Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/22/20230327

Correlation Analysis of Factors Influencing Insurance Claim

Boheng Yi[1,a,*]

[1]*College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China*
a. henryyi@nuaa.edu.cn
**corresponding author*

Abstract: In this work, many information concerning body indexes as well as living habits were used in order to analyze which factor may be most corelated with insurance claims. Many mathematical methods were used in order to gain more precise conclusions. By analyzing these key factors, it can be predicted that who is more likely to receive higher insurance claim. Therefore, it may give insurance companies valuable insight and help them to make decisions while considering potential for their services. On a broader scale, it can inform public policy by allowing for more targeted support for those who are most in need and vulnerable.

Keywords: health insurance claim, BMI, gender, age, smoke, blood pressure

1. Introduction

For the insurance companies, it is critical to determine the risk of insurance claims through the condition of the insured in order to set premiums and manage capital more accurately. Among all kinds of insurance, health insurance is particularly complicated, with adverse selection and moral hazard problems more serious than other types of insurance, such as Einav, Liran, Amy Finkelstein, Stephen P. Ryan [1] and Handel, Benjamin R. [2]. At the same time, health insurance is often accompanied by higher insurance claims because there is no limit on the amount of insurance coverage. Therefore, insurance companies need to understand the health status of their customers through medical examinations and other means, so as to more accurately determine the risk of insurance claims.

However, this may not be enough. Customers may hide their bad habits and deny their past history of illness in order to enjoy lower premiums. Although there are some ways to detect insurance fraud such as Melih Kirlidog and Cuneyt Asuk [3], by exploring the correlation between physical illness and lifestyle habits and insurance benefits, insurance companies can better predict losses and more scientifically set premiums for different customers.

In this work, I conducted a data analysis of a set of insurance claims data from Kaggle [4], exploring the impact of indicators on each other and their interpretation of the amount of insurance claims. These studies allow insurers to know which factors are more correlated with insurance claims and which factors are not strongly correlated with insurance claims, and thus to better predict and manage risk.

DOI: 10.54254/2754-1169/22/20230327

2. Materials and Method

2.1. Hypothesis Testing

This data analysis uses a hypothesis testing analysis method to formulate hypotheses about the data to be analyzed and verify or disprove the hypotheses through the data. Hypothesis testing allows me to explore the intrinsic relationship between each variable and the amount of insurance claims, and then to better determine and analyze which indicators are more correlated with the amount of insurance claims.

First, I propose the hypothesis that age is positively correlated with the amount of insurance claims because older people are more likely to be in poor health.

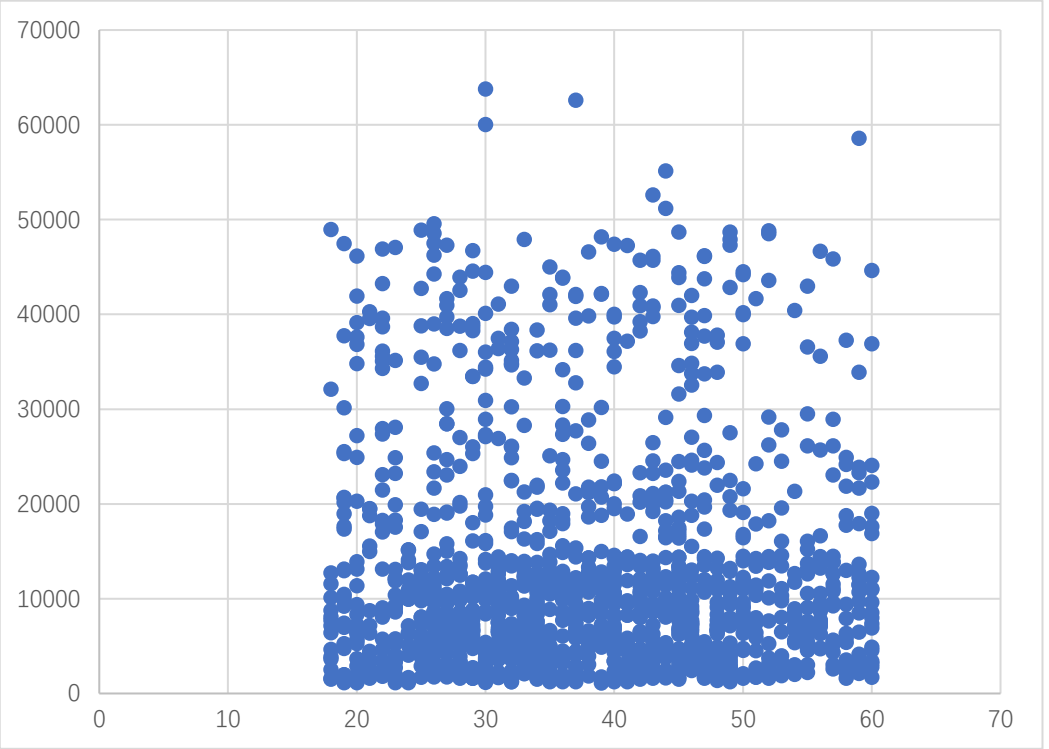

Figure 1: Age and claim distribution.

