Statistical Analysis

School year 2019/2020

Fall Semester

December of 2019

Project

Analysis of the Automobile Insurance Premium in different countries

Student:

Bruna Ribeiro | student number: m20190226

Professor:

Ana Cristina Costa

Index

1. Introduction

The main objective of this project is to study if the average amount of the three-month premiums for automobile insurance charged by an insurance company in five countries is the same. In order to do this, it´s necessary to test the equality of the 5 populations' means and also if there are different means it is essential to evaluate which of the means are different.

2. Methodology

As said before, the main objective of this work is to **test the equality of the 5 populations' means**. In order to achieve this goal, it is necessary to examine several aspects of the different populations and respective samples.

The populations under study are:

$X_1 -$ Amount of three-month premiums charged by the insurance firm for a specific type and amount of coverage in a given risk category in country 1.

$X_2 -$ Amount of three-month premiums charged by the insurance firm for a specific type and amount of coverage in a given risk category in country 2.

$X_3 -$ Amount of three-month premiums charged by the insurance firm for a specific type and amount of coverage in a given risk category in country 3.

$X_4 -$ Amount of three-month premiums charged by the insurance firm for a specific type and amount of coverage in a given risk category in country 4.

$X_5 -$ Amount of three-month premiums charged by the insurance firm for a specific type and amount of coverage in a given risk category in country 5.

It was collected one sample of 50 observations for each of the countries. It is assumed that **the 5 samples are independent of each other**.

For all the following tests it was considered a **significance level of 5%**, this is, $\alpha = 5\%$. In order to see the critical values of the tests, the respective tables will be consulted.

First of all, it will be presented a **descriptive analysis** of the data. In this analysis will be observed several aspects, such as the **sample mean, sample variance, confidence intervals**, among others.

In order to **test the equality of the 5 populations' means** using the **ANOVA** it is necessary to guarantee 3 assumptions:

- The samples and also the observations used in the study are independent;
- The samples come from normal populations;
- The variance of the 5 populations is the same.

If any of these assumptions are not verified then it will be needed to use another test.

As said before, it is assumed that the **samples are independent** of each other and that the **observations are also independent**.

In order to **test if the samples come from Normal populations**, it will be performed a **distribution fitting test**. In this case, the best test to use is the **Shapiro-Wilk test** because the **parameters of the population are unknown**. This test has the following hypotheses:

$H_0: The\ sample\ comes\ from\ a\ Normal\ population, with\ \mu\ and\ \sigma\ unknown.$

$H_1: The\ sample\ does\ not\ come\ from\ a\ Normal\ population.$

These test will be performed for each of the 5 samples and the decision that will be made will have in consideration that we must **reject H_0 if $\boldsymbol{W_{obs} < W_{crit}}$**.

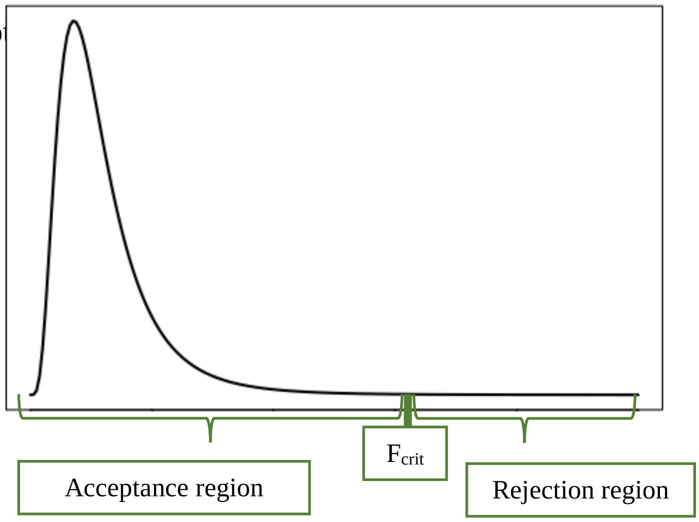
Where W_{obs} is the observed value of the test statistic and W_{crit} is the critical value of the test.

After checking this, it will be performed a **test for equality of variances of the different populations**. In this case, it was used the **Levene's test** with the **mean** as the center of each group. This because as we will see in the results the normality assumption was verified, if it wasn't it would be better to perform Levene's test using the median, which is less sensitive to variations. This test has the following hypotheses:

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$

$H_1: \exists_{i,j\ (i \neq j)} \sigma_i^2 \neq \sigma_j^2, i, j = 1,2,3,4,5$

The decision will take into consideration that we should **reject H$_0$ if $F_{obs} > F_{crit}$**, because this is a **right-sided test**, as shown below. So, if F_{obs} falls in the rejection area we must reject H$_0$, o



After checking positively all the assumptions of the **One-way ANOVA with fixed effects**, it will be finally possible to perform this test.

