Mugla Journal of Science and Technology

A STUDY ON COVID-19 DATA WITH LOG-LINEAR MODEL APPROACH

Gökçen Altun*, Department of Computer Technology and Information *Systems*,Bartin University, Bartin, Turkey,
gokcenefendioglu@gmail.com
https://orcid.org/0000-0003-4311-6508

Received: 03.12.2020, Accepted: 22.01.2021
*Corresponding author
Research Article
DOI: 10.22531/muglajsci.835562

Abstract

The researchers have done several studies to explore the important variables on the novel coronavirus (COVID-19) disease. However, there is no direct research to examine the interaction between the variables affecting the COVID-19 disease. In this study, the log-linear models are used to explore the significant interactions between the country, gender and age variables. Empirical findings of the presented study show that the main effects are found statistically significant for country, gender and age variables. The two-way interactions between the gender and country and between the interaction of the country and age are found statistically significant. The risk of the China for catching the COVID-19 disease is 1.7 times higher than the risk of the South Korea for catching the COVID-19 disease. The risk of male individuals is 1.2 times higher than the risk of female individuals for catching the COVID-19 disease. Additionally, the individuals having the 41-60 age group has higher risk than the individuals having 70 and above age group. We believe that the empirical results of the presented study will be helpful to the policymakers for decreasing the spread of the COVID-19 disease.

Keywords: Categorical data analysis, Log-linear models, COVID-19, Outbreak.

LOG-LINEER MODEL YAKLAŞIMI ILE COVID-19 VERISI ÜZERINE BİR ÇALIŞMA

Özet

Araştırmacılar, yeni koronavirüs (COVID-19) hastalığı üzerindeki önemli değişkenleri araştırmak için çeşitli çalışmalar yaptılar. Fakat, COVID-19 hastalığını etkileyen değişkenler arasındaki etkileşimi incelemek için doğrudan bir araştırma yoktur. Bu çalışmada, ülke, cinsiyet ve yaş değişkenleri arasındaki önemli etkileşimleri incelemek için log-lineer modeller kullanılmıştır. Çalışmanın sonucunda, ülke, cinsiyet ve yaş ana etkenlerinin istatistiksel olarak anlamlı bulunduğunu göstermektedir. Aynı zamanda cinsiyet ve ülke ve ülke ile yaş arasındaki iki yönlü etkileşimler istatistiksel olarak anlamlı bulunmuştur. Parametre tahmin değerlerinden elde edilen sonuçlara göre, Çin'de COVID-19 hastalığına yakalanma riski, Güney Kore'de COVID-19 hastalığına yakalanma riskinden 1,7 kat daha yüksek ve erkek bireylerin COVID-19 hastalığına yakalanma riski, kadın bireylerin hastalığa yakalanma riskinden 1,2 kat daha yüksektir. Ayrıca 41-60 yaş grubundaki bireyler 70 ve üzeri yaş grubuna göre daha yüksek risk taşımaktadır sonucuna ulaşılmıştır. Çalışmanın sonuçlarının, COVID-19 hastalığının yayılmasını azaltmak için yardımcı olacağına inanıyoruz.

Anahtar Kelimeler: Kategorik veri analizi, Log-doğrusal modeler, COVID-19, Salgın

Cite

ALTUN, G., (2021). "A Study on COVID-19 Data with Log-Linear Model Approach", Mugla Journal of Science and Technology, 7(1), 52-58.

1. Introduction

Coronaviruses have an important place in the history of the 21st century. COVID-19 caused by SARS-CoV-2 is a highly contagious disease. The World Health Organization (WHO) has announced that the ongoing outbreak is a global public health situation. Since the outbreak of severe acute respiratory syndrome (SARS), taken place before 18 years ago, a large number of SARS-related coronaviruses (SARSr-CoVs) have been discovered in their natural reservoir host, bats [1-4]. Coronaviruses have caused two large-scale pandemics in the past two decades, SARS and Middle East respiratory syndrome (MERS). It has generally been thought that SARSr-CoV—which is mainly found in bats—could cause a future disease outbreak. [5-7]. Typical clinical symptoms of these patients are fever, dry cough, breathing difficulties (dyspnoea), headache and pneumonia. Disease onset may result in progressive respiratory failure owing to alveolar damage (as observed by transverse chest computerized-tomography images) and even death [8].

