

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/316548846>

Statystyka od podstaw z systemem SAS (Basic Statistics with SAS)

Book · January 2013

CITATIONS

2

READS

1,712

2 authors:



[Ewa Fratzczak](#)

SGH Warsaw School of Economics

92 PUBLICATIONS 308 CITATIONS

[SEE PROFILE](#)



[Adam Korczyński](#)

SGH Warsaw School of Economics

5 PUBLICATIONS 7 CITATIONS

[SEE PROFILE](#)

Statystyka od podstaw z systemem SAS

Ewa Frątczak
Adam Korczyński



OFICyna WYDAWNICZA
SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE

**STATYSTYKA
OD PODSTAW
Z SYSTEMEM SAS**

wersja 9.2 i 9.3

**Szkoła Główna Handlowa
w Warszawie**

Spis treści

Wprowadzenie	6
1. Podstawowe informacje o systemie SAS	9
1.1. Informacje ogólne.....	9
1.2. Analityka	14
1.3. Środowisko pracy	26
1.4. Zbiory danych, biblioteki	30
2. SAS Enterprise Guide, wersja 4.3	34
2.1. Informacje ogólne.....	34
2.2. Organizacja pracy w EG.....	36
2.3. Podstawowe operacje na zbiorach danych	40
2.3.1. Wczytywanie zbiorów danych	40
2.3.2. Transformacja zmiennych	42
2.3.3. Tworzenie podstawowych zapytań SQL.....	54
2.3.4. Raportowanie na podstawie zbiorów danych	59
3. Wybrane zagadnienia metody reprezentacyjnej	77
3.1. Wprowadzenie.....	77
3.2. Wybrane schematy losowania próby.....	79
3.2.1. Losowanie proste.....	79
3.2.2. Losowanie warstwowe	81
3.2.3. Losowanie systematyczne	84
4. Analiza struktury	86
4.1. Wprowadzenie.....	86
4.2. Tabele liczebności i częstości.....	87
4.3. Miary położenia rozkładu, zróżnicowania, asymetrii oraz koncentracji.....	92
4.4. Obserwacje odstające	100
5. Rozkłady teoretyczne zmiennych losowych	108
5.1. Graficzna prezentacja wybranych rozkładów zmiennych losowych	108
5.2. Badanie zgodności rozkładu empirycznego z rozkładem teoretycznym.....	142
6. Wybrane zagadnienia wnioskowania statystycznego.....	151
6.1. Istota wnioskowania statystycznego	151
6.2. Wnioskowanie statystyczne	158
6.3. Testy parametryczne.....	159
6.3.1. Test istotności dla średniej	159
6.3.2. Test istotności dla różnicy średnich	170
6.3.3. Test istotności dla wariancji	175
6.3.4. Test istotności dla dwóch wariancji	180
6.3.5. Test istotności dla frakcji	182
6.4. Testy nieparametryczne.....	185
6.4.1. Test znaków (test istotności dla mediany)	185
6.4.2. Test serii	192
6.4.3. Test niezależności χ^2	196
7. Analiza wariancji i kowariancji	202
7.1. Wprowadzenie.....	202
7.2. Analiza wariancji z pojedynczą klasyfikacją	204
7.3. Analiza wariancji z podwójną klasyfikacją.....	218
7.4. Analiza kowariancji.....	227