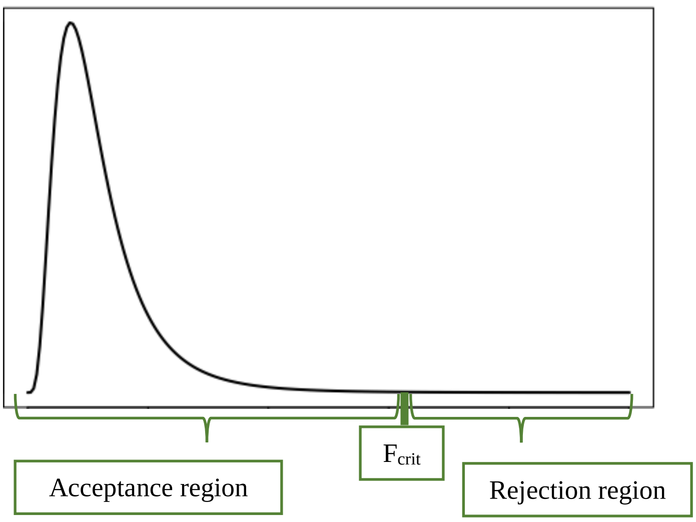
The **factor** of the one-way ANOVA is the country, which has **5 levels** (country 1 to 5) and the **experimental unit** is the three-month premium.

Its hypotheses are the following:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$H_1: \exists_{i,j\ (i \neq j)} \mu_i \neq \mu_j, i, j = 1,2,3,4,5$

The decision will take into consideration that we should **reject H$_0$ if $F_{obs} > F_{crit}$**, because as the Levenes' Test, this one is also **a right-sided test**, as shown below. So, if F_{obs} falls in the rejection area we must reject H$_0$, otherwise, we should not reject H$_0$.



As it will be shown afterward, after exploring the previous test it was concluded that not all the populations have the same mean, however, we weren't able to conclude which means have different means. Given that, it will also be needed to perform a **multiple comparison test.** In this case, the sample size is the same for the 5 countries ($n_1 = n_2 = n_3 = n_4 = n_5 = 50$) so the best test to use is **Tukey's HSD test**, which helps us to understand which pair or pairs of populations have different means. The hypotheses of this test are the following:

$H_0: \mu_i = \mu_j, i, j = 1,2,3,4,5$

$H_1: \mu_i \neq \mu_j, i, j = 1,2,3,4,5$

The decision will take into consideration that we should reject H_0 if $W_{obs} \geq q(k; n-k)$.

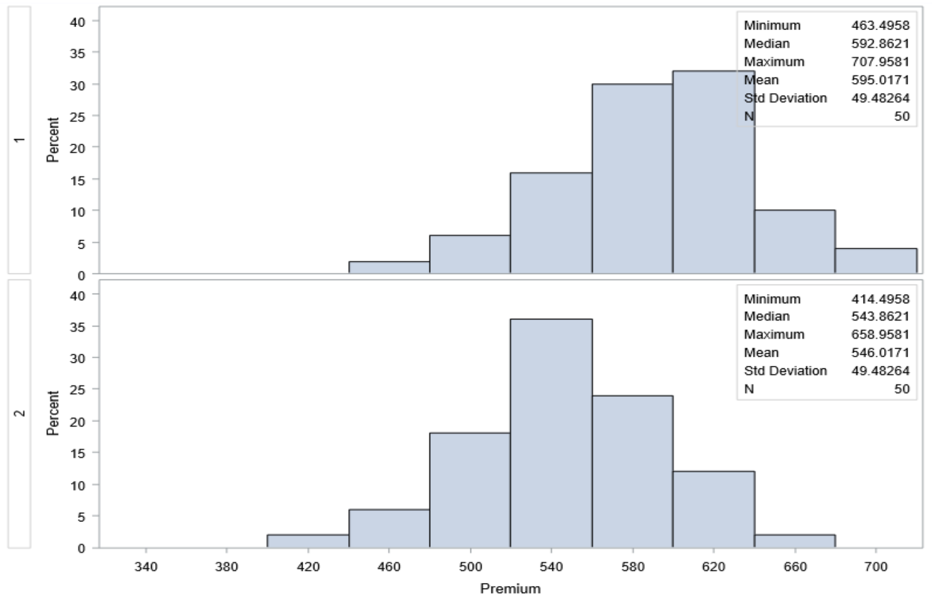
All the tests presented will be performed using the SAS Enterprise Guide, bellow is the code used. Additionally, it was also used some other tools from SAS Enterprise Guide and SPSS.

