

<https://doi.org/10.17816/MAJ17737>

## SAS ENTERPRISE GUIDE 6.1 FOR PHYSICIANS: CORRELATION ANALYSIS

N.S. Bunenkov, G.F. Bunenkova, V.V. Komok, O.A. Grinenko, A.S. Nemkov

Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia

For citation: Bunenkov NS, Bunenkova GF, Komok VV, et al. SAS Enterprise Guide 6.1 for physicians: correlation analysis. *Medical Academic Journal*. 2020;20(1):51-56. <https://doi.org/10.17816/MAJ17737>

Received: November 14, 2019

Revised: January 13, 2020

Accepted: March 2, 2020

**Objective:** to develop algorithm of correlation analysis of prospective non-randomized clinical trial AMIRI–CABG (ClinicalTrials.gov Identifier: NCT03050489) data using SAS Enterprise Guide 6.1.

**Materials and methods.** Data collection was performed according prospective non-randomized clinical trial AMIRI–CABG (ClinicalTrials.gov Identifier: NCT03050489) in Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia between 2016–2019 years with 336 patients. There is database with clinical, laboratory and instrumental data. Correlation analysis was performed with SAS Enterprise Guide 6.1.

**Results.** There was developed algorithm of correlation analysis data of prospective non-randomized clinical trial AMIRI–CABG (ClinicalTrials.gov Identifier: NCT03050489). This algorithm could be useful for physicians and researchers for data analysis.

**Conclusion.** Presented algorithm of correlation analysis could make easier and improve efficient data analysis with SAS Enterprise Guide 6.1.

**Keywords:** SAS Enterprise Guide 6.1; statistical analysis; clinical trials; correlation analysis; Pearson coefficient correlation; Spearman coefficient correlation.

## SAS ENTERPRISE GUIDE 6.1 ДЛЯ ВРАЧЕЙ: КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Н.С. Буненков, Г.Ф. Бунenkova, В.В. Комок, О.А. Гриненко, А.С. Немков

Федеральное государственное бюджетное образовательное учреждение высшего образования «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова»  
Министерства здравоохранения Российской Федерации, Санкт-Петербург

Для цитирования: Буненков Н.С., Бунenkova Г.Ф., Комок В.В., и др. SAS Enterprise Guide 6.1 для врачей: корреляционный анализ // Медицинский академический журнал. – 2020. – Т. 20. – № 1. – С. 51–56. <https://doi.org/10.17816/MAJ17737>

Поступила: 14.11.2019

Одобрена: 13.01.2020

Принята: 02.03.2020

**Цель** — разработать алгоритм обработки базы данных проспективного нерандомизированного исследования Assessment of Myocardial Ischemic-Reperfusion Injury During Off- and On-Pump CABG (AMIRI–CABG, ClinicalTrials.gov Identifier: NCT03050489) в программном пакете SAS Enterprise Guide 6.1 для определения наличия и оценки силы корреляции между исследуемыми переменными.

**Материалы и методы.** В проспективное нерандомизированное исследование AMIRI–CABG (ClinicalTrials.gov Identifier: NCT03050489) в ПСПбГМУ им. И.П. Павлова с 2016 по 2019 г. включено 336 пациентов. Создана база данных с результатами клинических, лабораторных и инструментальных исследований. Статистическая обработка данных выполнена с использованием лицензионного программного обеспечения SAS Enterprise Guide 6.1. Корреляционный анализ проводили с применением коэффициентов Пирсона и Спирмена.

**Результаты.** Разработан алгоритм обработки данных проспективного нерандомизированного исследования AMIRI–CABG (ClinicalTrials.gov Identifier: NCT03050489), позволяющий выполнить корреляционный анализ. Данный алгоритм может быть использован врачами различных специальностей и научными сотрудниками для обработки результатов научных или клинических исследований.