8. Analiza korelacji i regresji	235
8.1. Analiza korelacji	235
8.1.1. Współczynnik korelacji liniowej Pearsona	235
8.1.2. Nieparametryczne miary zależności.....	241
8.1.2.1. Współczynnik korelacji rang Spearmana.....	241
8.1.2.2. Współczynnik τ -Kendalla	243
8.1.2.3. Miara zależności Hoeffdinga	244
8.2. Klasyczny model regresji liniowej.....	248
8.3. Regresja wieloraka	270
9. Analiza szeregów czasowych	279
9.1. Wprowadzenie.....	279
9.2. Szereg czasowy bez okresowości.....	285
9.2.1. Szereg czasowy i jego składniki.....	285
9.2.2. Wygładzenie szeregu za pomocą średnich ruchomych.....	288
9.2.3. Wyrównywanie wykładnicze	295
9.2.4. Wygładzenie szeregu za pomocą funkcji liniowej.....	298
9.3. Szereg czasowy z okresowością.....	305
9.3.1. Szereg czasowy z wahaniami addytywnymi.....	305
9.3.2. Dekompozycja szeregu czasowego z okresowością – model addytywny.....	310
9.3.3. Szereg czasowy z wahaniami multiplikatywnymi	316
9.3.4. Dekompozycja szeregu czasowego z okresowością – model multiplikatywny ..	317
9.4. Prognozowanie w szeregach czasowych – metody analityczne.....	322
9.4.1. Addytywny liniowy model tendencji rozwojowej	322
9.4.2. Model autoregresji (AR)	330
Bibliografia	338
Załącznik 1. Słowniczek poleceń w SAS Enterprise Guide.....	342
Załącznik 2. Słowniczek procedur 4GL wykorzystywanych przez EG.....	343
Załącznik 3. Sylabus do przedmiotu Statystyka – program obowiązujący w roku akademickim 2012/2013.....	344

Contents

Introduction

1. Basic Information about SAS System

- 1.1. General Information
- 1.2. Analytics
- 1.3. Work Environment
- 1.4. Data Sets, Libraries

2. SAS Enterprise Guide 4.3

- 2.1. General Information
- 2.2. Work in EG
- 2.3. Basic Operations on Data Sets
 - 2.3.1 Reading Data Sets into EG
 - 2.3.2 Transforming Variables
 - 2.3.3 Creating SQL Queries
 - 2.3.4 Reporting with Data Sets

3. Sampling – Selected Aspects

- 3.1. Introduction
- 3.2. The Sample Selection Schemes
 - 3.2.1. Simple Sampling
 - 3.2.2. Stratified Sampling
 - 3.2.3. Systematic Sampling

4. Descriptive Statistics

- 4.1. Introduction
- 4.2. Frequency Distribution Tables
- 4.3. Measures of Location, Measures of Variation, Measures of Asymmetry and Measures of Concentration
- 4.4. Outliers

5. Discrete and Continuous Probability Distributions

- 5.1. Graphical Representation of Probability Distribution
- 5.2. Goodness-of-Fit Tests

6. Inferential Statistics – Selected Aspects

- 6.1. Introduction to the Concept of Statistical Inference
- 6.2. Inferential Statistics Using EG
- 6.3. Parametric t-Tests
 - 6.3.1. Test of the Mean
 - 6.3.2. Test of Difference between Two Means
 - 6.3.3. Test of the Variance
 - 6.3.4. Test of Difference between Two Variances
 - 6.3.5. Test of the Proportion
- 6.4. Non-Parametric Tests
 - 6.4.1. Sign Test (test of the median)
 - 6.4.2. Runs Test
 - 6.4.3. χ^2 Tests for independence

7. Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA)

- 7.1. Introduction
- 7.2. One-Way Analysis of Variance
- 7.3. Two-Way Analysis of Variance
- 7.4. Analysis of Covariance

8. Correlation and Regression Analysis

- 8.1. Correlation Analysis
 - 8.1.1. Pearson Correlation Coefficient
 - 8.1.2. Non-Parametric Measures of Correlation and Dependence
 - 8.1.2.1. Spearman Correlation Coefficient
 - 8.1.2.2. Kendall's τ Coefficient
 - 8.1.2.3. Hoeffding Measure of Dependence
- 8.2. Simple Linear Regression Model
- 8.3. Multiple Linear Regression Model

9. Time Series Analysis

- 9.1. Introduction
- 9.2. Non-Seasonal Time Series
 - 9.2.1. Components of Time Series
 - 9.2.2. Moving Averages Smoothing
 - 9.2.3. Exponential Smoothing
 - 9.2.4. Linear Regression Smoothing
- 9.3. Seasonal Time Series
 - 9.3.1. Additive time series
 - 9.3.2. Decomposition of a Seasonal Time Series - Additive Model
 - 9.3.3. Multiplicative time series
 - 9.3.4. Decomposition of a Seasonal Time Series - Multiplicative Model
- 9.4. Forecasting Time Series Using Analytical Methods
 - 9.4.1. Linear Trend Model for Additive Time Series
 - 9.4.2. Autoregressive Models (AR)

