Test of modeling sleepdisorders:quality of sleep

AbdellatifELLOUMI[1],Imen FARHAT[2], Mohammed DOGUI[3] and Abdeljelil FARHAT[1]

[1]EAS-Mahdia,Faculté des Sciences Economiques et de Gestion de Mahdia,
Université Monastir ,Tunisia.
[2]Faculté de Médecine, Université Laval, Canada.
[3]Chef de Service d'Explorations Fonctionnelles du Système Nerveux
CHU"Sahloul", Sousse, Tunisia.

ABSTRACT: We are trying to explain the quality of sleep. We have selected the Total Time of sleep as the endogenous and the variables of the structure of sleep as exogenous. Then, we have integrated the qualitative variable "quality of snoring" in the model. What attracts our attention is that the quality of the snoring is the most important variable in the estimate of the quality of sleep. This variable has the highest coefficient. If we want to quickly 1improve the quality of our sleep, it is enough to monitor our snoring.

KEYWORDS: Biostatistics, help to the medical decision, sleep disorder, modeling of the qualitative variables, tests of distribution of variables, explanation of the quality of sleep.

RESUME : Nous cherchons à expliquer la qualité de sommeil. Nous avons sélectionné le Temps Total du Sommeil comme étant l'endogène et les variables de la structure du sommeil comme des exogènes. Ensuite, nous avons intégré la variable qualitative "qualité du ronflement" dans le modèle. Ce qui attire notre attention est que la qualité du ronflement est la variable la plus importante dans l'estimation de la qualité de sommeil. Cette variable possède le coefficient le plus élevé. Si nous voulons améliorer rapidement notre qualité de sommeil, il suffit de contrôler notre ronflement.

MOTS-CLEFS: Bio-statistique, Aide à la décision médicale, Troubles de sommeil, Modélisation des variables qualitatives, Tests de distribution des variables, Explication de la qualité de sommei

1 INTRODUCTION

"Decreasing the amount of sleep does not mean decreasingthe need for sleep". This was the main result announced bySchafer and Olsen (2000). The human need for sleep depends on many different factors and life conditions (such as whereabouts of sleeping, napping, etc.). The daily need for sleep varies considerably between individuals. On average, humans need eight hours of sleep a day. Although, it has been shown in Dufour et al. (1998)that the time allocated to sleep has considerably decreased during the last decades, the variability for the need for sleep still exists. For instance, there are "very short sleepers" who only need four to five hours of sleep per day and there are well as "very long sleepers" who cannot do without eleven hours of sleep per day.
Staying awake for longer time periods and performing physical activities, in general, lead to longer period of sleep, and vice versa. This is an adaptive mechanism known a homeostatic mechanism. In practice, it is possible to tolerate the lack of sleep (night work shift, exceptional sleepless nights) if we follow the rules of the "predictive hemostasis" with the practice of preventive or curative naps.

2 DESCRIPTION OF THE DATA BASE

Our data base is constructed in 2015 at (Service d'ExplorationsFonctionnelles du SystèmeNerveux au Centre HospitalierUniversitaire "Sahloul")Functional Exploration of the Nervous System Service of Teaching Hospital of Sahloul, Sousse, Tunisia[1]. Our objective is to further explain sleeping troubles in patients focusing the on the quality of sleep in function of different health variables. This data base was constructed during 2012-2015 and includes 811 subjects and 35 variables.

2.1 BASE VARIABLES

The definitions of the variables are mostly derived from the works of Baguet et al. (2003) and Brun (2013)
Les définitions des variables sont en majorité extraites des travaux deKinugawa et al. (2014) et Baguet et al. (2003).

- Patient: name of the patient.
- Year: year of the exam.
- Age: age of the patient on the exam date.
- Sex: sex of the patient: (M or F)
- Medical history: a list of probable illnesses for every patient presented in the form of binary variables (0 or 1). The list is as follows:
    - Diabetes: Diabetes is disorder of digestion, usage and storage of sugars taken in as food. This causes a higher sugar levels in in the blood which is called hyperglycemia.
    - Hypertension (HTN), also known as high blood pressure (HBP) is a cardiovascular disease defined by a persistent high pressure of the blood in the arteries.
    - Dyslipidemia (Dys) is characterized by an abnormal (higher or lower) amount of lipids (cholesterols, tri-glycerides, phospholipids or free fatty acids) in the blood.
    - Stroke, also known as cerebrovascular accident (CVA) and sometime called brain attack is the sudden death of some brain cells due to the lack of oxygen because the blood flow is impaired by rupture of blockage of an artery in the brain.
    - Coronary artery disease (CAD) is a group of disorders in which the coronary arteries, the blood vessels that feed the heart are narrowed or blocked.
    - Other: which include other medical conditions such as depression, nasal congestion, breast cancer, osteoarthritis (OA), …