However, according to the figure 1, there is no significant relations between age and insurance claims, since the proportion of high insurance claims in the lower age group is roughly the same as in the higher age group.

Second, I propose the hypothesis that male is more likely than female to get higher insurance claims because they are more likely to have health-adverse habits such as smoking and alcohol abuse and, according to statistics, men live on average less than women.

Table 1: Gender.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	662	49.4	49.4	49.4
	male	678	50.6	50.6	100.0
	Total	1340	100.0	100.0	

DOI: 10.54254/2754-1169/22/20230327

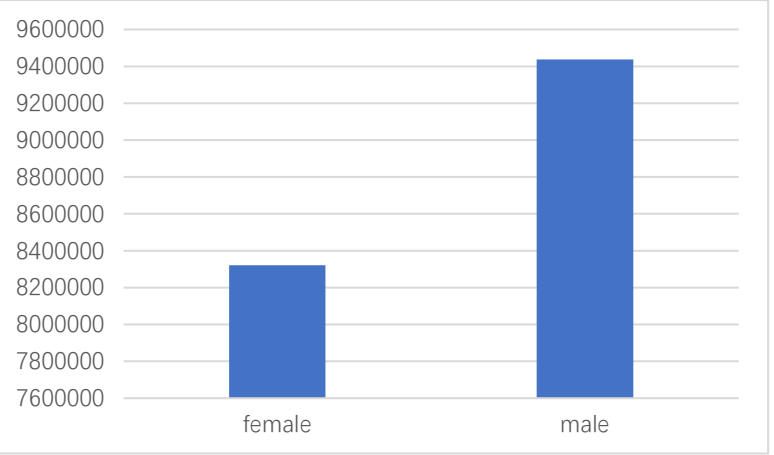

Figure 2: Total amount of claim for both sex.

The table 1 and figure 2 show that when the number of men and women is approximately equal, women receive significantly lower total insurance benefits compared to men, with around $8,300,000 for female and more than $9,400,000 for male, which strongly support my hypothesis. The result is consistent with the findings of by Thomas Heisser and Andreas Simon [5], which observed higher costs in men versus women.

Third, I propose the hypothesis that BMI is positively correlated with the amount of insurance claims because for people who are overweight, it is more likely that they get diseases such as diabetes, hypertensions, coronary heart diseases, which will lead to higher insurance claims.

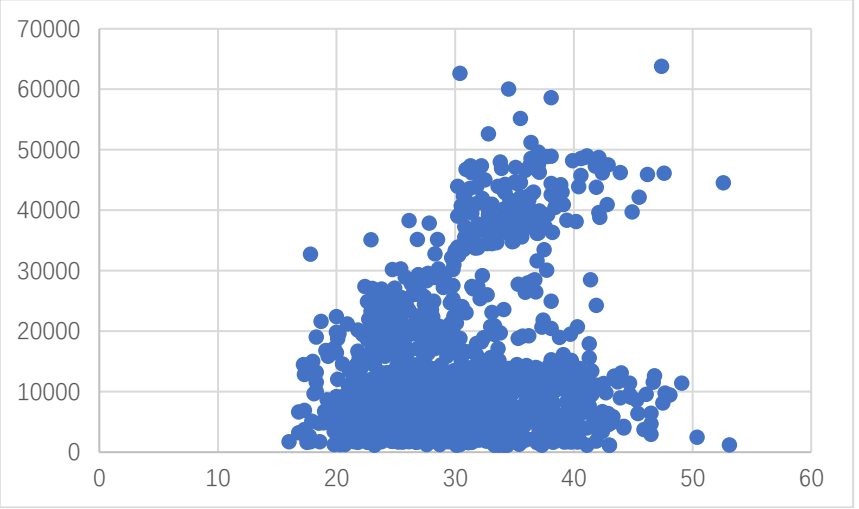

Figure 3: Scatter plot of BMI and claim distribution.

According to the figure 3, few people with a BMI less than 30 have insurance benefits greater than $40,000, while a large number of people with a BMI greater than 30 have insurance benefits greater than $40,000. Meanwhile, there is a rising trend for the maximum value of insurance claim when BMI gets larger, which all strongly support my hypothesis. According to Jay Bhattacharya and Neeraj Sood [6], when underwriting on weight is not allowed, there is an obesity externality, all plan participants face inefficient incentives to undertake unpleasant dieting and exercise.