References

Appendix 1. Index of Basic Tools of SAS Enterprise Guide

Appendix 2. Index of Basic Procedures Used in SAS Enterprise Guide

Appendix 3. Statistics Course Syllabus

Wprowadzenie

W procesie dydaktycznym, obecnie na każdym poziomie nauczania, codziennością prawie staje się korzystanie z różnego rodzaju narzędzi informatycznych począwszy od prostych programów, kończąc na zaawansowanych aplikacjach. Mają one ułatwić, a czasami usprawnić i przyspieszyć proces przetwarzania informacji, proces nauczania i wreszcie proces badawczy. Do takich narzędzi czy systemów wspomagających należy system **SAS**.

SAS jest jednym z wielu funkcjonujących systemów, pakietów komputerowych, które mogą być wykorzystywane w procesie dydaktycznym wyższej uczelni, w tym w procesie nauczania takiego przedmiotu jakim jest statystyka.

Podręcznik *Statystyka od podstaw z systemem SAS* został przygotowany bazując na naszych wcześniejszych dwóch wydaniach podręcznika: *Statystyka od podstaw z systemem SAS*, wersja 8.1.– wydanie pierwsze z 2002 roku oraz *Statystyka od podstaw z systemem SAS*, wersja 9.1. – wydanie drugie z roku 2005.

Punktem wyjściowym do przygotowania podręcznika, były funkcjonujące programy nauczania statystyki na uczelniach ekonomicznych, w tym w Szkole Głównej Handlowej (SGH). Wiodącym podręcznikiem do nauczania statystyki według standardowego programu w SGH (aktualny sylabus programu nauczania wraz z obowiązującą literaturą zamieszcza załącznik 3), jest *Statystyka od podstaw*, J. Joźwiak i J. Podgórskiego, PWE, Wyd. VI, Warszawa 2012.

Zakładamy, że informacje ogólne o systemie **SAS** oraz przykłady zawarte w tym podręczniku wraz z algorytmami rozwiązań mogą posłużyć jako pomoc do ćwiczeń ze statystyki w programie podstawowym statystyki, wspomaganym systemem **SAS**. Zdajemy sobie sprawę, że tekst tego podręcznika prezentuje zaledwie mały wycinek możliwości obu wykorzystanych narzędzi w zakresie „*Statystyki od podstaw*”. Mamy nadzieję, że ocena odbioru tego podręcznika przez użytkowników i wszelkie uwagi pozwolą nam w przyszłości przygotować jego kolejną, ulepszoną i poszerzoną wersję.

Uważamy, że wykorzystanie tak bogatego narzędzia statystycznego, jakim jest **SAS** w procesie dydaktycznym, powinno się zacząć od podstaw statystyki, a następnie przechodzić do coraz to bardziej zaawansowanych procedur, algorytmów i analiz, które mogą stanowić przedmiot wykładów specjalistycznych.

Logika konstrukcji tego podręcznika jest taka, że nie omawiamy procedur statystycznych, ani podstaw teorii statystyki, bo Czytelnik powinien je znaleźć w wielu podręcznikach ze statystyki, w tym m.in. w podręczniku wspomnianych już autorów

J. Józwiak i J. Podgórskiego (2012). Wyjątek od tej zasady stanowią krótkie informacje teoretyczne zawarte we wprowadzeniu do rozdziałów 3, 5, 6, 7 i 9. w zakresie m.in. metody reprezentacyjnej, wnioskowania statystycznego, analizy wariancji i kowariancji oraz szeregów czasowych.

Struktura podręcznika jest następująca: najpierw omówione jest narzędzie wraz z podstawowymi informacjami o możliwościach jego wykorzystania do obróbki i wizualizacji danych wraz z raportowaniem, a następnie w kolejnych rozdziałach podawane są przykłady analizy zaczynając od metody reprezentacyjnej i wybranych schematów losowania próby a kończąc na analizie szeregów czasowych. Z wielu istniejących możliwości, w pracy przedstawiono analizy z wykorzystaniem narzędzia *SAS Enterprise Guide*, wersja 4.3.