Ai et al. [9] used statistical analysis methods in their study. They expressed continuous variables as mean ± standard deviation and categorical variables as count and percentage. Comparison was made with the chi-square test to identify COVID-19 by different age groups (<60 years and ≥ 60 years) and gender. They compared Chest CT and RT-PCR Test and they stated that chest CT has high sensitivity for the diagnosis of COVID-19. Li et al. [10] assessed the psychological stress and traumatization caused by the COVID-19 outbreak, 214 general publics and 526 nurses (234 frontline nurses and 292 frontline nurses) were used to assess traumatization scores through a mobile-designed questionnaire. They used chi-square test and Fisher exact test to evaluate group differences. There was no significant difference between groups in comparing general characteristics (such as gender, age, marital status). However, in terms of traumatization scores, the difference between the general public, the nurses working in the front line and the nurses who did not take the task in the front line were found statistically significant. Zhao et al. [11] conducted a study to investigate the relationship between chest CT findings and the clinical conditions of COVID-19 pneumonia. Basic clinical features and detailed imaging features were evaluated and compared between mild and severe disease groups. Emergency (severe disease) and non-emergency (mild disease) groups compared with chi-square or Fisher exact test, and p <0.05 was considered statistically significant. Linton et al. [12] made predictions of incubation time and incubation time for COVID-19 using the statistical analysis. In making the analyzes, they used the history of three critical points during the infection, such as disease onset, hospitalization, and death to calculate four time periods. They made inferences based on gender and age variables. Li et al. [13] showed that disease distribution was higher in males using meta-analysis and the mortality rate of COVID-19 patients was 5%. Lippi and Henry [14] investigated the possible relation between the smoking and severity of the COVID-19. For this purpose, the studies on this subject have been scanned. Only one study explained that smoking was an important predictor of COVID-19 severity, while in other studies the relationship between active smoking and COVID-19 was not statistically significant. Liu et. al [15] found that elderly patients with COVID-19 are more likely to progress to severe disease. Likassa [16] also showed that age variation has a statistical significant associated with COVID-19. Cesar et al. [17] classified the COVID 19 cases and the number of deaths by age and gender in Peru. At the end of the study, they concluded that men were more affected by COVID-19. In addition, it has been concluded that individuals over the age of 80 have a higher mortality rate. Pieh et al. [18] conducted a study involving 1005 people to investigate the effect of COVID-19 on mental health in Australia. Statistical analyzes such as ANOVA, t test and post-hoc tests were applied to the data obtained from the survey. As a result of the analysis, they showed that COVID-19 increased depression and decreased quality of life compared to pre-COVID-19.

As seen from the above literature review, COVID-19 strongly depends on the variables such as age, gender, presence of chronic disease, country of residence. The aim of this study is to make new empirical inferences from the literature using current COVID 19 data. In this study, 652 COVID-19 test patients were evaluated according to age, gender and countries they live in. The three-way cross-classified COVID 19 data has been analyzed based on log-linear models. The most suitable model was found and the results was interpreted based on the estimated parameters. The data source is https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge.

The remaining parts of the presented study are organized as follows. Section 2 deals with the used statistical model and its theoretical properties. Section 3 is devoted to the empirical results of the study. Section 4 contains the concluding remarks.

2. Methods

2.1. Log-linear Models

The concept of log-linear analysis in contingency tables is analogous to the concept of analysis of variance (ANOVA) for the continuously distributed factor-response variables. While response observations are assumed to be continuous with underlying normal distributions in ANOVA, the log-linear analysis assumes that the response observations are counts having Poisson distributions [19].