Next, I propose the hypothesis that there is no significant difference in insurance claims between four regions, as no relevant studies proved a significant correlation between regions and claims.

DOI: 10.54254/2754-1169/22/20230327

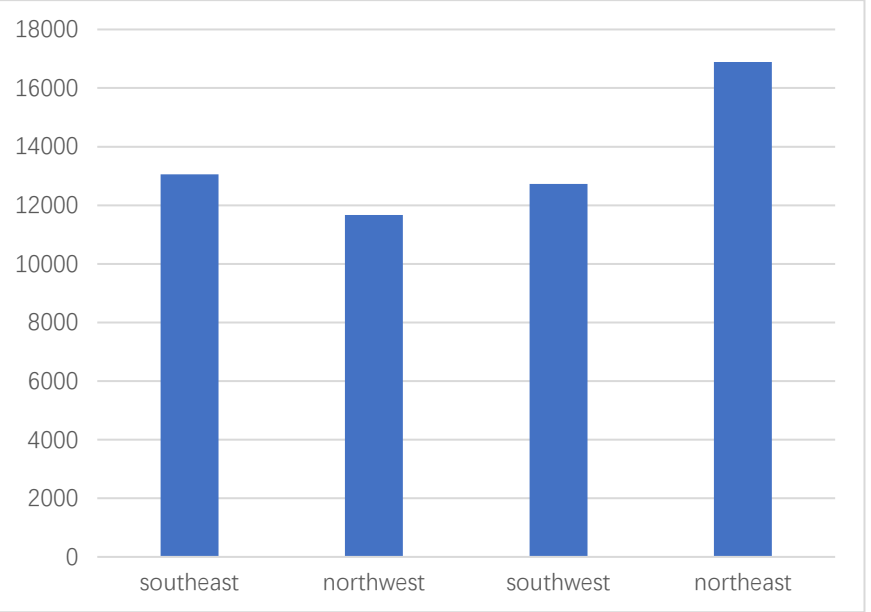

Figure 4: Claim per capita for all regions.

From the figure 4, it is clear that people from northeast have highest average insurance claim of over 16,000 dollars, while other people from other areas have the average insurance claim of around 12,000 dollars.

Then, I propose the hypothesis that smokers tend to have significantly higher insurance claims comparing with non-smokers, because cigarettes contain substances such as nicotine, which can have a negative effect on the human body, and long-term exposure to these substances can lead to lung cancer and other diseases.

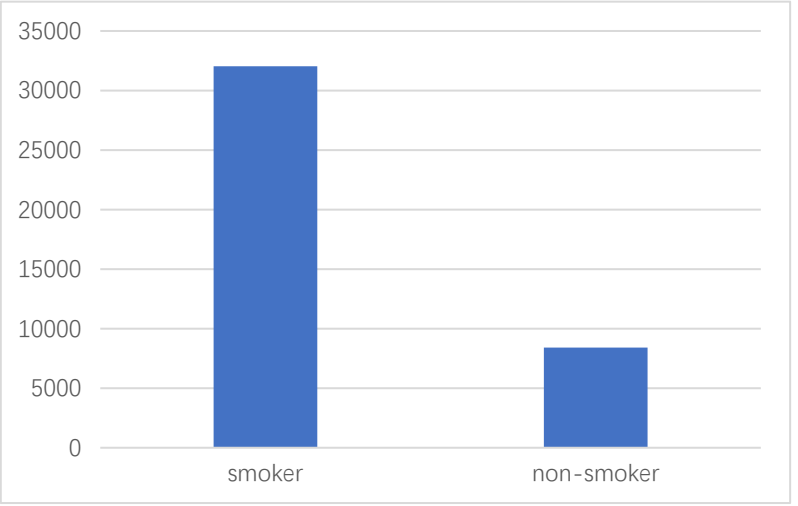

Figure 5: Claims per capita for both smokers and non-smokers.

The figure 5 shows that insurance benefits are significantly higher for smokers than for non-smokers, reaching over $30,000 for the former, more than three times as much as for the latter, which strongly support my hypothesis.

Next, I propose the hypothesis that the number of children people have are positively corelated with insurance claims because more children bring more housework and labor, and thus parents will have less time for exercise and diet.