- Weight: patient weight in kg.
- Height: patient height in cm.
- Body mass index (BMI): in kg/m^2. BMI= $\frac{weight}{(Taille)^2}$.
- Snoring (Ronf) (Sno): binary variable indicating if the patient snore or not.
- History of snoring (Anct) (HisS): period measured in years starting when snoring first
- Quality of snoring
    - 0: No snoring
    - M: moderate snoring
    - L: lightly sever snoring
    - F: loud snoring
    - I: irregular snoring

We built a model for the quality of snoring using a decreasing scale. We gave “loud snoring” the number 1 and “lightly severe snoring” number 2. We followed with the rest of the label until “no snoring” with number 5.

- Excessive daytime sleepiness complaints (EDSC): it is a binary variable indicating if the patient is suffering from excessive sleep during the day or not. In certain cases, patients sleep on their chairs or on their steering wheel….
- Sleepiness scale « EPWORTH » (EPW): this variable is built on the bases of the answers provided to 8 questions where each is graded on a scale from 0 to 3. The sum of the answers

provided which is between 0 and 24 defines the "EPWORTH" sleepiness scale. Generally, if the value is less than 10, the person is considered normal. Otherwise, the patient is considered drowsy.

- Cephalalgia (Ceph): headaches. On symptom of the condition is waking up in the morning with a headache after sleeping well.
- Asthenia (Asth): weakness of the organism, physical fatigue.
- Latency of falling asleep (Lat) : is defined by the duration in minutes between closing the eyes (once the subject is in optimal environmental conditions for sleep) and the effective to sleep. Entry. This duration is compatible and observed using an electroencephalogram.
- Total time of sleep (TTS) : this variable is divided into four phases :
  - Light sleep stage 1 (S1) ;
  - Medium sleep stage 2 (S2) ;
  - Long or deep sleep stages 3 and 4 (S3) ;
  - Paradoxal Sleep (PS) : it the stage of sleep where dreams that we remember occur.

Each of these sleep stages are measured in minutes and percentages.

- Number of apnea (Nap) : the number of suspension of external breathing.
- Number of Hypopnea (Nhyp) : the number of anomaly low respiratory rate.
- Apnea-hypopnea respiratory index (IAH)[1] : it is and index used to measure the severity of the ampea sleep syndrome. $IAH = \frac{Nap + Nhyp}{temps} \times 60$

2.2 HANDLING THE MISSING VALUES :

The data base has many missing values. At the beginning, the data base included 811 observations. Based on different medical work but mainly Brun et al. (2013) andHorton and Stuart (2001), we started by localising and eliminating subjects whom the majority of valued were missing. From this process, we had 726 observations left. Then, we noticed that the majority of the observations had missing values in one or two variables. To avoid losing half of the data base because of missing variables, we used the work of Schafer and Olsen (2000) to adopt the following method:

- For quantitative variables, we graphically represented each variable to determine the nature of their probability law. Then, we generated the necessary number of missing observations following their respective probability law.
- For qualitative variables, we generated values using the uniform law while respecting each variable law properties.

3 DESCRIPTIVE STATISTICS OF SOME VARIABLES FROM THE DATA BASE

We start with the qualitative variables from the data base. We represent each variable with the appropriate graphic, box plots and density function. We discuss the nature of each variable depending on the shape of the graph and the normality coefficient, skewness and kurtosis.

Looking at the work of Papoulis and Pillai (2002), Bourbonnais and Terraza (2008) andDufour et al. (1998), the two most frequent normality tests in the literature are:

- Sh-W: normality test of Shapiro-Wilk. It is able to test the null hypothesis from which the sample is taking from a normality distributed population. It was published in 1965 by Samuel Shapiro and Martin Wilk
- K-S: Kolmogorov-Smirnov normality test is a hypothesis test to determine if a given sample is following a certain law given by its known cumulative distribution function or if two given samples follow the same law.

Our data base includes 726 observations of which 356 are males and 370 are females. The population is almost balanced in terms of the sex, since it contains 51% females and 49% males.

We notice that the majority of patients suffer from snoring (87.6%). Among this majority, 72% of the patients have loud snoring.

On average, these patients have 6 hours of sleep per night, with a minimum of 2h and 15minutes and a maximum of 10hours and a half. The S3 lasts one hour, and the S1 lasts for 57 minutes only. The paradoxal sleep (PS) is in general the phase that lasts the least for every most subjects. On average, it lasts 12 minutes which is almost equal to the median. Only the TTS and S2 variables are normally distributed since their p-values from the normality test are respectively (K-S (TTS) = 0,06> 0,05 and Sh-W (S2) = 0,2151 > 0,05 ; K-S (S2) = 0,3676 > 0,05).