In the case of more than two categorical variables, the use of chi-square independence tests in the determination of the relationship between the variables in the contingency tables becomes difficult or sometimes impossible. In this case, log-linear models, which allow the testing of a much larger number of hypotheses compared to the chi-square, which do not impose restrictions on the number of rows and columns in both the two-dimensional tables where chi-square can be applied, and the three-dimensional tables where chi-square is insufficient, is preferred. In the multi-dimensional contingency tables in log-linear models, a model is created to investigate the relationships between the variables, the parameters in the model are estimated and the significance of this model is tested. The overall goodness-of-fit of a model is assessed by comparing the expected frequencies to the observed cell frequencies for each model. The Pearson chi-square statistics or the likelihood ratio statistic (G^2) can be used to test a model fit. G^2 is more commonly used because it is the statistic that minimized in maximum likelihood estimation [20].

$$Pearson's\chi^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{K}\frac{\left(n_{ijk}-E_{ijk}\right)^2}{E_{ijk}}$$

$$Loglikelihood\ G^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{K} n_{ijk}\log\left(\frac{n_{ijk}}{E_{ijk}}\right)$$

9 different log-linear models can be created for three dimensional contingency tables. It is possible to collect these models in five groups. These; log-linear models with complete independence, log-linear models with partial independence, log-linear models with conditional independence, log- linear models containing all two-way interactions and log-linear models containing all interactions. These models are called progressive (hierarchical) models.

Table 1: Hierarchical model representation and equations.

Model	Model Equation
0 (X,Y,Z)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$
1(X,YZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$
2(Y,XZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$
3(Z,XY)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$
4(XZ,YZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
5(XY,YZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$
6(XY,XZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$
7(XY,XZ,YZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
8(XYZ)	$logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$

X, Y and Z show the variables in a three-dimensional contingency table, R, C, and K represent the level numbers of these variables. The explanations of the terms in the models are given below.

E_{ij} is the expected frequency for cell, (i=1,...,R, j=1,...,C, and k=1,...K), which is calculated over the model. λ reflects the constant term. The parameter λ_i^X shows the effect of ith level of factor X. The parameter λ_j^Y shows the effect of jth level of factor Y . The parameter λ_k^Z shows the effect of kth level of factor Z where i,j,k=1,...R. $\lambda_{ij}^{XY}, \lambda_{ik}^{XZ}, \lambda_{jk}^{YZ}$ and λ_{ijk}^{XYZ} indicate the interaction parameter. For the log-linear models in Table 1, the following constraints must be met.

$$\sum_{i=1}^{R} \lambda_i^X = \sum_{j=1}^{C} \lambda_j^Y = \sum_{k=1}^{K} \lambda_k^Z = \sum_{i=1}^{R} \lambda_{ij}^{XY} = \sum_{j=1}^{C} \lambda_{ij}^{XY} = \cdots$$
$$= \sum_{k=1}^{K} \lambda_{ijk}^{XYZ} = 0$$

$\lambda_{ij}^{XY}, \lambda_{ik}^{XZ}, \lambda_{jk}^{YZ}$ and λ_{ijk}^{XYZ} show two-way and three-way interactions, respectively.

2.2. Statistical Analysis

The study included 652 COVID-19 infected individuals. The data source is https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset. the numerical results of the study are obtained by using the SPSS (Statistical Package for the Social Sciences) . Three-dimensional cross-table was obtained according to the gender (female=1, male=2), countries (China=1, Hong Kong=2, Japan=3, Singapore=4, South Korea=5) and age (0-20=1, 21-40=2, 41-60=3, 61-70=4, 70+=5) of the patients. The best model was decided by backward elimination methods. After parameter estimates of the best model, the variables that are significant are interpreted ($p<0.05$).

3. Results

3.1. Comparison of patients according to gender, country and age

Table 2: Number of patients infected positive COVID-19 test according to gender, country and age.

