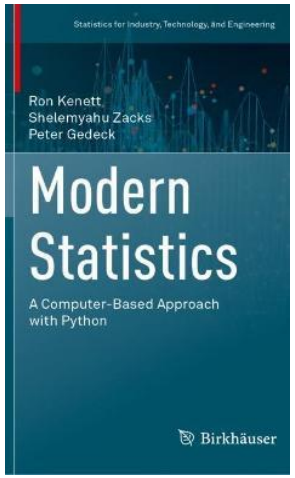
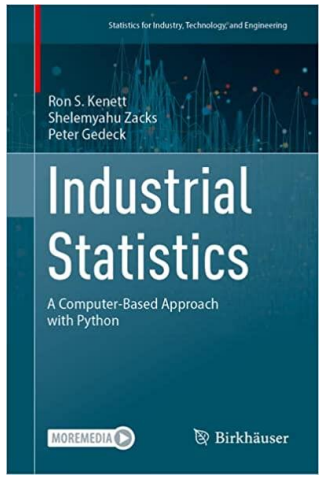
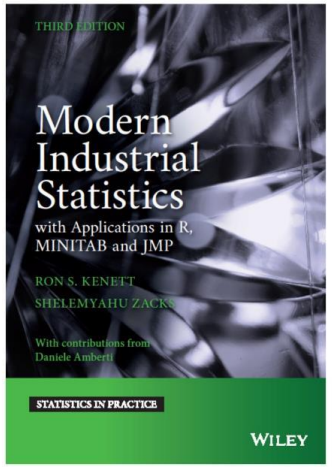
A Biomed Data Analyst Training Program

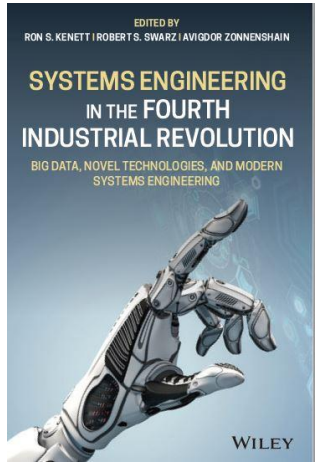
Introduction

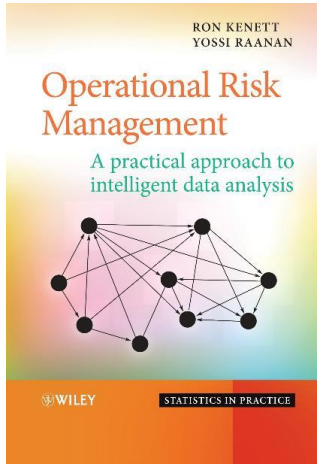
Professor Ron S. Kenett

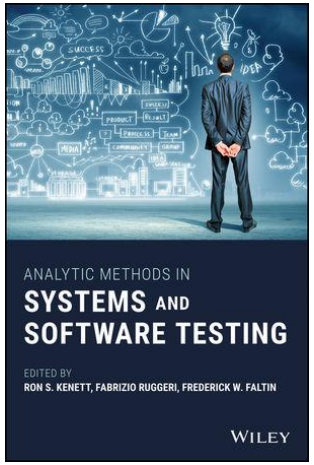












Professor Ron Kenett, Chairman of the *KPA Group*, Senior Research Fellow, *Samuel Neaman Institute,* Technion and chairman of the *Data Science Society*, Israel. Ron authored and co-authored over 250 papers and 18 books on topics ranging from industrial statistics, biostatistics, data science, surveys and statistical methods in healthcare. He was awarded the 2015 Greenfield Medal by the Royal Statistical Society (RSS), the 2018 Box Medal by the European Network for Business and Industrial Statistics (ENBIS), life achievement award from the Israel Society for Quality (2022) and Alan Tayler Memorial Lecture, European Consortium for Mathematics in Industry (2023). He is editor in chief of Wiley's *StatsRef electronic Encyclopedia*, on the advisory board of *Data Centric Engineering* published by Cambridge University Press and the book series *Statistics for Industry, Technology and Engineering*, published by Springer. Ron is Member of the International Board of the UniSR University Centre for Statistics in the Biomedical Sciences (Cussb).

KPA Insights through analytics

UNIVERSITÀ DEGLI STUDI DI TORINO

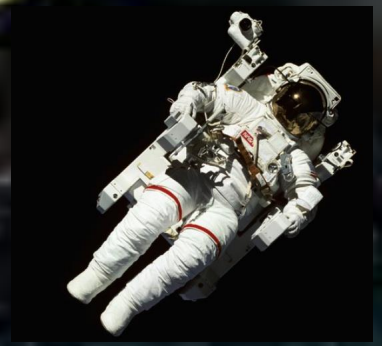
Samuel Neaman Institute for National Policy Research

AEAi ASSOCIATION OF ENGINEERS, ARCHITECTS AND GRADUATES IN TECHNOLOGICAL SCIENCES IN ISRAEL

DATA SCIENCE SOCIETY

The application

The theory



The need

The data analyst competencies - 1

1. Have a genuine desire to solve real problems and help others to solve problems.
2. Be able to help investigators formulate their problem in quantifiable terms.
3. Be able to listen carefully and to ask probing questions.
4. Have a broad knowledge and true understanding of statistical and scientific methods.
5. Be able to adapt existing statistical procedures to novel environments.
6. Be able to locate or develop good statistical procedures in a timely fashion.
7. Be able to keep abreast of developments in statistics.
8. Be willing to meet deadlines, even if it requires substantial extra effort.
9. Be able to understand something about the clients' subject matter and speak a bit of the clients' language.