4 QUALITY OF SLEEP MODELING

We chose to explain the quality of sleep by the total time of sleep (TTS) per night. As exogenous, based on the linear correlation criteria between the endogenous and the rest of the variables, we have selected the following variables: the latency of falling asleep (Lat) and the sleep cycle variables (S1, S2, S3 and PS)

4.1 MODEL SPECIFICATION

We examined the linear correlation coefficients between the exogenous two by two and we decided to build the following model:

$$TTS = \beta_0 + \beta_1 Lat + \beta_2 S2 + \beta_3 S3 + \beta_4 PS + \epsilon \qquad (1)$$

Le résultat de l'estimation du modèle (1) est le suivant :
Estimating the model (1) gives the following results for the different variables:

Table 1 : Model (1) Estimations

$R^2 = 0,8742$	Estimate	Std. Error	t-value	Pr(> |t|)	
(Intercept)	116.8673	4.7863	24.42	0.0000	***
Lat	-0.3045	0.0476	-6.39	0.0000	***
S2	0.8976	0.0198	45.28	0.0000	***
S3	0.5317	0.0281	18.95	0.0000	***
SP	0.9040	0.0427	21.16	0.0000	***

We can write the model as follows:

$$TTS = 116,87 - 0,3Lat + 0,89S2 + 0,53S3 + 0,9PS \qquad (2)$$

To validate the estimates, we conducted autocorrelation and heteroscedasticity tests. The obtained results are the following:

- For the autocorrelation test (DW = 2.007701 and p-value = 0.4404): we can conclude that there is no autocorrelation in the errors.
- For the heteroscedasticity test ($LM = nR^2 = 638.0859375$): we can admit the homoscedasticity of the errors.

As we can notice, the model is a very good linear model with which we can explain 87.42% of the total sleeping time per night with four explicative variables.
Each coefficient (including the constant) is significant event at 1% level. Given the nature of our variables (the majority being binary), we have explained the quality of sleep (measured by TTS) by the sleeping

cycle variable only. With suc a model, we cannot see the influence of other variables (useful in practice) on the quality of sleep. WE have decided to follow the next procedure for all the qualitative variables:

- For every modality, we will estimate the linear model (1).
- Then, we will examine the coefficients' stability :
  - If the coefficients are stable from one modality to another, we can conclude that the variable has no effect on the model.
  - Otherwise, the variable is influential on the model and we have to introduce it into the final model.

We first highlight that every coefficient must always be significant in each estimation (23 estimations in total). In the light of these results, we want integrate the most influential variables into the previous linear regression. The final model would be written in the following form:

$$TTS = \beta_0 + \beta_1 Lat + \beta_2 S2 + \beta_3 S3 + \beta_4 PS + \beta_5 QSnor + \beta_6 CVA + \beta_7 Dys + \beta_8 Ceph + \epsilon \qquad (3)$$

The results of the estimation of the model (3) is the following:

Table 2. Results of Model (3) estimations

R^2 = 0,875	**Estimate**	**Std. Error**	**t-value**	**Pr(> |t|)**	
(Intercept)	110.2650	5.8451	18.86	0.0000	***
Lat	-0.3073	0.0477	-6.44	0.0000	***
S2	0.8949	0.0201	44.63	0.0000	***
S3	0.5378	0.0283	19.01	0.0000	***
PS	0.9153	0.0433	21.12	0.0000	***
QSnor	2.3791	1.1915	2.00	0.0462	*
CVA	2.1156	6.6966	0.32	0.7522	
Dys	-0.5426	3.3307	-0.16	0.8706	
Ceph	-0.5647	2.6844	-0.21	0.8335	

We notice the three last variables (of binary type) are not significant in our estimated model. To avoid other potential problems, we tried to integrate them to the model one by one. The results remain the same. None of these variables is significant.