Gender	Country	Age					
		0-20	21-40	41-60	61-70	70+	Total
Female	China	7 (2.6)	23 (8.6)	21 (7.8)	15 (5.6)	6 (2.2)	72 (26.9)
	Hong Kong	0 (0)	8 (3.0)	16 (6.0)	12 (4.5)	10 (3.7)	46 (17.2)
	Japan	1 (0.4)	13 (4.9)	25 (9.3)	11 (4.1)	14 (5.2)	64 (23.9)
	Singapore	1 (0.4)	13 (4.9)	18 (6.7)	1 (0.4)	4 (1.5)	37 (13.8)
	South Korea	2 (0.7)	17 (6.3)	19 (7.1)	9 (3.4)	2 (0.7)	49 (18.3)
	Total	11 (4.1)	74 (27.6)	99 (36.9)	48 (17.9)	36 (13.4)	268 (100)
Male	China	3 (0.8)	38 (9.9)	44 (11.5)	16 (4.2)	17 (4.4)	118 (30.7)
	Hong Kong	1 (0.3)	12 (3.1)	16 (4.2)	11 (2.9)	7 (1.8)	47 (12.2)
	Japan	4 (1.0)	16 (4.2)	46 (12.0)	27 (7.0)	28 (7.3)	121 (31.5)
	Singapore	2 (0.5)	26 (6.8)	20 (5.2)	4 (1.0)	1 (0.3)	53 (13.8)
	South Korea	2 (0.5)	16 (4.2)	17 (4.4)	4 (1.0)	6 (0.3)	45 (11.7)
	Total	12 (3.1)	108 (28.1)	143 (37.2)	62 (16.1)	59 (15.4)	384 (100)

* The values in the parenthesis are the percentages.

When Table 2 is examined in terms of gender, it can be concluded that the disease is more common in male individuals. The number of diseases as a country is higher for both sexes in China and Japan. When the ages of COVID-19 infected individuals are examined, it is seen that individuals between the ages of 41-60 are more at risk of disease. It may be thought that the reason for this is that individuals in this age range are actively working individuals and they have to go out in this process. Studies have shown that older individuals are more affected by the disease. However, since these individuals live more isolated in the process, the number of infected individuals in the 70+ age group is less than the other groups, except for 0-20 years old individuals. It can be concluded that the age range that caught the least affected was 0-20 age group individuals. The reason for this can be predicted as the fact that these individuals can live isolated and have less chronic ailments compared to older age groups.

3.2. Determination of the best model with backward elimination method

The log-linear models are applied to data used and the main effects and interaction terms are hierarchically tested to decide which terms will be included in the model. Therefore, in the three-way table, the significance of the main effect terms, two-way interaction and three-way interactions are examined.

Table 3: Degrees of freedom (df), p-value and test statistic values for K-Way and higher-order effects.

	K	df	G^2	p-value	χ^2	p-value
K-way and Higher Order Effects	1	49	440.119	<0.001	445.393	<0.001
	2	40	94.864	<0.001	90.364	<0.001
	3	16	21.901	0.146	21.149	0.173
K-way Effects	1	9	345.255	<0.001	355.028	<0.001
	2	24	72.963	<0.001	69.215	<0.001
	3	16	21.901	0.146	21.149	0.173

H_{01}: One-way and higher effects are not significant.
$p = 0.000 < 0.05$ H_0 is rejected; one-way and higher interactions are significant.
H_{02}: Two-way and three-way interactions are not significant.
$p = 0.000 < 0.05$ H_0 is rejected; Two-way and three-way interactions are significant.
H_{03}: Three-way interaction is not significant.
$p = 0.173 > 0.05$ H_0 is not rejected; Three-way interaction is not significant.
H_{04}: One-way effect is not significant.
$p = 0.000 < 0.05$ H_0 is rejected; one-way effect is significant.
H_{05}: Two-way interactions are not significant.
$p = 0.000 < 0.05$ H_0 is rejected; Two-way interactions are significant.
H_{06}: Three-way interaction is not significant.
$p = 0.173 > 0.05$ H_0 is not rejected; Three-way interaction is not significant.
Table 4 was obtained to analyze whether two-way interactions and main effects were significant.

Table 4: The degree of freedom, p value and test statistic values of two-way interaction and main effects terms.

Effect	df	χ^2	p-value
gender*country	4	11.361	0.023
gender*age	4	1.109	0.893
country *age	16	60.432	0.000
gender	1	20.748	0.000
country	4	81.298	0.000
age	4	243.208	0.000

It can be concluded from Table 4 that gender and age interaction is not statistically significant (p>0.05). Country and age interactions with gender and country are statistically significant (p<0.05). In addition, three of the main effects were found significant (p<0.05). The parameter estimates of the main effects are as given in Table 5.