DOI: 10.54254/2754-1169/22/20230327

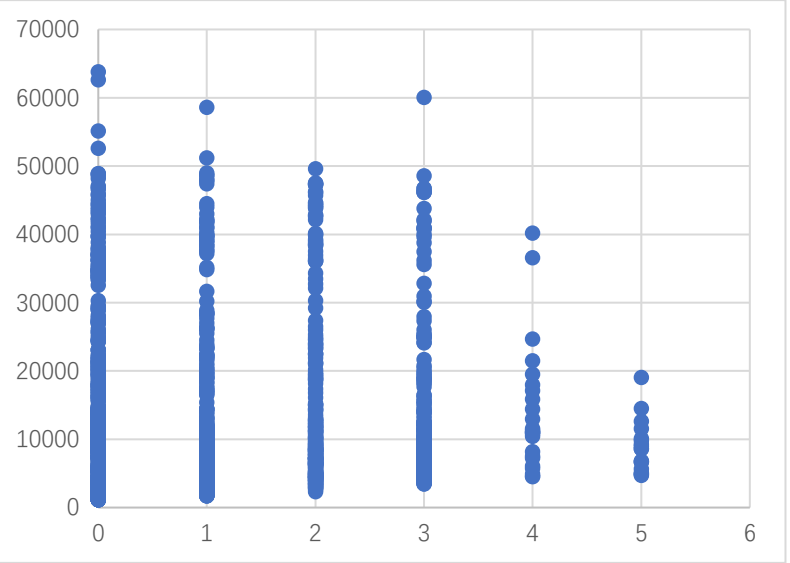

Figure 6: Children and claim distribution.

However, from the figure 6, most people with 4 or 5 kids have insurance claims of less than $30,000, while there are many people who have less than 4 kids have claims of over $30,000. Meanwhile, people who have 2 or 3 kids do not significantly have higher insurance claims than those who have less than 2 kids. This may be due to the fact that families with larger numbers of children tend to be better off financially and therefore spend more money on health monitoring and hiring people to help them with household chores.

After that, I propose the hypothesis that diabetics tend to have significantly higher insurance claims comparing with non-diabetics, because people who have diabetics are more likely to have worse living habits and diet, and diabetic will lead to other serious diseases.

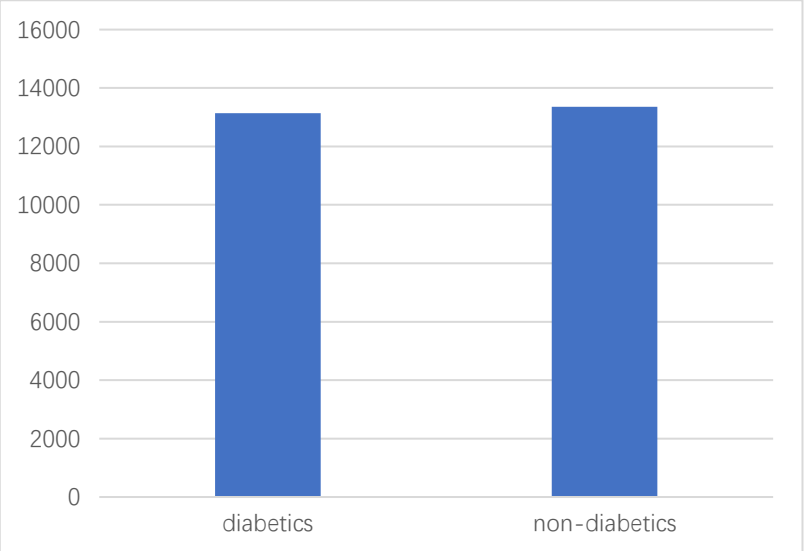

Figure 7: Claims per capita for both diabetics and non-diabetics.

However, from the figure 7, average claims for diabetics and non-diabetics show no significant difference as they are both around $13,000.

Last, I propose the hypothesis that blood pressure is positively corelated with insurance claims because high blood pressure will possibly damage brains and hearts.

DOI: 10.54254/2754-1169/22/20230327

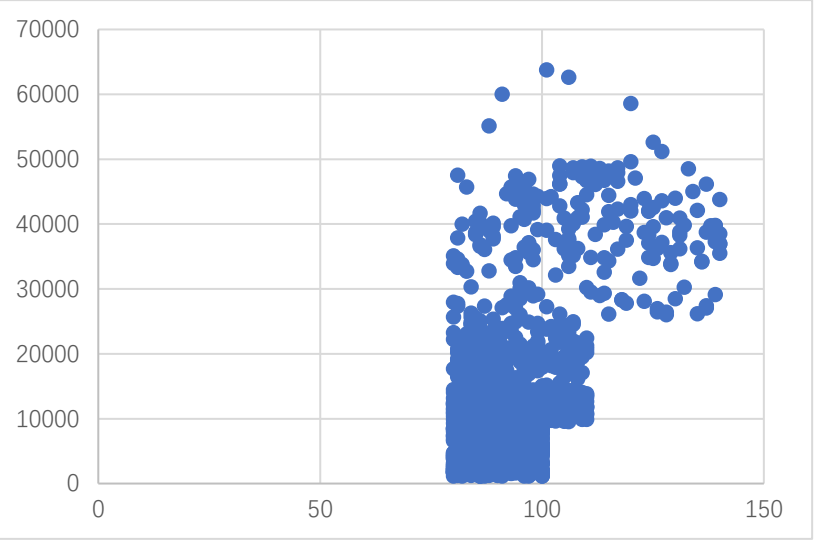

Figure 8: Blood pressure and claim distribution.

