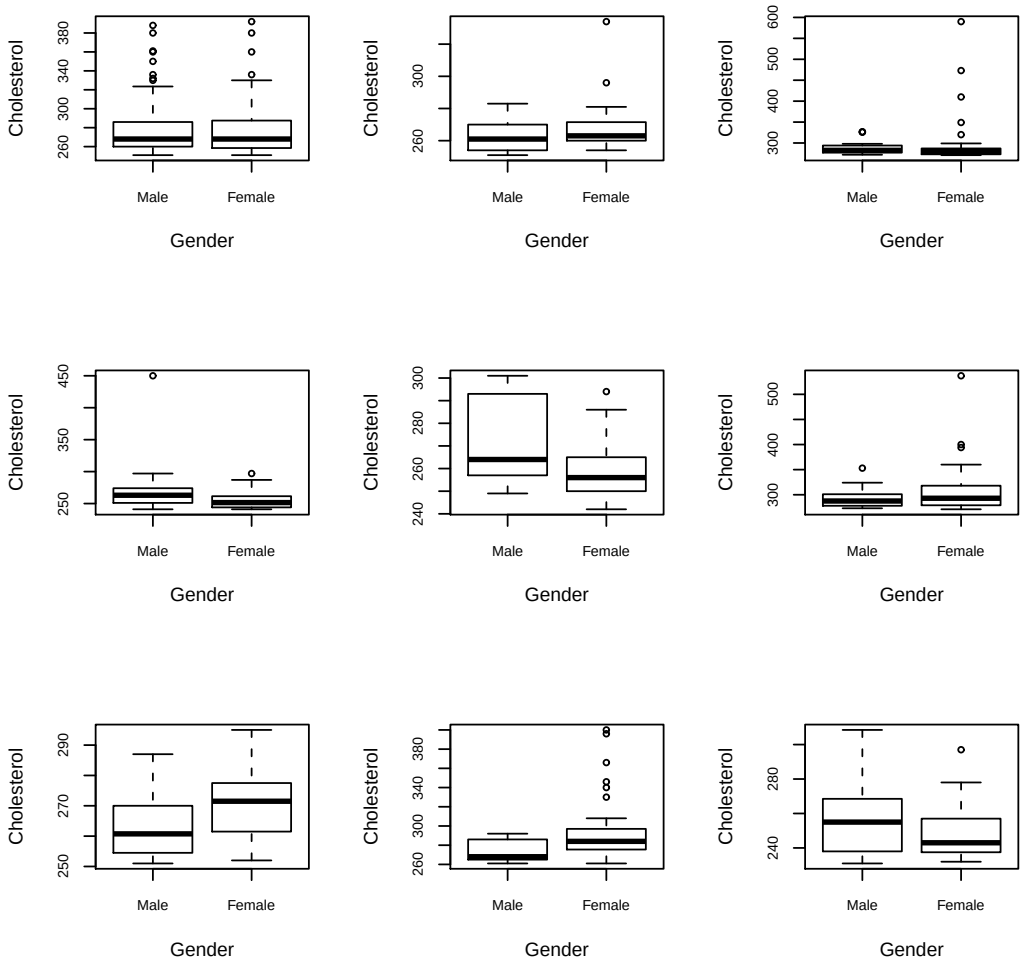

Figure 19: High total cholesterol levels distributed by gender and district - Aveiro, Beja, Braga, (top row, left to right), Bragança, Castelo Branco, Coimbra (central row, left to right), Évora, Faro and Guarda (bottom row, left to right)