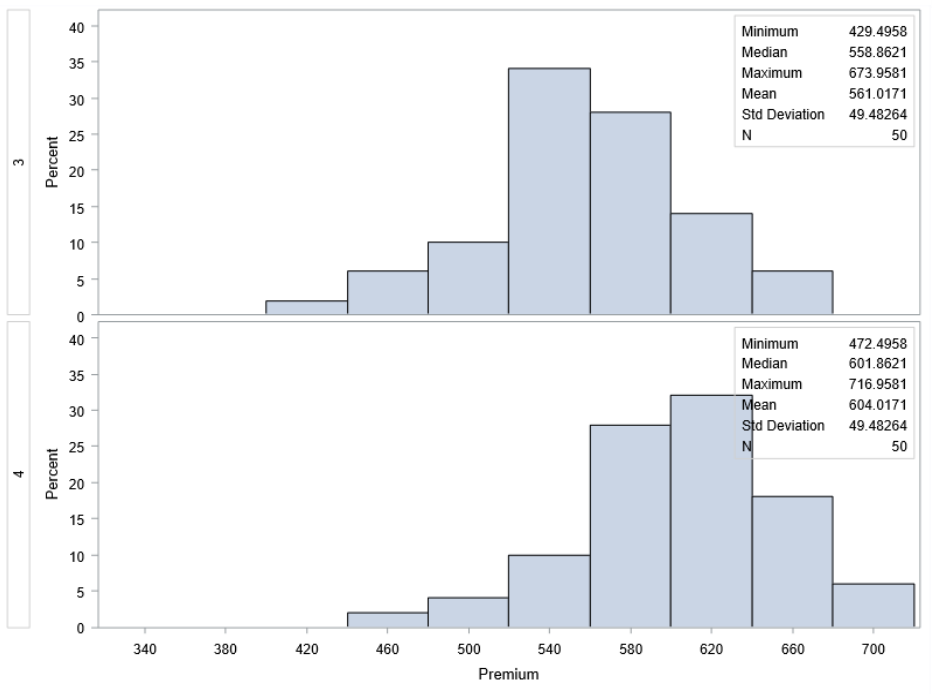
```
LIBNAME stat 'C:\Users\Utilizador\Desktop\Statistics Project';

PROC UNIVARIATE DATA=stat.automobile_10;
      CLASS Country;
      VAR Premium;
RUN;

PROC UNIVARIATE DATA=stat.automobile_10 NORMAL;
      CLASS Country;
      VAR Premium;
      HISTOGRAM Premium;
      INSET MIN MEDIAN MAX MEAN STD N / POSITION=NE;
RUN;

PROC glm DATA=stat.automobile_10;
Class Country;
Model Premium=Country;
MEANS Country/HOVTEST=LEVENE(type=abs);
RUN;

PROC ANOVA DATA=stat.automobile_10;
      CLASS Country;
      MODEL Premium=Country;
RUN;
QUIT;

proc anova data=stat.automobile_10;
     class Country;
     model Premium=Country;
     means Country/tukey;
run;
```

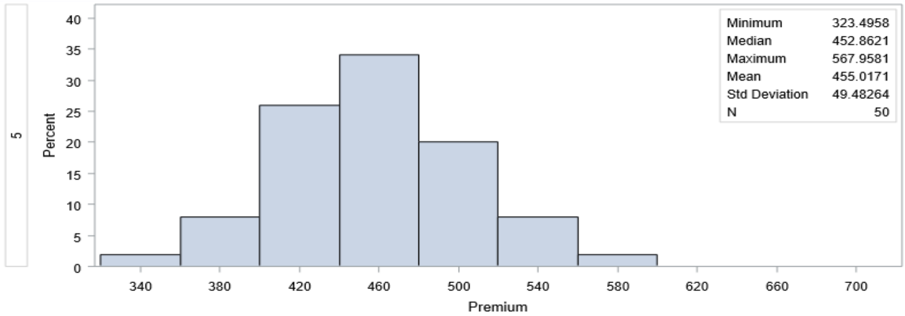
3. Results

3.1 Exploratory data analysis

In the dataset used for the analysis, there are **5 samples** of 5 different populations. Below follows the **histograms** from each of the samples, where it is possible to get a first look at the data. It is possible to check that all the samples have the **same size** (n=50) and the **mean is slightly different** in all of them, there is none with exactly the same value. The **standard deviation appears to be the same** for all of the samples (Std Deviation = 49.48264). The sample from country 4 is the one with the higher values of the three-month premiums, followed by the country 1. In opposition to country 5, which is the one with the smallest values. Finally, it is also possible to conclude that the **data doesn't seem to be far from a Normal Distribution**.