From the figure 8, most people who have insurance claims of less than $10,000 have the blood pressure around 80 to 100, while few people whose blood pressure is higher than 110 have claims of less than $20,000, which strongly support my previous hypothesis. The data results are consistent with the findings of Koshi Nakamura and Tomonori Okamura [7]: high blood pressure was a useful predictor for excess medical costs.

To summarize, I can initially get the following conclusions that BMI and blood pressure are positively corelated with insurance claims, and people who smoke, live in northeast, have less kids and are male are more likely to have higher claims, while diabetic and age seem to have no strong relation with insurance claims.

2.2. Normality Test

Comparative analyses and normality tests allow me to explore whether there are significant differences in the data between different groups classified by different types of people, and thus further investigate the relationship between the data.

First, the test for normality was performed on continuous variables in order to find out how cigarette affects insurance claim as well as body index, and the results were as follows Table 2:

Table 2: Tests of Normality.

	smoker	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
age	0	.063	1061	.000	.974	1061	.000
	1	.084	274	.000	.966	274	.000
bmi	0	.026	1061	.099	.994	1061	.000
	1	.053	274	.061	.991	274	.085
Blood pressure	0	.080	1061	.000	.928	1061	.000
	1	.076	274	.001	.951	274	.000
claim	0	.110	1061	.000	.872	1061	.000
	1	.121	274	.000	.940	274	.000

DOI: 10.54254/2754-1169/22/20230327

Considering that the sample data size exceeds 50, the significance is based on Kolmogorov-Smirnova. From the table 2-1, it is obvious that the significance of BMI is greater than 0.05, which means BMI follows normal distribution, while other variables do not.

So, I started t-tests on the BMI data and have the following observations (table 3):

Table 3: Independent samples test.

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
BMI	Equal variances assumed	1.122	.290	-.134	1338	.893
	Equal variances not assumed			-.131	411.037	.896

According to table 3, because significance(2-tailed) is greater than 0.05, there is no significant difference in BMI values between smokers and non-smokers. This may be due to the fact that BMI is primarily related to the amount of calories people consume and exercise, as well as genetics, rather than whether or not they smoke.

Next, a non-parametric test was performed on the non-normally distributed parameters.

Table 4: Hypothesis test summary.

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of blood pressure is the same across categories of smoker.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.
2	The distribution of claim is the same across categories of smoker.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.
3	The distribution of age is the same across categories of smoker.	Independent-Samples Mann-Whitney U Test	.249	Retain the null hypothesis.

According to table 4, it could be concluded that the distributions of blood pressure and claim are different across categories of smokers, while the distribution of age is the same across categories of smokers, which also support the conclusion that insurance claims of smokers are significantly higher than claims of non-smokers. The data results can be supported by Paola Primatesta [8], who found out that cigarette smoking causes various adverse cardiovascular events and acts synergistically with hypertension and dyslipidemia to increase the risk of coronary heart disease.

Second, the test for normality was performed on continuous variables in order to find out how diabetic affects insurance claim as well as body index, and the results were as follows (table 5):

DOI: 10.54254/2754-1169/22/20230327

Table 5: Tests of normality.

	diabetic	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
claim	0	.190	696	.000	.811	696	.000
	1	.190	639	.000	.817	639	.000
age	0	.064	696	.000	.972	696	.000
	1	.069	639	.000	.974	639	.000
bmi	0	.030	696	.198	.994	696	.007
	1	.033	639	.158	.992	639	.001
bloodpressure	0	.121	696	.000	.888	696	.000
	1	.114	639	.000	.874	639	.000
children	0	.237	696	.000	.826	696	.000
	1	.256	639	.000	.821	639	.000

From the table, BMI follows normal distribution while other variables do not follow normal distribution as their significance are less than 0.05. So, I started t-tests on the BMI data and have the following observations:

Table 6: Independent samples test.

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig.(2-tailed)
BMI	Equal variances assumed	.067	.795	1.353	1338	.176
	Equal variances not assumed			1.353	1328.127	.176

According to table 6, because significance(2-tailed) is greater than 0.05, there is no significant difference in BMI values between diabetics and non-diabetics.

Next, a non-parametric test was performed on the non-normally distributed parameters.

Table 7: Hypothesis test summary.