**Заключение.** С помощью разработанного алгоритма обработки базы данных научного исследования и программного пакета SAS Enterprise Guide 6.1 врачи и научные сотрудники смогут упростить и ускорить корреляционный анализ результатов исследований.

**Ключевые слова:** SAS Enterprise Guide; статистическая обработка данных; статистика; клинические исследования; научные исследования; корреляционный анализ; коэффициент корреляции Пирсона; коэффициент корреляции Спирмена.

## Introduction

One of the aims of scientific research is to establish a relationship between studied parameters; for example, a marker in blood plasma and a drug concentration. It is widely believed that to identify the relationship between two variables, a correlation should be evaluated. The word “relationship” is often replaced by the word “correlation.” However, the presence of a correlation does not imply a direct causal relationship, just as the absence of a correlation does not exclude the existence of a relationship between two variables, including a causal one [3]. One of the question of the clinical trial at the cardiac surgery center of the Pavlov First Saint Petersburg State Medical University was whether there is a relationship between duration of aortic cross-clamping during coronary artery bypass grafting (CABG) and the troponin I level after procedure?

In our example, we used the following variables for analysis: TnIEndOp as the troponin I level at the end of procedure, TnI1 as the troponin I level on day 1 after CABG, and AoClamp as the duration of aortic cross-clamping in minutes. Patients were allocated into the three groups: group 1 — off-pump CABG, group 2 — on-pump CABG, and group 3 — CABG with parallel cardiopulmonary bypass.

**This study aimed** to develop an algorithm for processing the database of the AMIRI–CABG prospective nonrandomized clinical trial (ClinicalTrials.gov Identifier: NCT03050489) using the SAS Enterprise Guide 6.1 software package.

## Materials and methods

The AMIRI–CABG prospective nonrandomized clinical trial (ClinicalTrials.gov Identifier: NCT03050489) was performed at the Research Center for Cardiovascular Surgery of the Pavlov First Saint Petersburg State Medical University between 2016 and 2019 and enrolled 336 patients with coronary heart disease who had indications for coronary revascularization surgery. A database with the results of clinical, laboratory, and instrumental studies was created. Statistical database processing was performed using SAS Enterprise Guide 6.1 licensed software.

## Results and discussion

The type of distribution of the studied variables was determined prior to the correlation analysis [1]. If the distribution differed from normal, it can often be transformed to a normal distribution, which has been described previously [1]. In our case, the distribution of the variable TnIEndOp (level of troponin I at the end of procedure) was became normal after logarithmic transformation. The variables TnI1 (level of troponin I on day 1 after procedure) and AoClamp (duration of aortic cross-clamping) were not normally distributed. The correlation was calculated for the normally distributed variables using Pearson’s correlation coefficient analysis. Spearman’s correlation coefficient analysis was used for the non-normally distributed variables. It should be noted that Pearson’s correlation coefficient enables identifying a linear relationship. If the relationship is nonlinear, then this method will reveal the absence of a correlation [2].

Then a code which is presented can be used to determine the nature of the distribution as well as the distribution can be brought to normal if it differs from normal.

```
ods graphics on;
/* The normality of the distribution
is tested according to the Kolmogorov-
Smirnov test or the Shapiro-Wilk test,
if p>0.05 then the distribution is
normal. */
Proc UNIVARIATE DATA=WORK.'20_06_2019
work'n normaltest plots;
  where CPBType=2;
  VAR TnIEndOp TnI1 AoClamp;
run;
/* Bringing to the normal distribution
using the logarithm */
DATA NEWDATASET; /* Create a new data
table */
  SET WORK.'20_06_2019 work'n;
/* Transfer all our data to the new
table */
  LGTnI1=LOG10(TnI1); /* Logarithm to
“normalize” */
  LGTnIEndOp=LOG10(TnIEndOp);
  LGAoClamp=LOG(AoClamp);
RUN;
/* TnIEndOp - troponin level at the
end of the surgery
TnI1 - troponin level on the
1st day after the surgery
AoClamp - duration of the aortic
cross-clamping; */
```

We created a scatter plot after determining the nature of the variable distributions (Fig. 1). The scatter plot shows whether there was a corre-

lation between the increase in troponin I level at the end of surgery and on day 1 after procedure.