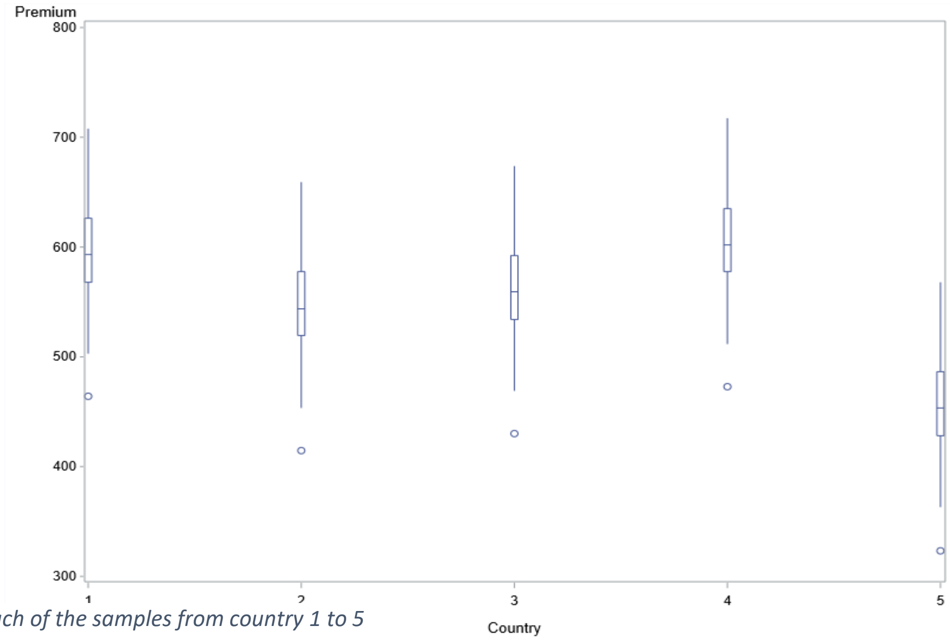




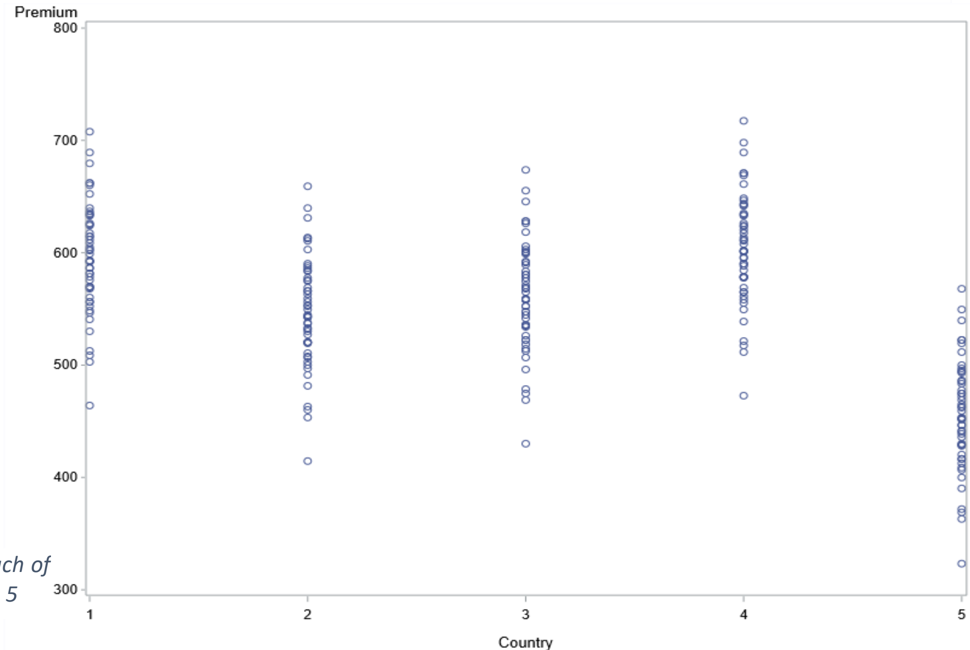


Graph 1 - Histograms from each of the samples for country 1 to 5

According to the **boxplots** that are below, it is possible to confirm that the **means from each of the samples appear to be different**, and also that all of the samples have 1 individual that has a three-month premium lower than the usual (point outside the boxplot), which means that all of the samples have **one possible outlier**. This fact will not have an impact on the analysis of this project, however, in the future analysis developed by the insurance company they may be taken into consideration. The **scatter plots** that follow confirm the previous facts and allows another view of the values.



Graph 2 - Boxplot from each of the samples from country 1 to 5



Graph 3 – Scree-Plot from each of the samples from country 1 to 5

The last analysis made in this topic was the **confidence interval for the mean, standard deviation, and variance of the populations.** This analysis was made after performing the Shapiro-Wilk test. So, it was used the test statistic for normal populations and σ^2 unknown.

Basic Confidence Limits Assuming Normality for country 1			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	595.01711	580.95430	609.07992
Std Deviation	49.48264	41.33453	61.66197
Variance	2449	1709	3802

Table 1 - Confidence intervals for the parameters of the first population

For country 1, we can say with 95% confidence that the population mean is between 580.95430 and 609.07992. The standard deviation is with 95% confidence between 41.33453 and 61.66197. At last, we have evidence to believe that the variance is between 1709 and 3802 with 95% confidence.

Basic Confidence Limits Assuming Normality for country 2			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	546.01711	531.95430	560.07992
Std Deviation	49.48264	41.33453	61.66197
Variance	2449	1709	3802

Table 2 - Confidence intervals for the parameters of the second population

For country 2, we can say with 95% confidence that the population mean is between 531.95430 and 560.07992. The standard deviation is with 95% confidence between 41.33453 and 61.66197. At last, we have evidence to believe that the variance is between 1709 and 3802 with 95% confidence.

Basic Confidence Limits Assuming Normality for country 3			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	561.01711	546.95430	575.07992
Std Deviation	49.48264	41.33453	61.66197
Variance	2449	1709	3802

Table 3 - Confidence intervals for the parameters of the third population

For country 3, we can say with 95% confidence that the population mean is between 546.95430 and 575.07992. The standard deviation is with 95% confidence between 41.33453 and 61.66197. At last, we have evidence to believe that the variance is between 1709 and 3802 with 95% confidence.

Basic Confidence Limits Assuming Normality for country 4			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	604.01711	589.95430	618.07992
Std Deviation	49.48264	41.33453	61.66197
Variance	2449	1709	3802

Table 4 - Confidence intervals for the parameters of the fourth population

For country 4, we can say with 95% confidence that the population mean is between 580.95430 and 609.07992. The standard deviation is with 95% confidence between 41.33453 and 61.66197. At last, we have evidence to believe that the variance is between 1709 and 3802 with 95% confidence.

Basic Confidence Limits Assuming Normality for country 5			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	455.01711	440.95430	469.07992
Std Deviation	49.48264	41.33453	61.66197
Variance	2449	1709	3802

Table 5 - Confidence intervals for the parameters of the fifth population

At last, for country 5 we can say with 95% confidence that the population mean is between 440.95430 and 469.07992. The standard deviation is with 95% confidence between 41.33453 and 61.66197. At last, we have evidence to believe that the variance is between 1709 and 3802 with 95% confidence.