Null Hypothesis	Test	Sig.	Decision
The distribution of blood pressure is the same across categories of diabetic.	Independent-Samples Mann-Whitney U Test	.375	Retain the null hypothesis.
The distribution of age is the same across categories of diabetic.	Independent-Samples Mann-Whitney U Test	.329	Retain the null hypothesis.
The distribution of claim is the same across categories of diabetic.	Independent-Samples Mann-Whitney U Test	.863	Retain the null hypothesis.

DOI: 10.54254/2754-1169/22/20230327

From the table 7, it could be concluded that the distributions of age, blood pressure and claim are all the same across categories of diabetics, which also support the conclusion that diabetic seems to have no strong relation with insurance claims. The data results are consistent with the findings of K. Kähm and R. Stark [9], which suggest that absolute excess costs of diabetics were approximately the same in all age groups.

Third, the test for normality was performed on continuous variables in order to find out how gender affects insurance claim as well as body index, and the results were as follows:

Table 8: Tests of normality.

	gender	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
age	0	.082	662	.000	.943	662	.000
	1	.086	673	.000	.948	673	.000
bmi	0	.035	662	.054	.993	662	.003
	1	.030	673	.200	.993	673	.002
Blood pressure	0	.114	662	.000	.885	662	.000
	1	.112	673	.000	.878	673	.000
claim	0	.184	662	.000	.805	662	.000
	1	.198	673	.000	.823	673	.000
children	0	.253	662	.000	.820	662	.000
	1	.240	673	.000	.828	673	.000

From the table 8, BMI follows normal distribution while other variables do not follow normal distribution as their significance are less than 0.05. So, I started t-tests on the BMI data and have the following observations:

Table 9: Independent samples test.

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
BMI	Equal variances assumed	.016	.899	-1.714	1338	.087
	Equal variances not assumed			-1.715	1337.953	.087

According to the table 9, because significance(2-tailed) is greater than 0.05, there is no significant difference in BMI values between male and female.

Next, a non-parametric test was performed on the non-normally distributed parameters.

DOI: 10.54254/2754-1169/22/20230327

Table 10: Hypothesis test summary.

Null Hypothesis	Test	Sig.	Decision
The distribution of blood pressure is the same across categories of gender.	Independent-Samples Mann-Whitney U Test	.951	Retain the null hypothesis.
The distribution of age is the same across categories of gender.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.
The distribution of claim is the same across categories of gender.	Independent-Samples Mann-Whitney U Test	.799	Retain the null hypothesis.

From the table 10, it could be concluded that the distributions of blood pressure and claim are both the same across categories of gender, while the distribution of age is different across the category of gender, which refute the conclusion that male tend to have higher insurance claims compared to female.

Last, the test for normality was performed on continuous variables in order to find out how number of children affects insurance claim as well as body index, and the results were as follows:

Table 11: Tests of normality.

	children	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
age	0	.073	571	.000	.974	571	.000
	1	.059	324	.008	.973	324	.000
	2	.085	240	.000	.965	240	.000
	3	.085	157	.007	.967	157	.001
	4	.108	25	.200	.974	25	.743
	5	.116	18	.200	.965	18	.692
bmi	0	.030	571	.200	.993	571	.006
	1	.053	324	.030	.986	324	.003
	2	.045	240	.200	.990	240	.091
	3	.045	157	.200	.991	157	.443
	4	.170	25	.059	.926	25	.071
	5	.157	18	.200	.920	18	.131
Blood pressure	0	.120	571	.000	.868	571	.000
	1	.112	324	.000	.899	324	.000
	2	.116	240	.000	.900	240	.000
	3	.132	157	.000	.856	157	.000
	4	.115	25	.200	.958	25	.385
	5	.166	18	.200	.941	18	.304
claim	0	.184	571	.000	.817	571	.000
	1	.230	324	.000	.775	324	.000
	2	.220	240	.000	.795	240	.000
	3	.203	157	.000	.806	157	.000
	4	.201	25	.011	.829	25	.001
	5	.143	18	.200	.891	18	.040

DOI: 10.54254/2754-1169/22/20230327

From the table 11, no variance follows normal distribution as their significance are less than 0.05. So a non-parametric test was performed on all parameters.

Table 12: Hypothesis test summary.

Null Hypothesis	Test	Sig.	Decision
The distribution of blood pressure is the same across categories of children.	Independent-Samples Kruskal-Wallis Test	.275	Retain the null hypothesis.
The distribution of age is the same across categories of children.	Independent-Samples Kruskal-Wallis Test	.687	Retain the null hypothesis.
The distribution of claim is the same across categories of children.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
The distribution of BMI is the same across categories of children.	Independent-Samples Kruskal-Wallis Test	.774	Retain the null hypothesis.

From the table 12, it could be concluded that the distributions of blood pressure, age and BMI are all the same across categories of children, while the distribution of claim is different across the category of children.