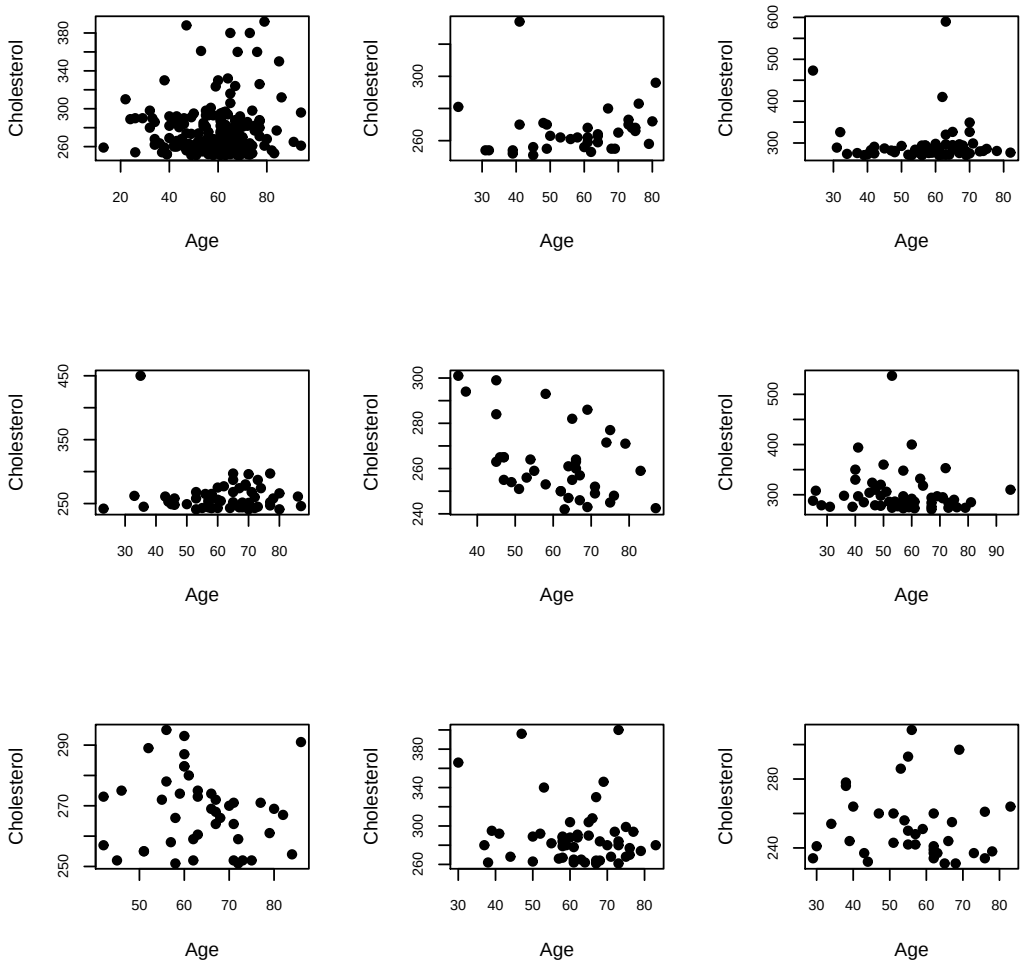

Figure 20: High total cholesterol levels distributed by age and district - Aveiro, Beja, Braga, (top row, left to right), Bragança, Castelo Branco, Coimbra (central row, left to right), Évora, Faro and Guarda (bottom row, left to right)