3.2 Distribution fitting tests

To check if the samples come from Normal populations it was performed a Shapiro-Wilk test. The results given by the SAS Enterprise guide follows below.

Tests for Normality for Country 1				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.993984	Pr < W	0.9965

Table 2 – Shapiro-Wilk test in SAS

In the table of this test, we can see that the critic value of W_{crit} is 0.947.

$W_{obs} = 0.993984$

W_{crit} = 0.947 ($W_{(n=50,\alpha=0.05)}$)

Hence, $W_{obs} > W_{crit}$

Additionally, the value of the p-value is 0.9965. Which is bigger than the value of $\alpha =$ 5%.

So, we should not reject H_0, for $\alpha = 5\%$. There is evidence that the sample data of the country 1 comes from a Normal distribution.

Tests for Normality for Country 2				
Test		**Statistic**	**p Value**	
Shapiro-Wilk	**W**	0.993984	**Pr < W**	0.9965

Table 3 - Shapiro-Wilk test in SAS

$W_{obs} = 0.993984$

W_{crit} = 0.947 ($W_{(n=50,\alpha=0.05)}$)

Hence, $W_{obs} > W_{crit}$

Additionally, the value of the p-value is 0.9965. Which is bigger than the value of $\alpha =$ 5%.

So we should not reject H_0, for $\alpha = 5\%$. There is evidence that the sample data of the country 2 comes from a Normal distribution.

Tests for Normality for country 3				
Test		**Statistic**	**p Value**	
Shapiro-Wilk	**W**	0.993984	**Pr < W**	0.9965

Table 4 - Shapiro-Wilk test in SAS

$W_{obs} = 0.993984$

W_{crit} = 0.947 ($W_{(n=50,\alpha=0.05)}$)

Hence, $W_{obs} > W_{crit}$

Additionally, the value of the p-value is 0.9965. Which is bigger than the value of $\alpha =$ 5%.

So we should not reject H_0, for $\alpha = 5\%$. There is evidence that the sample data of the country 3 comes from a Normal distribution.

Tests for Normality for country 4				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.993984	**Pr < W**	0.9965

Table 5 - Shapiro-Wilk test in SAS

$W_{obs} = 0.993984$

$W_{crit} = 0.947$ ($W_{(n=50,\alpha=0.05)}$)

Hence, $W_{obs} > W_{crit}$

Additionally, the value of the p-value is 0.9965. Which is bigger than the value of $\alpha = 5\%$.

So we should not reject H_0, for $\alpha = 5\%$. There is evidence that the sample data of the country 4 comes from a Normal distribution.

Tests for Normality for country 5				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.993984	**Pr < W**	0.9965

Table 6 - Shapiro-Wilk test in SAS

$W_{obs} = 0.993984$

$W_{crit} = 0.947$ ($W_{(n=50,\alpha=0.05)}$)

Hence, $W_{obs} > W_{crit}$

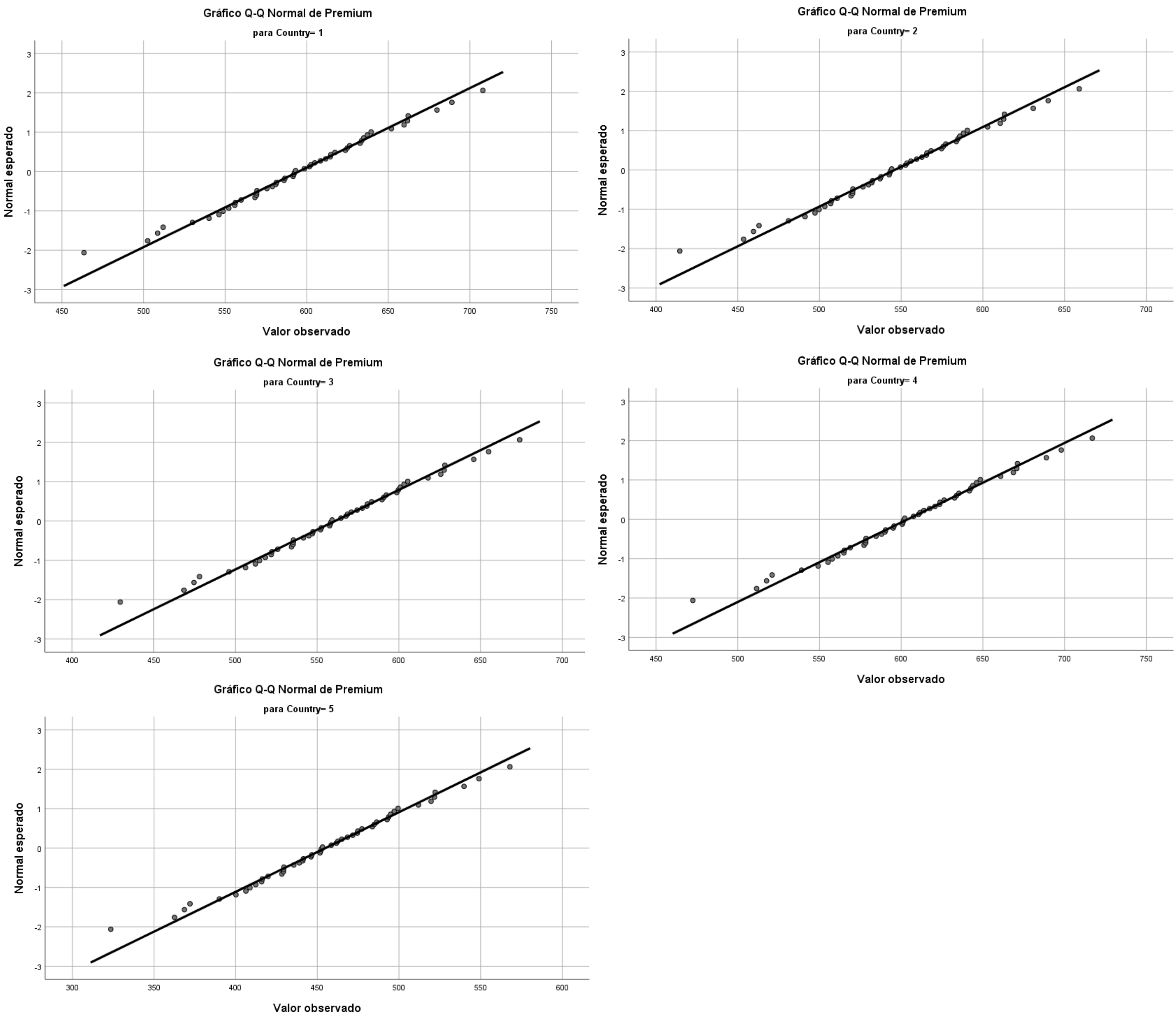
Additionally, the value of the p-value is 0.9965. Which is bigger than the value of $\alpha = 5\%$.