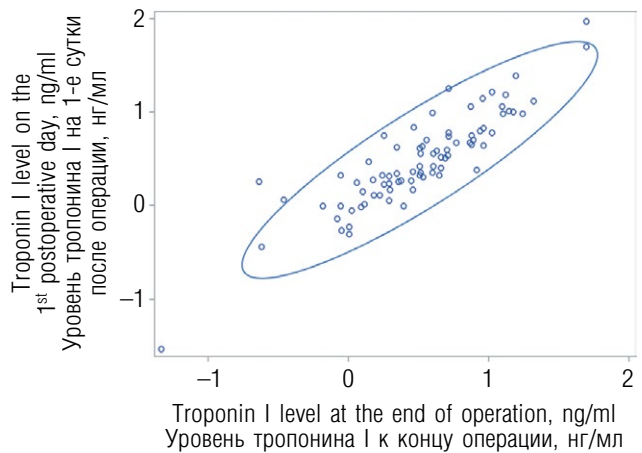
```
/* We evaluate the nature of the
relationship between the level of
troponin I at the end of the
surgery and the level of troponin I
on day 1 after the surgery */
proc sgplot data=NEWDATASET;
  WHERE CPBTYPE=2;
  title "Scattergram";
  scatter x=LGTnIEndOp y=LGTnI1;
  ellipse x=LGTnIEndOp y=LGTnI1;
  label LGTnIEndOp = 'troponin I
level at the end of the surgery';
  label LGTnI1 = 'troponin I level
on day 1 after the surgery';
run;
```

Fig. 1 shows that the points tended to be distributed along an oblique straight line; therefore, the dependence of the two variables was linear and a linear correlation analysis can be applied.

```
TITLE 'Correlation of the initial
values of troponin I at the end of
the surgery and on day 1 after the
surgery';
proc corr DATA=WORK.'20_06_2019 work'n
pearson spearman kendall hoeffding
fisher;
  WHERE CPBTYPE=2;
  var TnI1;
  with TnIEndOp;
run;
```

#### Notes.

pearson — indicates to calculate Pearson's correlation coefficient (two variables under study are normally distributed);



- Troponin I level at the end of operation, ng/ml  
Уровень тропонина I к концу операции, нг/мл
- 95% prediction ellipse  
95 % доверительный эллипс

Fig. 1. Scatter plot for correlation analysis of troponin I level at the end of operation and on the 1<sup>st</sup> postoperative day

Рис. 1. Скатерограмма зависимости уровня тропонина I на первые сутки после операции от уровня тропонина I к концу операции

spearman — indicates to calculate Spearman's correlation coefficient (one or two of the variables under study have a distribution different from normal);

kendall — indicates to calculate Kendall's correlation coefficient;

fisher — indicates to calculate the confidence interval of the correlation.

Click the “Run” button (Fig. 2).