From Table 5, we conclude that the risk of males being infected by the COVID-19 is 1.2 times higher than female individuals. The risk of catching COVID-19 in China is 1.7 ($e^{0.531} = 1.7$) times higher than in South Korea, and the risk of catching COVID-19 infected in Japan is 1.5 times higher than in South Korea. In addition, in South Korea, the risk of catching COVID-19 is 1.6 times higher in Singapore. When the table is interpreted in terms of age variable, the individuals over the age of 70 are 4 times more likely to be infected with COVID-19 than those under the age of 20. In addition, individuals aged 21-40 are approximately 2 times more likely than individuals over the age of 70, and individuals aged 41-60 are 2.5 times more likely to be infected disease than individuals over 70. The best model selection with backward elimination method shown in Table 6.

In Table 6, the generating class expression tests the compatibility with the model, while the deleted effect expression tests the significance of the interactions. Steps continue until all of the interactions examined are statistically significant. In this case, gender-age and country-age interactions are statistically significant and should be included in the model. So the best model for these three variables is the model represented by XY, YZ. In this case, the model equation is $logE_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$. Although it is expressed by the variables itself, it is in the form of $logE_{ijk} = constant + gender_i + country_j + age_k + gender_i \times country_j + country_j \times age_k$. The test statistic value of the best model is given in Table 7 which shows that the model is statistically significant.

Table 7: Test statistics and p-value of XY,YZ model.

	Tets statistics	df	p-value
G^2	23.009	20	0.288
χ^2	22.417	20	0.318

The parameter estimates of the best fitting model are given in Table 8.

The contribution of interaction terms with p> 0.05 to the model is not statistically significant. If the interaction parameters are interpreted according to the results in Table 8, it is possible to say that the gender variable has a significant effect in Japan and the risk of male catching COVID-19 is higher than female. The common age group that makes sense in every country is individuals between the ages of 41-60. Compared to individuals over the age of 70, the risk of COVID-19 infection among individuals aged 41-60 is 2.8 times in China, 1.9 times in Hong-Kong, 1.7 times in Japan, 7.6 times in Singapore and 4.5 times in South Korea.

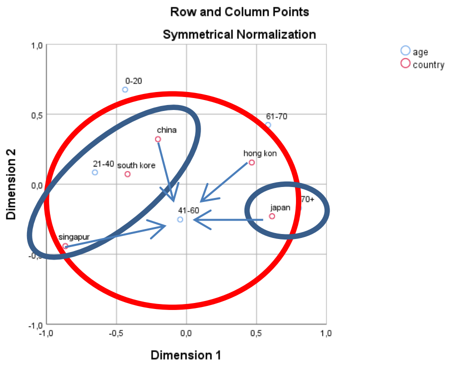

Figure 1: Bi-plot graphics of the correspondence analysis for the age and country

For age and country variables whose interaction was significant, the correspondence analysis was implemented and the results are graphically displayed in Figure 1 which shows that the age group 41-60 is approximately the same distance from all countries and is important. 21-40 years old COVID-19 infected patients are observed to be close to China, Singapore and South Korea. The results of the log-linear model are consistent with the graphic obtained.

4. Conclusion

There are several advantages of utilizing efficiently the special models which use the ordinal information. Such models are more parsimonious and have simpler interpretations than the standard models for nominal variables. If at least one of the variables in a two or higher dimensional contingency table, we'd rather some order restricted models. In this study, 652 COVID-19 infected patients were grouped according to their country, age and gender and analyzed using log-linear models. According to the results of the analysis, age and country and gender and country interactions were significant. Studies and explanations about the coronavirus show that individuals over 60 years old and patients with chronic conditions are more at risk. However, the risk of transmission of the virus and the number of infected patients were shown in the study conducted between 21-60 years of age. It is thought that the reason for this situation is due to the fact that individuals aged 21-60 are active working age groups. The fact that death in this age group is less than individuals over the age of 70 and patients with chronic diseases prevents the consequences of pandemic from being much more terrible.

It is important to collect data regularly and continuously against the epidemic. Studies similar to this study, which will be carried out regularly with the data obtained, are very important to provide a scientific understanding of the preventive measures that can be used to control and reduce the spread of infections and epidemics. Based on these and similar studies, the groups heavily affected by the virus can be easily identified. Thus, the measures to be taken can be customized according to the age, gender and profession of the persons.