So we should not reject H_0, for $\alpha = 5\%$. There is evidence that the population of the country 5 has a Normal distribution.

Hence, we can conclude that there is evidence to believe that all the samples come from a Normal distribution.

It is possible to confirm these results with the observation of the Q-Q Plot. This shows that all the sample values have almost a perfect fit with the expected values of the Normal distribution.

Note: The Q-Q Normal graphs are in Portuguese because they were built with SPSS, which was in Portuguese.



Graph 2 - Normal Q-Q plots for countries 1 to 5

3.3 Tests for equality of variances

After checking the normality assumption it was performed the Levens' Test. Which presented the following results.

Levene Test for Homogeneity of Premium Variance ANOVA of Absolute Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Country	4	1.24E-25	3.09E-26	0.00	1.0000
Error	245	232090	947.3		

Table 7 - Levene Test for Homogeneity of variance

$F_{obs} = 0.00$

$F_{crit} = 2{,}4085$ (calculate using Excel)

$F_{obs} < F_{crit}$

P-value = 1.00

$\alpha = 5\%$

P-value > α

Hence, the F_{obs} falls in the acceptance region, so we should not reject H_0 for the 5% significance level. There is evidence to believe that all the populations have the same variance.

3.4 ANOVA F-test

After checking positively all the assumptions of the ANOVA, it is time to perform this test. Its output follows below.

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	4	703940.000	175985.000	71.87	<.0001
Error	245	599890.294	2448.532		
Corrected Total	249	1303830.294			

Table 8 – ANOVA F-test

$F_{obs} = 71.87$

$F_{crit} = 2{,}4085$ (calculate using Excel)

$F_{obs} > F_{crit}$

P-value = 0.0001

$\alpha = 5\%$

P-value < α

Hence, F_{obs} falls in the rejection region, so we should reject H_0 for $\alpha = 5\%$. There is evidence to believe that at least one of the population means is different from the others.

3.5 Multiple Comparison tests

In order to check which populations means are different, it was performed a Tukey's HSD test, with the following results. With the first table, it is possible to conclude that countries with significantly different means are the ones with a difference of more than 27.198. Countries with a difference of less than 27.198 have means that are not significantly different.

Alpha	0.05
Error Degrees of Freedom	245
Error Mean Square	2448.532
Critical Value of Studentized Range	3.88654
Minimum Significant Difference	27.198

Table 9 - Tukeys' HSD test

At last, the table below shows that the population means of countries 1 and 4 are not significantly different. The same statement is valid for countries 2 and 3. In opposition to country 5 that doesn´t have any similar population mean.

This means that we have evidence not to reject the null hypothesis for the pairs of means of countries 1 and 4 and also of countries 2 and 3. For all the remaining pairs of means, we should reject the null hypothesis for the 5% significance level.

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	Country
A	604.017	50	4

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**Country**
A			
A	595.017	50	1
B	561.017	50	3
B			
B	546.017	50	2
C	455.017	50	5

Table 3 - Tukey Grouping of the countries

4. Conclusion

Now that all the tests are performed it is possible to conclude that there is clear evidence that the insurance company charges different premiums according to the country of the clients. One can say that there is also evidence for the equality of the charged premiums in countries 1 and 4 and also in countries 2 and 3. Given that, this fact may be due to the similarity risk that the residents of these countries are exposed to. This means that clients from countries 1 and 4 might have the same risk which is also verified for clients in countries 2 and 3. In contrast, country 5 is not charged the same premiums as in any other country and additionally this country is the one with the lowest mean values. This may indicate that clients from country 5 have a lower exposition to risk.