The following table was generated as results (Fig. 3).

```
ods graphics on;
/*
Выполняем проверку на нормальность распределения по
Критериям Колмогорова-Смирнова или Шапиро-Уилка, если p>0,05 то распределение нормальное
*/
proc UNIVARIATE DATA=WORK.'20_06_2019 work'n normaltest plots;
  where CPBType=2;
  VAR TnIEndOp TnI1 AcClamp;
run;
/*
приведем к нормальному распределению с помощью логарифмирования
*/
DATA NEWDATASET; /* Создаем новую таблицу данных */
SET WORK.'20_06_2019 work'n; /* переносим в новую таблицу все наши данные */
LGTnI1=LOG10(TnI1); /* логарифмируем */
LGTnIEndOp=LOG10(TnIEndOp);
LGAcClamp=LOG(AcClamp);
RUN;
/*
TnIEndOp - уровень тропонина к концу операции
TnI1 - уровень тропонина на 1 сутки после операции
AcClamp - время пережатия аорты;
Оценим характер связи между уровнем тропонина I к концу операции и уровнем тропонина I на 1 сутки
*/
proc sgplot data=WORK.'20_06_2019 work'n ;
  WHERE CPBTYPE=2;
  title "Скатерограмма";
  scatter x=TnIEndOp y= TnI1;
  ellipse x=TnIEndOp y=TnI1;
  label TnIEndOp = 'уровень тропонина I к концу операции';
  label TnI1 = 'уровень тропонина I на 1-е сутки после операции';
run;
/*
Рассчитаем корреляцию уровня тропонина I к концу операции и на 1 сутки после операции
*/
TITLE 'Корреляция исходных значений тропонина I к концу операции и на 1 сутки после операции';
proc corr DATA=WORK.'20_06_2019 work'n pearson spearman kendall hoeffding fisher;
  WHERE CPBTYPE=2;
  var TnI1;
```

Fig. 2. Correlation analysis

Рис. 2. Расчет корреляции

Pearson Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
TnI1	TnIEndOp	81	0.90346	1.49072	0.00565	0.90241	0.851930	0.936278	<.0001

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
TnI1	TnIEndOp	81	0.87681	1.36179	0.00548	0.87553	0.812515	0.918325	<.0001

Fig. 3. Correlation analysis results

Рис. 3. Результаты корреляционного анализа

As one of the variables had a distribution that differed from normality, we obtained the results from the Spearman's correlation statistics section (see Fig. 3). The correlation value was 0.88 (strong correlation), the 95% confidence interval was 0.81–0.92, and the significance level was  $p < 0.0001$ . The coefficient of determination was  $0.88^2 = 0.77$ , so one variable explained 77% of the variability of the other, which indicates a strong relationship between the two variables.

Thus, the level of troponin I on day 1 after procedure strongly correlated with the level of troponin I at the end of surgery. This result was expected based on the pathophysiology of the myocardial ischemic reperfusion injury.

We performed a correlation analysis to identify the relationship between the duration of aortic cross-clamping (AoClamp) and the troponin I level at the end of surgery (TnIEndOp). The scatter plot was generated as follows:

```
TITLE 'Dependence of troponin I level
on the duration of the aortic cross-
clamping';
```



Fig. 4. Scatter plot aorta clamping time – troponin I level

Рис. 4. Скатерограмма между временем пережатия аорты и уровнем тропонина I к концу операции

```
proc sgplot data=NEWDATASET;
WHERE CPBType=2;
title "Scattergram";
scatter x=AoClamp y=LGTnIEndOp;
ellipse x=AoClamp y=LGTnIEndOp;
label AoClamp = 'aortic compression
time';
label TnIEndOp = 'troponin I level
at the end of the surgery';
run;
```

We obtained a scatter plot (Fig. 4) of the dependence of the troponin I level at the end of surgery on duration of the aortic cross-clamping.

The wide scatter of points from the boundaries of the ellipse is noteworthy, so it is rather difficult to draw a line along where most points were located, and the ellipse tended to be in the contour of a circle (see Fig. 4). The scatter plot revealed no correlation between duration of the aortic cross-clamping and the increase of troponin I at the end of procedure.

We calculated Spearman's correlation coefficient using the following:

```
TITLE 'Correlation of troponin I level
at the end of surgery and duration of
the aortic cross-clamping';
proc corr DATA=NEWDATASET pearson
spearman kendall hoeffding fisher;
where CPBTYPE=2;
var AoClamp;
with LGTnIEndOp;
label LGTnIEndOp = 'troponin I
level at the end of the surgery';
run;
```

The result obtained is presented in Fig. 5.