4.2 Estimation and discussion of the results.

We can conclude that, although the variables (CVA, Dys and Ceph) can alter minimally the coefficients of the model, they are not significant enough to be explicative. Therefore, we decide to eliminate them and only integrate QSnor into our model. Thereafter, the final model can be written as follows:

$$TTS = \beta_0 + \beta_1 Lat + \beta_2 S2 + \beta_3 S3 + \beta_4 PS + \beta_5 QSnor + \epsilon \qquad (4)$$

The results of model (4) estimation are the following :

Table 3. Results of model (4) estimation

R^2 = 0,8749	**Estimate**	**Std. Error**	**t-value**	**Pr(> |t|)**	
(Intercept)	110. 2823	5. 7906	19.05	0.0000	***

Lat	-0.3073	0.0476	-6.46	0.0000	***
S2	0.8941	0.0199	45.02	0.0000	***
S3	0.5378	0.0282	19.10	0.0000	***
PS	0.9154	0.0430	21.29	0.0000	***
QSnor	2.3877	1.1871	2.01	0.0447	*

Finally, we can write the model in the following way :

$$TTS = 10{,}2823 - 0{,}3073Lat + 0{,}8941S2 + 0{,}5378S3 + 0{,}9154PS + 2{,}3877QSnor \quad (5)$$

We notice that all the coefficients are significant (without exception). The TTS depends negatively on the Lat variable. This is completely normal since the more time we spend in bed trying to fall asleep, the less we actually sleep. The TTS depends positively on the rest of the variables. First, we analyse the link between the endogenous with the tree variables of sleeping structure (S2, S3 and PS). We notice that the highest coefficient is the paradox sleep coefficient. This means that that's the paradoxal sleep is the most important phase in determining sleep quality. If we wish to improve sleep quality, we should increase PS. Second, we notice that the QSnor coefficient is positive and the largest coefficient. This means that the more QSnor increases (by the variable construction, the snoring decreases), the better we sleep, which is logical. What is important to notice is the fact that QSnor has the largest coefficient. This translate to the QSnor variable being the most important variable in the estimation.

5 Conclusion

In this work, we sought to explain the quality of sleep. Considering the variables available in our data base, we have selected the TTS as being the endogenous. Then, to determine the explicative variables, we have performed a series of linear correlations tests between the variables. The results were unexpected, since based on this model we only able to integrate of the sleep structure (Lat, S1, S2, S3, and PS) to the model.

Since non to the qualitative variables is significant in the model, we have decided to redo the estimation on each of the terms of these variables. A qualitative variable would be considered important in the model if, the coefficients of the estimate would change from one term to the other. Only the QSnor variable was worth including in the model.

When examining the coefficients' significance and the signs, we notice that all the explicative variables in the model are significant and with the right sign. What catches our attention the most is that the quality of snoring is the most important variable in the estimation of sleep quality. This variable has the highest coefficient. If we want to improve sleep quality, one needs to control snoring.

References

[1] J. P. Baguet, J. L. Pépin, L. Hammer, P. Lévy and J. M. Mallion, "Conséquences cardiovasculaires du syndrome d'apnées obstructives du sommeil", La Revue de Médecine Interne, vol. 24 no. 8, pp 530–537, 2003.
[2] R. Bourbonnais and M. Terraza, "Analyse des séries temporelles, 2e édition", Dunod, Paris, 2008.

[3] J. F. Brun, C. Fedou, M. Desplan, A. Avignon and J. Mercier, "Évaluation d'un dépistage du syndrome d'apnées et hypopnées du sommeil chez des patients diabétiques et/ou obèses par RU sleeping", Médecine du Sommeil, vol. 10 no. 2, pp 62–63, 2013.
[4] M. Chavance and R. Manfredi, "Modélisation d'observations incomplètes", Revue Épidémie et Santé, vol. 48 no. 4, pp 389–400, 2000.
[5] J. M. Dufour, A. Farhat, L. Gardiol and L. Khalaf, "Simulation-based finite sample normality tests in linear regressions", Econometrics Journal, Royal Economic Society, vol. 1(Conference), pp C154–C173, 1998.
[6] N. J. Horton and S. R. Stuart, "Multiple imputation in practice: comparison of software packages for regression model swith missing variables", The American Statistician, vol. 55 no. 3, pp 244–254, 2001.
[7] k. Kinugawa, V. H. Nguyen-Michel andJ. Mariani, "Le syndrome d'apnées obstructives du sommeil : une cause de troubles cognitifs chez les personnes âgées ?", La Revue de Médecine Interne, vol. 35 no. 10, pp 664–669, 2014.
[8] A. Papoulis and S. U. Pillai, "Probability, Random Variables and Stochastic Processes, Fourth Edition", McGraw-Hill, New York, 2002.
[9] J. L. Schafer and M. K. Olsen, "Multiple imputations for multivariate missing data problems: A data analyst's perspective", Multivariate Behavioral Research, vol. 33 no. 4, pp 545–571, 2000.
[10] V. Viot-Blanc, "Durée de sommeil et métabolisme", Revue des Maladies Respiratoires, vol. 32 no. 10, pp 1047–1058, 2015.