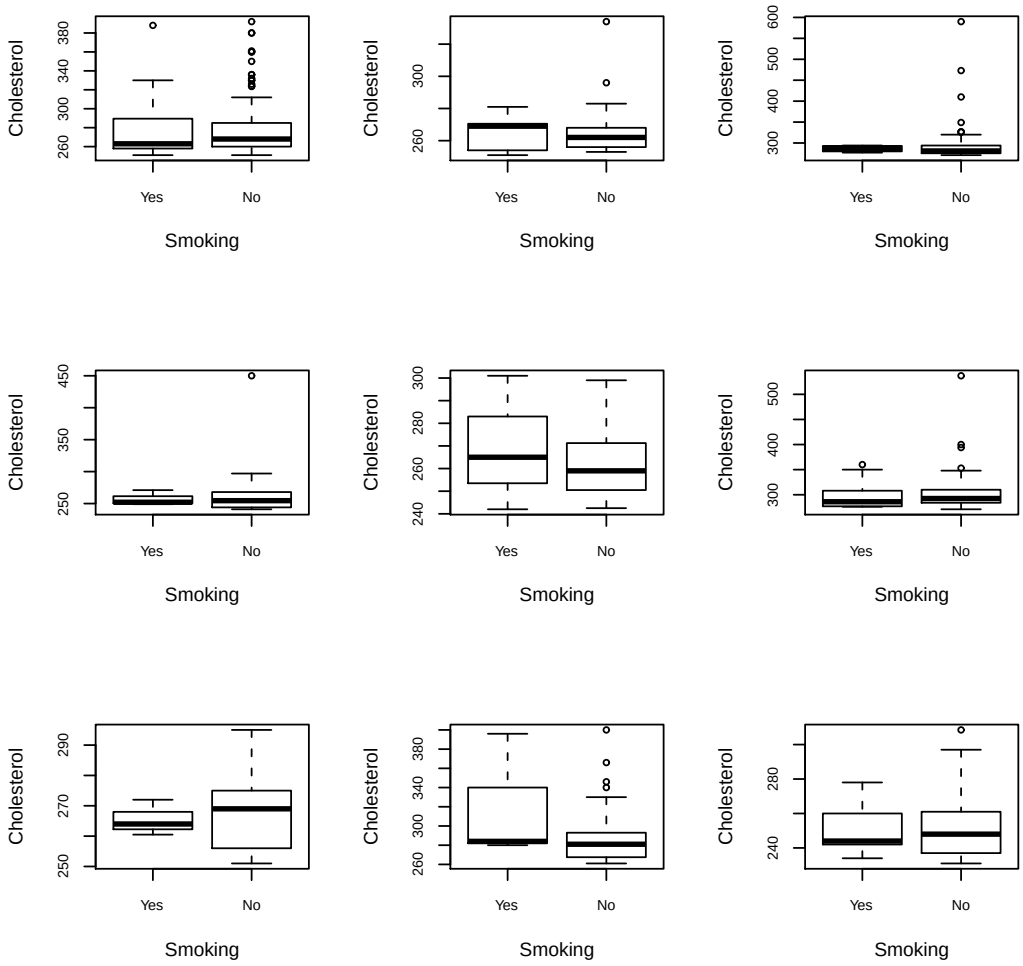

Figure 21: High total cholesterol levels distributed by smoking habits and district - Aveiro, Beja, Braga, (top row, left to right) Bragança, Castelo Branco, Coimbra (central row, left to right) Évora, Faro e Guarda (bottom row, left to right)