The correlation between duration of the aortic cross-clamping and the level of troponin I at the end of procedure was 0.4 (weak). The 95% confidence interval was 0.19–0.57 which is relatively wide, but the significance level was  $p = 0.0002$  (correlation is statistically significant). The coefficient of determination was  $0.4^2 = 0.16$ . Based on the calculations, one variable explained only 16% of the variability of the other variable, which is extremely low. Thus, we concluded that duration of the aortic cross-clamping does not affect the increase in the level of troponin I at the

Pearson Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
AoClamp	LGTnlEndOp	80	0.45889	0.49591	0.00290	0.45660	0.263293	0.614650	<.0001

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
AoClamp	LGTnlEndOp	80	0.39768	0.42089	0.00252	0.39556	0.192582	0.566080	0.0002

Fig. 5. Correlation analysis aorta clamping time – troponin I level

Рис. 5. Корреляция между временем пережатия аорты и уровнем тропонина I к концу операции

end of procedure.  $r$  values  $< 0.4$  are considered as a weak correlation;  $0.4 < r < 0.8$ , the shows medium strength of correlation, and  $r > 0.8$  is a strong correlation.

As was mentioned above, the absence of a correlation may indicate the absence of a linear relationship between the two variables, but a nonlinear association is also possible. The scatter plot presented in Fig. 4 tended to be uniform, which indicates the absence of linear and nonlinear relationships. The correlation analysis showed the absence of correlation in the case of nonlinear dependence; therefore, scatter plot should be created before correlation analysis. An example of a nonlinear relationship between two variables is presented in Fig. 6.

Correlation analysis is widely referred to in the Russian literature, but knowledge of biostatistics is necessary to understand the limits of its applicability and to interpret the results [2].

Despite the absence of a correlation, the presence of a relationship between the studied variables was not excluded. Perhaps it is worth trying to bring one of the variables under study to a binary form and apply logistic regression.

The correlation analysis algorithm developed is presented in such a way that a researcher who has never worked with the SAS Enterprise Guide 6.1 software product can process a database. One of the advantages of SAS is the ability to calculate the confidence interval for the correlation coefficient. Term “Fisher” should only be specified in the correlation analysis procedure, while this is not possible in other software products, and confidence intervals for the correlation coefficients must be calculated manually [2]. This article discusses the issues of practical application of a correlation analysis in the SAS software package; the theoretical aspects are reviewed only to a small extent, as they are represented well in the literature [2–5].

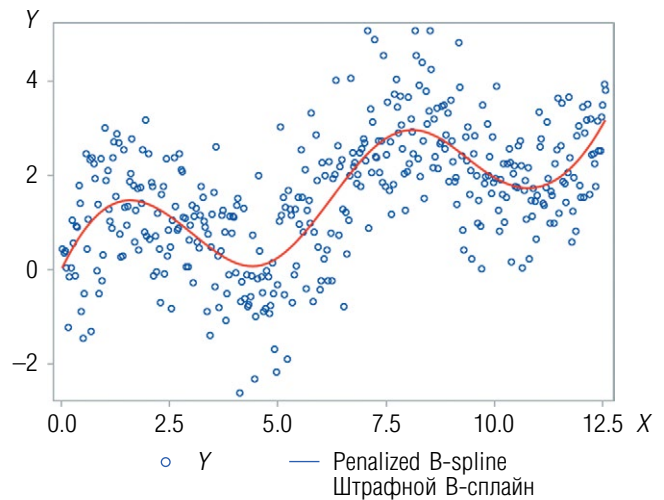


Fig. 6. Nonlinear relationship

Рис. 6. Нелинейная связь между двумя переменными

Thus, SAS Enterprise Guide 6.1 provides a complete set of modern data processing methods necessary for the medical researcher.

## Conclusion

1. SAS Enterprise Guide 6.1 enables quick and convenient correlation analyses, which makes this software package useful for researchers.
2. The correlation analysis algorithm developed can be used by researchers to process various databases from scientific studies and clinical trials.