Table 5: Parameter estimates of the main effects

Effects	Parameter	Estimate	Standard error	Z-value	p-value	95% Confidence Interval	
						Lower bound	Upper bound
gender	1	-0.147	0.065	-2.264	0.024	-0.274	-0.020
country	1	0.531	0.098	5.424	<0.001	0.339	0.723
	2	-0.259	0.157	-1.654	0.098	-0.566	0.048
	3	0.427	0.110	3.879	<0.001	0.211	0.642
	4	-0.495	0.150	-3.294	0.001	-0.789	-0.200
age	1	-1.386	0.198	-6.997	<0.001	-1.775	-0.998
	2	0.652	0.089	7.284	<0.001	0.476	0.827
	3	0.941	0.084	11.253	<0.001	0.777	1.105
	4	0.002	0.118	0.021	0.983	-0.230	0.235

Table 6: Determination of the best model with backward elimination method

Step		Effects	χ^2	df	p-value
0	Generating class	gender* country *age	0.000	0	.
	Deleted effect	gender* country *age	21.901	16	0.146
1	Generating class	gender*country, gender*age, country*age	21.901	16	0.146
	Deleted effect	gender*country	11.361	4	0.023
		gender*age	1.109	4	0.893
		country*age	60.432	16	0.000
2	Generating class	gender*country, country*age	23.009	20	0.288
	Deleted effect	gender*country	11.392	4	0.022
		country*age	60.463	16	0.000
3	Generating class	gender*country, country*age	23.009	20	0.288

Table 8: Parameter estimates for XY,YZ model

Parameter	Estimate	Standard error	z-value	p-value	95% Confidence Interval	
					Lower bound	Upper bound
Constant	1.343	0.370	3.633	<0.001	0.618	2.067
[gender = 1] * [country = 1]	0.822	0.434	1.893	0.058	-0.029	1.674
[gender = 1] * [country = 2]	0.786	0.454	1.731	0.083	-0.104	1.677
[gender = 1] * [country = 3]	1.333	0.413	3.228	0.001	0.524	2.143
[gender = 1] * [country = 4]	-0.622	0.594	-1.048	0.295	-1.786	.0541
[gender = 1] * [country = 5]	0.085	0.206	0.412	0.680	-0.320	.0490
[gender = 2] * [country = 1]	1.316	0.428	3.075	0.002	0.477	2.155
[gender = 2] * [country = 2]	0.808	0.454	1.780	0.075	-0.081	1.697
[gender = 2] * [country = 3]	1.970	0.404	4.876	<0.001	1.178	2.762
[gender = 2] * [country = 4]	-0.263	0.587	-0.448	0.654	-1.413	0.887
[country = 1] * [age = 1]	-0.833	0.379	-2.199	0.028	-1.575	-0.091
[country = 1] * [age = 2]	0.975	0.245	3.986	<0.001	0.496	1.455
[country = 1] * [age = 3]	1.039	0.243	4.282	<0.001	0.563	1.514
[country = 1] * [age = 4]	0.298	0.275	1.085	0.278	-0.241	0.838
[country = 2] * [age = 1]	-2.833	1.029	-2.754	0.006	-4.850	-0.817
[country = 2] * [age = 2]	0.163	0.330	0.493	0.622	-0.484	0.809
[country = 2] * [age = 3]	0.633	0.300	2.108	0.035	0.044	1.221
[country = 2] * [age = 4]	0.302	0.320	0.945	0.345	-0.325	0.929
[country = 3] * [age = 1]	-2.128	0.473	-4.499	<0.001	-3.055	-1.201
[country = 3] * [age = 2]	-0.370	0.241	-1.534	0.125	-0.844	0.103
[country = 3] * [age = 3]	0.525	0.195	2.697	0.007	0.143	0.907
[country = 3] * [age = 4]	-0.100	0.224	-0.447	0.655	-0.539	0.339
[country = 4] * [age = 1]	-0.511	0.730	-0.699	0.484	-1.942	0.921
[country = 4] * [age = 2]	2.054	0.475	4.324	<0.001	1.123	2.985