Given that we are studying an automobile insurance, probably country 5 is a smaller country than the others, where there are fewer cars and less traffic and consequently, less exposition to risk of having an accident. In opposition to countries, 1 and 4 that have the highest amount of three-month premiums, so probably these are big countries, with a lot of traffic and it is possible that the vehicle drivers have bad driving habits, which is a synonym for higher exposition to risk.

5. References

Kent State University (2019). SPSS TUTORIALS: DESCRIPTIVE STATS FOR ONE NUMERIC VARIABLE (EXPLORE) [online] Available at:

https://libguides.library.kent.edu/SPSS/Explore

[Accessed 18 Dec. 2019].

Support.sas.com (2019). SAS/STAT(R) 13.1 User's Guide, The ANOVA Procedure. [online] Available at:

https://support.sas.com/documentation/onlinedoc/stat/131/anova.pdf

[Accessed 16 Dec. 2019].

Support.sas.com (2019). SAS/STAT(R) 14.3 User's Guide, The GLM Procedure. [online] Available at:

https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_glm_syntax15.htm&docsetVersion=14.3&locale=en#statug.glm.means_opt_hov

[Accessed 16 Dec. 2019].

6. Appendices

Outputs from the code used in SAS

The UNIVARIATE Procedure

Variable: Premium

Country = 1

Moments			
N	50	**Sum Weights**	50

Moments

Mean	595.017106	Sum Observations	29750.8553
Std Deviation	49.4826415	Variance	2448.53181
Skewness	-0.1543126	Kurtosis	0.27674129
Uncorrected SS	17822245.9	Corrected SS	119978.059
Coeff Variation	8.31617125	Std Error Mean	6.99790227

Basic Statistical Measures

Location		Variability	
Mean	595.0171	Std Deviation	49.48264
Median	592.8621	Variance	2449
Mode	.	Range	244.46229
		Interquartile Range	58.05489

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	**t**	85.02792	**Pr > |t|**	<.0001
Sign	**M**	25	**Pr >= |M|**	<.0001
Signed Rank	**S**	637.5	**Pr >= |S|**	<.0001

Tests for Normality

Test		Statistic	p Value	
Shapiro-Wilk	**W**	0.993984	**Pr < W**	0.9965
Kolmogorov-Smirnov	**D**	0.054417	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.015822	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.121938	**Pr > A-Sq**	>0.2500

Quantiles (Definition 5)

Level	Quantile
100% Max	707.958
99%	707.958
95%	679.834
90%	660.691
75% Q3	626.325
50% Median	592.862

Quantiles (Definition 5)	
Level	**Quantile**
25% Q1	568.270
10%	535.103
5%	508.602
1%	463.496
0% Min	463.496

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
463.496	42	661.720	13
502.521	25	662.162	44
508.602	32	679.834	39
512.000	20	688.973	30
530.037	28	707.958	46

The UNIVARIATE Procedure

Variable: Premium

Country = 2

Moments			
N	50	Sum Weights	50

Moments

Mean	546.017106	Sum Observations	27300.8553
Std Deviation	49.4826415	Variance	2448.53181
Skewness	-0.1543126	Kurtosis	0.27674129
Uncorrected SS	15026712.1	Corrected SS	119978.059
Coeff Variation	9.06247093	Std Error Mean	6.99790227

Basic Statistical Measures

Location		Variability	
Mean	546.0171	Std Deviation	49.48264
Median	543.8621	Variance	2449
Mode	.	Range	244.46229
		Interquartile Range	58.05489

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	78.02583	Pr > |t|	<.0001
Sign	M	25	Pr >= |M|	<.0001
Signed Rank	S	637.5	Pr >= |S|	<.0001

Tests for Normality

Test		Statistic	p Value	
Shapiro-Wilk	W	0.993984	Pr < W	0.9965

Tests for Location: Mu0=0

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.054417	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.015822	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.121938	Pr > A-Sq	>0.2500

Quantiles (Definition 5)

Level	Quantile
100% Max	658.958
99%	658.958
95%	630.834
90%	611.691
75% Q3	577.325
50% Median	543.862
25% Q1	519.270
10%	486.103
5%	459.602
1%	414.496
0% Min	414.496

Extreme Observations

Lowest	Highest

Quantiles (Definition 5)				
Level	Quantile			
	Value	Obs	Value	Obs
	414.496	92	612.720	63
	453.521	75	613.162	94
	459.602	82	630.834	89
	463.000	70	639.973	80
	481.037	78	658.958	96

The UNIVARIATE Procedure

Variable: Premium

Country = 3

Moments			
N	50	Sum Weights	50
Mean	561.017106	Sum Observations	28050.8553
Std Deviation	49.4826415	Variance	2448.53181
Skewness	-0.1543126	Kurtosis	0.27674129
Uncorrected SS	15856987.7	Corrected SS	119978.059
Coeff Variation	8.82016626	Std Error Mean	6.99790227

Basic Statistical Measures	
Location	Variability

Moments

Mean	561.0171	Std Deviation	49.48264
Median	558.8621	Variance	2449
Mode	.	Range	244.46229
		Interquartile Range	58.05489

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	80.16933	Pr > |t|	<.0001
Sign	M	25	Pr >= |M|	<.0001
Signed Rank	S	637.5	Pr >= |S|	<.0001