To summarize, there are significant differences in the distribution of blood pressure and claim between the smokers and non-smokers, while there is no significant difference in the distribution of BMI and age between the smokers and non-smokers. Also, there is no significant difference in the distribution of BMI, age, blood pressure and claim between the diabetics and non-diabetics. Meanwhile, there is no significant difference in the distribution of BMI, blood pressure and claim between male and female, and there are significant differences in the distribution of age between male and female. Last, there are significant differences in the distribution of claim between people have different amount of kids, and there is no significant difference in the distribution of BMI, blood pressure and age between them.

2.3. Factor Analysis

Because variables such as BMI and blood pressure might have some kind of potential relations with each other, it is possible for me to simplify the model by conducting factor analysis and thus reducing the independent variables to achieve dimensionality reduction. Firstly, the variables are analyzed for correlation between them.

Table 13: KMO and bartlett's test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.529
Bartlett's Test of Sphericity	Approx. Chi-Square	521.136
	df	10
	Sig.	.000

KMO measure of sampling adequacy is a test to assess the appropriateness of using factor analysis on the data set. Bartlett' test of sphericity is used to test the null hypothesis that the variables in the population correlation matrix are uncorrelated. From the table 13, Kaiser-Meyer-Olkin Measure of

DOI: 10.54254/2754-1169/22/20230327

Sampling Adequacy is not less than 0.5 and significance of Bartlett's Test of Sphericity is less than 0.05, so there is some correlation between the variables, which allows me to do the next step.

Table 14: Total variance explained.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.636	32.729	32.729	1.636	32.729	32.729	1.532	30.642	30.642
2	1.023	20.457	53.186	1.023	20.457	53.186	1.009	20.184	50.826
3	.980	19.610	72.796	.980	19.610	72.796	1.002	20.047	70.872
4	.905	18.099	90.895	.905	18.099	90.895	1.001	20.023	90.895
5	.455	9.105	100.000						

Table 15: Communalities.

	Initial	Extraction
age	1.000	.999
Blood pressure	1.000	.786
BMI	1.000	.997
children	1.000	.992
claim	1.000	.770

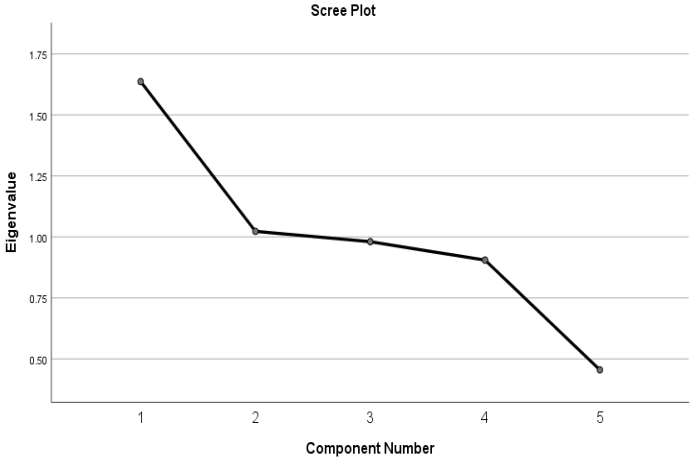

Figure 9: Scree plot.

Table 16: Component matrix.

	Component			
	1	2	3	4
claim	.846			
Blood pressure	.823			
children		.809	.574	
age		-.574	.793	
BMI	.465			.877

DOI: 10.54254/2754-1169/22/20230327

Table 17: Rotated component matrix.

	Component			
	1	2	3	4
Blood pressure	.881			
claim	.862			
children		.996		
BMI			.992	

From the scree plot, it is obvious that the first variable has the highest eigenvalue of over 1.5, while the last variable only has that of less than 0.5. Meanwhile, other variables have almost the same eigenvalue of around 1. So, I decided to remove the last variables and keep the rest.

From the table 14, the two variables with total greater than 1 explain only about 50% of the model, but four variables can explain the model with an accuracy of more than 90%. Meanwhile, from the table 15, the extractions of all variables are more than 75%, and three of them are even higher than 99%. In conclusion, the best choice considering accuracy and simplicity is to keep 4 components.

From the table 16 and table 17, the blood pressure and insurance claim can be combined into one variable, which suggest that there are some kinds of relations between them. From the result it can be possibly concluded that blood pressure is the variable most closely related to insurance claim than any other variable such as number of children, BMI or age, because there are more diseases related to high blood pressure. Claim and blood pressure together constitute the FIRST principal component with the strongest explanatory properties because claim is the most relevant to blood pressure compared to indicators such as children, BMI, and age.

2.4. Linear Regression

In order to find out how body index including blood pressure, BMI and age affect the amount of insurance claim being paid, I performed a linear regression analysis of the relevant data using stepwise, and obtained the following result:

Table 18: Linear regression result.