[country = 4] * [age = 3]	2.028	0.476	4.263	<0.001	1.096	2.961
[country = 4] * [age = 4]	4.7x10^{-17}	0.632	<0.001	1.000	-1.240	1.240
[country = 5] * [age = 1]	-0.693	0.612	-1.132	0.258	-1.893	0.507
[country = 5] * [age = 2]	1.417	0.394	3.596	<0.001	0.645	2.189
[country = 5] * [age = 3]	1.504	0.391	3.848	<0.001	0.738	2.270
[country = 5] * [age = 4]	0.486	0.449	1.080	0.280	-0.395	1.366

5. References

[1] Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., & Zhang, J., "Bats are natural reservoirs of SARS-like coronaviruse", *Science*, 310(5748), 676-679, 2005.

[2] Ge, X. Y., Li, J. L., Yang, X. L., Chmura, A. A., Zhu, G., Epstein, J. H., ... & Zhang, Y. J., "Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor", *Nature,* 503(7477), 535-538, 2013.

[3] Yang, L., Wu, Z., Ren, X., Yang, F., He, G., Zhang, J.& Zhang, S., "Novel SARS-like betacoronaviruses in bats, China, 2011", *Emerging infectious diseases*, 19(6), 989,2013.

[4] Hu, B., Zeng, L. P., Yang, X. L., Ge, X. Y., Zhang, W., Li, B., ... & Luo, D. S., "Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus", *PLoS pathogens*, 13(11), e1006698, 2017.

[5] Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H. R., Becker, S., ... & Berger, A., "Identification of a novel coronavirus in patients with severe acute respiratory syndrome" *New England journal of medicine*, 348(20), 1967-1976, 2003

[6] Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A., "Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia", *New England Journal of Medicine*, 367(19), 1814-1820, 2012

[7] Fan, Y., Zhao, K., Shi, Z. L., & Zhou, P., "Bat coronaviruses in China", *Viruses*, 11(3), 210, 2019

[8] Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D., "A pneumonia outbreak associated with a new coronavirus of probable bat origin", *Nature*, 579(7798), 270-273,2020

[9] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... & Xia, L.,"Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases", *Radiology*, 200642, 2020.

[10] Li, Z., Ge, J., Yang, M., Feng, J., Qiao, M., Jiang, R., ... & Zhou, Q., "Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control", *Brain, behavior, and immunity.* 2020

[11] Zhao, W., Zhong, Z., Xie, X., Yu, Q., & Liu, J., "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study", *American Journal of Roentgenology*, 214(5), 1072-1077,2020.

[12] Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S. M., ... & Nishiura, H., "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data", *Journal of clinical medicine*, 9(2), 538,2020

[13] Li, L. Q., Huang, T., Wang, Y. Q., Wang, Z. P., Liang, Y., Huang, T. B., ... & Wang, Y., "COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis", *Journal of medical virology*, 92(6), 577-583,2020

[14] Lippi, G., & Henry, B. M., "Active smoking is not associated with severity of coronavirus disease 2019 (COVID-19", *European journal of internal medicine*, 2020

[15] Liu, D., Li, L., Wu, X., Zheng, D., Wang, J., Yang, L., & Zheng, C., "Pregnancy and perinatal outcomes of women with coronavirus disease (COVID-19) pneumonia: a preliminary analysis*", American journal of roentgenology*, 1-6,2020

[16] Likassa, H. T., "Predictive Models on COVID 19: What Africans Should Do?", Infectious Disease Modelling, 2020

[17] Munayco, C., Chowell, G., Tariq, A., Undurraga, E. A., & Mizumoto, K., "Risk of death by age and gender from CoVID-19 in Peru, March-May, 2020*", Aging (Albany NY)*, 12(14), 13869-13881,2020

[18] Pieh, C., Budimir, S., & Probst, T., "The effect of age, gender, income, work, and physical activity on mental health during coronavirus disease (COVID-19) lockdown in Austria", *Journal of psychosomatic research*, 136, 110186,2020

[19] Lawal, HB., Categorical Data Analysis with SAS and SPSS Application. Lawrence ERlbawn Association Publisher, London, 2003

[20] Agresti, A., *Categorical data analysis* . John Wiley & Sons, New Jersey, 2003