Tests for Normality

Test		Statistic	p Value	
Shapiro-Wilk	W	0.993984	Pr < W	0.9965
Kolmogorov-Smirnov	D	0.054417	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.015822	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.121938	Pr > A-Sq	>0.2500

Quantiles (Definition 5)

Level	Quantile
100% Max	673.958
99%	673.958

Quantiles (Definition 5)

Level	Quantile
95%	645.834
90%	626.691
75% Q3	592.325
50% Median	558.862
25% Q1	534.270
10%	501.103
5%	474.602
1%	429.496
0% Min	429.496

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
429.496	142	627.720	113
468.521	125	628.162	144
474.602	132	645.834	139
478.000	120	654.973	130
496.037	128	673.958	146

The UNIVARIATE Procedure

Variable: Premium

Country = 4

Moments

N	50	Sum Weights	50
Mean	604.017106	Sum Observations	30200.8553
Std Deviation	49.4826415	Variance	2448.53181
Skewness	-0.1543126	Kurtosis	0.27674129
Uncorrected SS	18361811.3	Corrected SS	119978.059
Coeff Variation	8.1922583	Std Error Mean	6.99790227

Basic Statistical Measures

Location		Variability	
Mean	604.0171	Std Deviation	49.48264
Median	601.8621	Variance	2449
Mode	.	Range	244.46229
		Interquartile Range	58.05489

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	86.31402	Pr > |t|	<.0001
Sign	M	25	Pr >= |M|	<.0001
Signed Rank	S	637.5	Pr >= |S|	<.0001

Tests for Normality

Tests for Location: Mu0=0

Test	Statistic		p Value	

Test	Statistic		p Value	
Shapiro-Wilk	W	0.993984	Pr < W	0.9965
Kolmogorov-Smirnov	D	0.054417	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.015822	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.121938	Pr > A-Sq	>0.2500

Quantiles (Definition 5)

Level	Quantile
100% Max	716.958
99%	716.958
95%	688.834
90%	669.691
75% Q3	635.325
50% Median	601.862
25% Q1	577.270
10%	544.103
5%	517.602
1%	472.496
0% Min	472.496

Quantiles (Definition 5)

Level	Quantile

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
472.496	192	670.720	163
511.521	175	671.162	194
517.602	182	688.834	189
521.000	170	697.973	180
539.037	178	716.958	196

The UNIVARIATE Procedure

Variable: Premium

Country = 5

Moments

N	50	Sum Weights	50
Mean	455.017106	Sum Observations	22750.8553
Std Deviation	49.4826415	Variance	2448.53181
Skewness	-0.1543126	Kurtosis	0.27674129
Uncorrected SS	10472006.4	Corrected SS	119978.059
Coeff Variation	10.874897	Std Error Mean	6.99790227

Moments

Basic Statistical Measures

Location		Variability	
Mean	455.0171	Std Deviation	49.48264
Median	452.8621	Variance	2449
Mode	.	Range	244.46229
		Interquartile Range	58.05489

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	65.02193	Pr > |t|	<.0001
Sign	M	25	Pr >= |M|	<.0001
Signed Rank	S	637.5	Pr >= |S|	<.0001

Tests for Normality

Test		Statistic	p Value	
Shapiro-Wilk	W	0.993984	Pr < W	0.9965
Kolmogorov-Smirnov	D	0.054417	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.015822	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.121938	Pr > A-Sq	>0.2500

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	567.958
99%	567.958
95%	539.834
90%	520.691
75% Q3	486.325
50% Median	452.862
25% Q1	428.270
10%	395.103
5%	368.602
1%	323.496
0% Min	323.496

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
323.496	242	521.720	213
362.521	225	522.162	244
368.602	232	539.834	239
372.000	220	548.973	230
390.037	228	567.958	246

The GLM Procedure

Class Level Information

Class	Levels	Values
Country	5	1 2 3 4 5

Number of Observations Read	250
Number of Observations Used	250

Levene Test for Homogeneity of Premium Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Country	4	1.24E-25	3.09E-26	0.00	1.0000
Error	245	232090	947.3		

Level of Country	N	Premium	
		Mean	Std Dev
1	50	595.017106	49.4826415
2	50	546.017106	49.4826415
3	50	561.017106	49.4826415
4	50	604.017106	49.4826415
5	50	455.017106	49.4826415

The ANOVA Procedure

Class Level Information

Class	Levels	Values
Country	5	1 2 3 4 5

Number of Observations Read	250
Number of Observations Used	250

Dependent Variable: Premium

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	703940.000	175985.000	71.87	<.0001
Error	245	599890.294	2448.532		
Corrected Total	249	1303830.294			

R-Square	Coeff Var	Root MSE	Premium Mean
0.539902	8.960722	49.48264	552.2171

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Country	4	703940.0000	175985.0000	71.87	<.0001

The ANOVA Procedure

Tukey Studentized Range (HSD) Test for Premium

Alpha	0.05
Error Degrees of Freedom	245
Error Mean Square	2448.532
Critical Value of Studentized Range	3.88654
Minimum Significant Difference	27.198

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	Country
A	604.017	50	4
A			
A	595.017	50	1
B	561.017	50	3
B			
B	546.017	50	2
C	455.017	50	5