Autorzy podręcznika składają serdeczne podziękowania Panu dr Zdzisławowi Piaście, pracownikowi Politechniki Świętokrzyskiej, za bardzo wnikliwą oraz obszerną recenzję wydawniczą, która była bardzo pomocna w przygotowaniu ostatecznej wersji podręcznika.

Przedkładany Czytelnikowi podręcznik jest jednym z wielu podręczników oferowanych studentom SGH, którzy podejmują trud ukończenia obowiązującej ścieżki przedmiotów niezbędnych do uzyskania Certyfikatu „Analityk Statystyczny SAS”.

W dniu 22 lutego 2005 roku w Warszawie, pomiędzy Szkołą Główną Handlową, reprezentowaną przez prof. dr hab. Marka Rockiego – Rektora SGH, a SAS Institute Sp. z o.o., reprezentowaną przez Alicję Wiecką – Dyrektora Zarządzającego, zostało podpisane porozumienie dotyczące nadawania certyfikatu „Analityk Statystyczny SAS”. Koordynatorami Certyfikatu zostali pracownicy SGH: dr hab. Ewa Frątczak i prof. dr Grzegorz Rawicz-Mańkowski z SAS Institute. Aktualnie koordynatorem ze strony SAS Institute jest Dyrektor Zarządzająca SAS Institute Polska - pani Alicja Wiecka.

Otrzymanie Certyfikatu „Analityk Statystyczny SAS” wymaga zaliczenia 210 godz. programu zajęć o tematyce „Statystyczna analiza danych z wykorzystaniem narzędzi SAS”. Program obejmuje 60 godzin wykładu oraz 150 godzin laboratorium komputerowego z następujących przedmiotów:

- I. Przetwarzanie danych w SAS – 30h
- II. Metodologia tworzenia hurtowni danych SAS – 30h
- III. Statystyka od podstaw z wykorzystaniem narzędzi SAS – 15h
- IV. Wielowymiarowa analiza statystyczna z wykorzystaniem narzędzi SAS – 15h
- V. Analiza historii zdarzeń z wykorzystaniem narzędzi SAS – 30h

- VI. Regresja logistyczna z wykorzystaniem narzędzi SAS – 15h
- VII. Analiza i prognozowanie szeregów czasowych z wykorzystaniem narzędzi SAS – 30h
- VIII. Zaawansowane metody analizy statystycznej z wykorzystaniem narzędzi SAS – 45h

Warto podkreślić, że certyfikat „Analityk Statystyczny SAS” jest uznawany nie tylko w Polsce, ale również we wszystkich krajach, w których wykorzystywany jest system **SAS**. Oprócz wspomnianego certyfikatu, na II poziomie studiów realizowana jest specjalność „Analizy Statystyczne i Data Mining” oraz dwa programy studiów podyplomowych: „Analizy Statystyczne i Data Mining w Biznesie” oraz „Akademia Analityka – Analizy Statystyczne i Data Mining w Biznesie”. Więcej na temat oferty programowej znajduje się na stronie Zakładu Analizy Historii Zdarzeń i Analiz Wielopoziomowych Instytutu Statystyki i Demografii SGH (<http://www.sgh.waw.pl/zaklady/zahziaw/>).

Dla większości wykładanych przedmiotów zostały przygotowane i opublikowane autorskie podręczniki. Informacja o naszych publikacjach jest na stronie internetowej Zakładu Analizy Historii Zdarzeń i Analiz Wielopoziomowych Instytutu Statystyki i Demografii Kolegium Analiz Ekonomicznych Szkoły Głównej Handlowej: <http://www.sgh.waw.pl/zaklady/zahziaw/ozakl/publikacje/>.

Warto na zakończenie dodać, iż cała dokumentacja do **SAS 9.2** i **SAS 9.3** jest dostępna na stronach internetowych:

<http://support.sas.com/documentation/onlinedoc/index.html>, w formacie html

<http://support.sas.com/documentation/93/index.html> i PDF

<http://support.sas.com/documentation/92/index.html>

Na stronach <http://support.sas.com/documentation/onlinedoc/code.samples.html> znajdują się kody przykładowych programów (*proc step* i *data step*) opublikowane przez SAS Institute, Cary, USA.