## Additional information

**Source of funding.** The study was conducted under the state task on the topic “Assessing the regenerative potential of a patient during heart surgery.”

**Ethical considerations.** The study protocol was approved by the local Ethics Committee of the Pavlov First Saint Petersburg State Medical University.

**Conflict of interests.** The authors declare no conflicts of interest.

## References

1. Гржибовский А.М. Корреляционный анализ // Экология человека. — 2008. — № 9. — С. 50–60. [Grjibovski AM. Correlation analysis. *Ecology, human*. 2008;(9):50-60. (In Russ.)]
2. Гржибовский А.М., Иванов С.В., Горбатова М.А. Корреляционный анализ данных с использованием программного обеспечения STATISTICA и SPSS // Наука и здравоохранение. — 2017. — № 1. — С. 7–36. [Grjibovski AM, Ivanov SV, Gorbatova MA. Correlation analysis of data using statistica and spss software. *Nauka i zdravo-okhranenie*. 2017;(1):7-36. (In Russ.)]
3. Гржибовский А.М., Иванов С.В., Горбатова М.А. Экологические (корреляционные) исследования в здравоохранении // Наука и здравоохранение. — 2015. — № 5. — С. 5–18. [Grjibovski AM, Ivanov SV, Gorbatova MA. Ecological (correlation) studies in health sciences. *Nauka i zdravo-okhranenie*. 2015;(5):5-18. (In Russ.)]
4. Унгурияну Т.Н., Гржибовский А.М. Корреляционный анализ с использованием пакета статистических программ STATA // Экология человека. — 2014. — Т. 9. — С. 60–64. [Unguryanu TN, Grjibovski AM. Correlation analysis using STATA. *Ecology, human*. 2014;9:60-64. (In Russ.)]

## Information about the authors / Сведения об авторах

*Nikolay S. Bunenkov* — aspirant, Department of Faculty Surgery, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0003-4331-028X>. E-mail: bunenkov2006@gmail.com.

*Gulnara F. Bunenkova* — resident, Department of Hospital Therapy, First Pavlov State Medical University, Saint Petersburg, Russia. E-mail: gulnara533@gmail.com.

*Vladimir V. Komok* — PhD, Cardiac Surgeon, Department of Cardiac Surgery No. 2, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0002-3834-7566>. E-mail: vladimir\_komok@mail.ru.

*Oleg A. Grinenko* — Doctor of Medical Science, Vice-rector, First Pavlov State Medical University, Saint Petersburg, Russia. E-mail: klinika@spb-gmu.ru.

*Alexander S. Nemkov* — Doctor of Medical Science, Professor, Cardiac Surgeon, Department of Cardiac Surgery No. 2, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0002-5152-0001>. E-mail: nemk\_as@mail.ru.

*Николай Сергеевич Буненков* — аспирант кафедры факультетской хирургии, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0003-4331-028X>. E-mail: bunenkov2006@gmail.com.

*Гульнара Физулиевна Буненкова* — клинический ординатор кафедры госпитальной терапии, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. E-mail: gulnara533@gmail.com.

*Владимир Владимирович Комок* — канд. мед. наук, кардиохирург отделения кардиохирургии № 2, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0002-3834-7566>. E-mail: vladimir\_komok@mail.ru.

*Олег Александрович Гриненко* — д-р мед. наук, проректор по лечебной работе, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. E-mail: klinika@spb-gmu.ru.

*Александр Сергеевич Немков* — д-р мед. наук, профессор кафедры факультетской хирургии, кардиохирург отделения кардиохирургии № 2, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0002-5152-0001>. E-mail: nemk\_as@mail.ru.

## ✉ Corresponding author / Контактное лицо

*Nikolay S. Bunenkov* / *Николай Сергеевич Буненков*  
E-mail: bunenkov2006@gmail.com