Model		Unstandardized Coefficients		Sig.
		B	Std. Error	
1	(Constant)	-39657.480	2329.970	.000
	Blood pressure	562.296	24.560	.000
2	(Constant)	-45487.360	2543.321	.000
	Blood pressure	542.968	24.561	.000
	BMI	249.559	45.970	.000
3	(Constant)	-46582.935	2550.842	.000
	Blood pressure	546.091	24.470	.000
	BMI	246.404	45.778	.000
	children	818.496	229.767	.000

DOI: 10.54254/2754-1169/22/20230327

From the table 18, insurance claims were related to blood pressure, BMI and number of children, but not to age. At the same time, the results yielded the same conclusions when using the other two approaches of forward and backward.

Based on the data results, it can be seen that blood pressure, BMI and number of children are all positively correlated with insurance claims, which is very obvious since higher the blood pressure, the more likely a person is to get a variety of diseases associated with high blood pressure, and thus more health insurance claims will be paid. As for BMI, when a person's BMI gets higher, he or she gets fatter and will probably lead to diabetes, hypertensions, coronary heart diseases and so on according to related research. It is worth paying attention that the data results of children contradict the conclusions above, which means people have more children tend to have higher insurance claims. For people who have more children, they usually spend more time and energy caring for their children and are therefore more likely to lack the time to exercise and the energy to diet.

3. Conclusion

From the research above, it is clear that variables such as BMI and blood pressure are strongly corelated with insurance claims, while for smokers and people who live in northeast, it is more likely for them to get higher insurance claims. However, gender, diabetic and age seem to have little relations with insurance claims. Meanwhile, blood pressure is the most related variance when explaining claim, and BMI is the second most related variance. Of all these factors, blood pressure was the most highly correlated with smoking, while age and sex were somewhat correlated. Region, child, and BMI were not correlated with the other factors, and they all affected the claim relatively independently. Diabetic was not only unrelated to all factors, but also unrelated to claim and should be excluded from the model.

For insurance companies, there are some suggestions: First, it is very important for them to test the client's blood pressure before selling them health insurance, because blood pressure is the first principal component when explaining claim. Second, when clients are not completely honest with whether they smoke, it is possible to extrapolate the answer to this question by their blood pressure, since the two factors are highly corelated. Last, it is necessary for them to introduce different premium systems in different areas, as people living in different places might have different possibility of receiving insurance claims.

For public policy makers, smoking should be banned in public places and restrictions on high-calorie junk food should be increased to improve people's high blood pressure and obesity, thus better protecting their health. Also, pay more attention to people who are overweight and have high blood pressure, if conditions allow, they can be compensated and rewarded when they improve their physical condition.

References

[1] *Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." American Economic Review, 103 (1): 178-219. DOI: 10.1257/aer.103.1.178*

[2] *Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." American Economic Review, 103 (7): 2643-82. DOI: 10.1257/aer.103.7.2643*

[3] *Melih Kirlidog, Cuneyt Asuk. 2012. A Fraud Detection Approach with Data Mining in Health Insurance. World Conference on Business.https://doi.org/10.1016/j.sbspro.2012.09.168*

[4] *https://www.kaggle.com/datasets/thedevastator/insurance-claim-analysis-demographic-and-health*

[5] *Thomas Heisser, Andreas Simon, Jana Hapfelmeier, Michael Hoffmeister and Hermann Brenner. 2022. Treatment Costs of Colorectal Cancer by Sex and Age: Population-Based Study on Health Insurance Data from Germany. https://doi.org/10.3390/cancers14153836*

DOI: 10.54254/2754-1169/22/20230327

[6] *Jay Bhattacharya, Neeraj Sood. 2006. The Economics of Obesity. (Advances in Health Economics and Health Services Research, Vol. 17), Emerald Group Publishing Limited, Bingley, pp. 279-318. https://doi.org/10.1016/S0731-2199(06)17011-9*

[7] *Nakamura, K., Okamura, T., Kanda, H. et al. Impact of Hypertension on Medical Economics: A 10-Year Follow-Up Study of National Health Insurance in Shiga, Japan. Hypertens Res 28, 859–864 (2005). https://doi.org/10.1291/hypres.28.859*

[8] *Paola Primatesta, Emanuela Falaschetti, Sunjai Gupta, Michael G. Marmot and Neil R. Poulter. 2001. Association Between Smoking and Blood Pressure https://doi.org/10.1161/01.HYP.37.2.187*

[9] *K. Kähm, R. Stark, M. Laxy, U. Schneider, R. Leidl. 2019. Assessment of excess medical costs for persons with type 2 diabetes according to age groups: an analysis of German health insurance claims data. https://doi.org/10.1111/dme.14213*