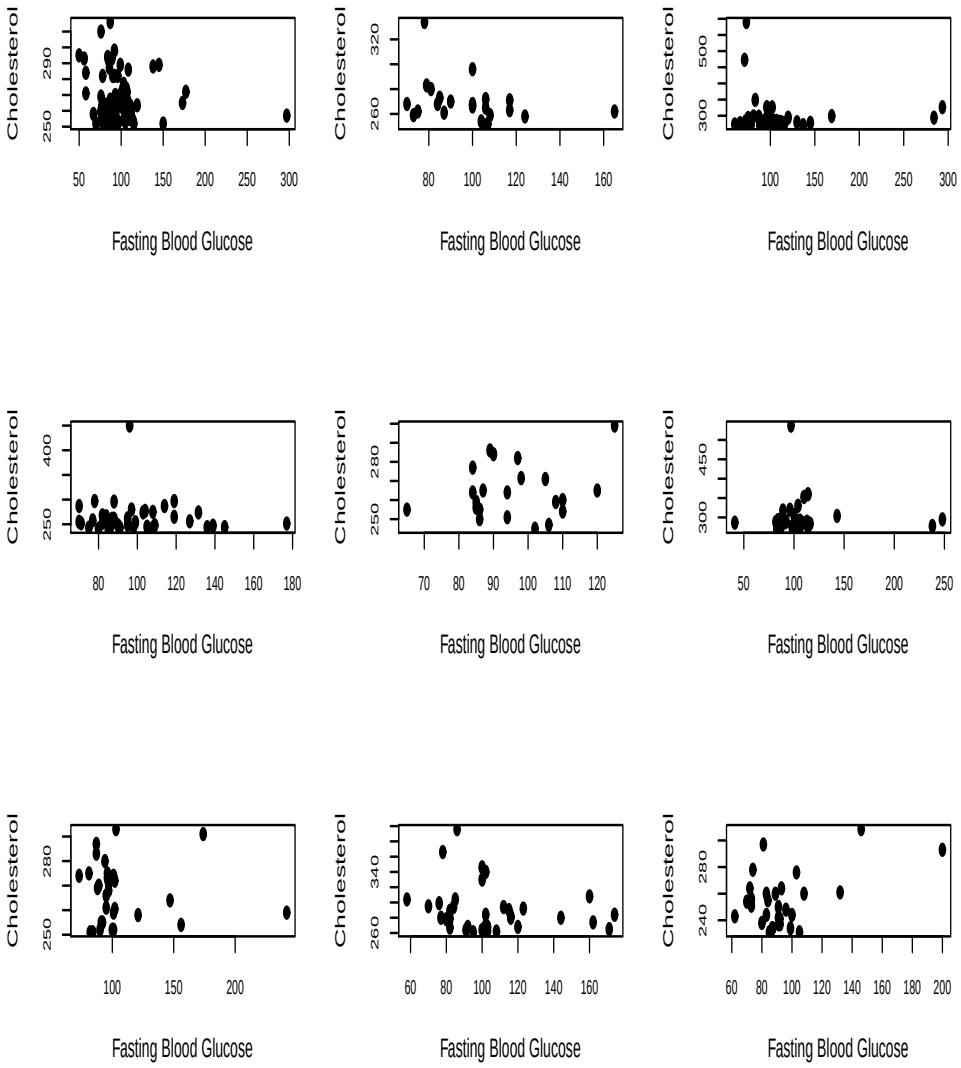

Figure 22: High total cholesterol levels distributed by fasting blood glucose and district - Aveiro, Beja, Braga, (top row, left to right), Bragança, Castelo Branco, Coimbra (central row, left to right), Évora, Faro and Guarda (bottom row, left to right)

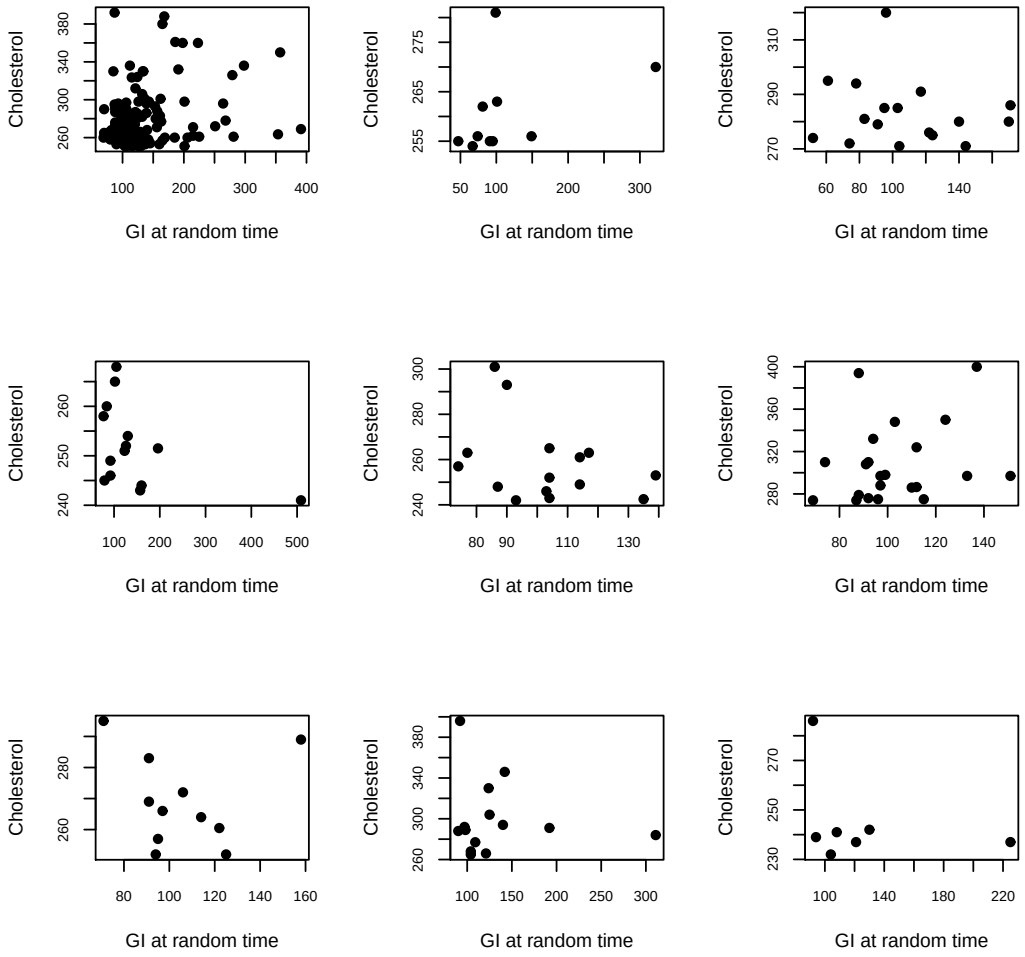

Figure 23: High total cholesterol levels distributed by blood glucose at random time and district - Aveiro, Beja, Braga, (top row, left to right), Bragança, Castelo Branco, Coimbra (central row, left to right), Évora, Faro and Guarda (bottom row, left to right)

u	**District**	**Variable**	**Mean**	**Median**	**Sd**
250	Aveiro	Triglyceride level	200.4	185	85.7
250	Beja	Triglyceride level	237.0	168	146.4
270	Braga	Triglyceride level	259.9	237	100.2
240	Bragança	Triglyceride level	274.4	250	116.1
240	Castelo Branco	Triglyceride level	234.6	248	86.1
270	Coimbra	Triglyceride level	212.3	209	63.8
250	Évora	Triglyceride level	275.4	233	136.7
260	Faro	Triglyceride level	228.9	184	115.2
230	Guarda	Triglyceride level	185.1	164	67.0
250	Aveiro	Fasting blood glucose	97.6	93	31.6
250	Beja	Fasting blood glucose	98.3	100	21.1
270	Braga	Fasting blood glucose	103.3	96	43.0
240	Bragança	Fasting blood glucose	99.9	95	23.3
240	Castelo Branco	Fasting blood glucose	95.9	94	13.9
270	Coimbra	Fasting blood glucose	106.9	100	40.1
250	Évora	Fasting blood glucose	105.2	96	33.3
260	Faro	Fasting blood glucose	104.2	101	28.1
230	Guarda	Fasting blood glucose	94.1	90	26.4
250	Aveiro	Blood glucose at random time	135.3	116	59.6
250	Beja	Blood glucose at random time	112.6	93	78.3
270	Braga	Blood glucose at random time	107.4	103	34.8
240	Bragança	Blood glucose at random time	145.1	114	110.2
240	Castelo Branco	Blood glucose at random time	102.7	104	19.1
270	Coimbra	Blood glucose at random time	102.8	97	20.1
250	Évora	Blood glucose at random time	105.8	97	23.3
260	Faro	Blood glucose at random time	132.1	115	58.1
230	Guarda	Blood glucose at random time	124.9	108	46.2

Table 16: Characteristics of the extreme individuals - Aveiro, Beja, Braga, Bragança, Castelo Branco, Évora, Faro and Guarda