The data analyst competencies - 2

10. Be a good teacher—much success in consulting depends on being able to help others understand statistical tools, and their strengths and weaknesses.
11. Be willing to settle for a reasonably correct approximate solution, then go on to the next problem.
12. Be able to identify important problems (and thus avoid spending too much time on projects of little significance).
13. Have the confidence to use as simple a procedure as will get the job done, be it design or analysis.
14. Be able to convince others of the validity of a solid solution and see to it that proper action is taken.
15. Be able to use computers effectively and direct others in their use.

The data analyst competencies - 3

16. Be a good problem solver.
17. Be willing to meet clients regularly on their home ground, and take the responsibility to meet and communicate with all members of the working team.
18. Be diplomatic and know when to bend, when to stand firm, and how to help smooth over conflicts among other team members.
19. Be willing to get some experience in the actual collection of the data.
20. Be willing to take the time to check and double-check procedures and results.
21. Be able to communicate effectively in writing as well as orally (this often includes helping clients write their reports as well).
22. Be able to make a good estimate of how much effort will be required to solve the problem without actually having to solve the problem itself.

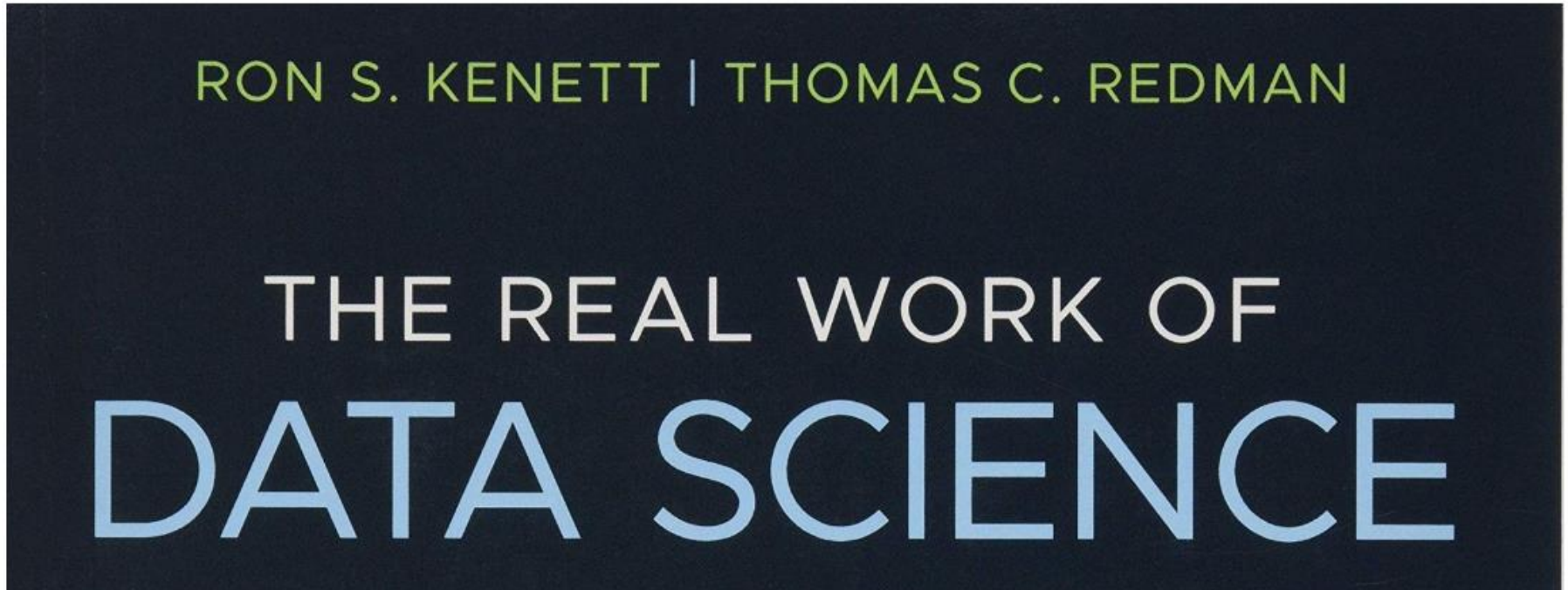
https://ceeds.unimi.it/prof-ron-kenett-short-course-in-data-science/

HOME ABOUT ACTIVITIES CEEDS SEMINARS & SHORT COURSES CEEDS EVENTS NEWS CEEDS DASHBOARD MASTER IN DATA SCIENCE

Prof. Ron Kenett – Short Course in Data Science

jmp STATISTICAL DISCOVERY

www.jmp.com



Lecture Series in Analytics (Sala Laurea) *[Materials : Kenett – Analytics]*

22/01 10.30-13.30

23/01 9.30-12.30

24/01 10.30-13.30

Lecture Series in Causality (Sala Laurea) *[Materials : Kenett – Causality]*

28/01 9.30-12.30

29/01 10.30-13.30

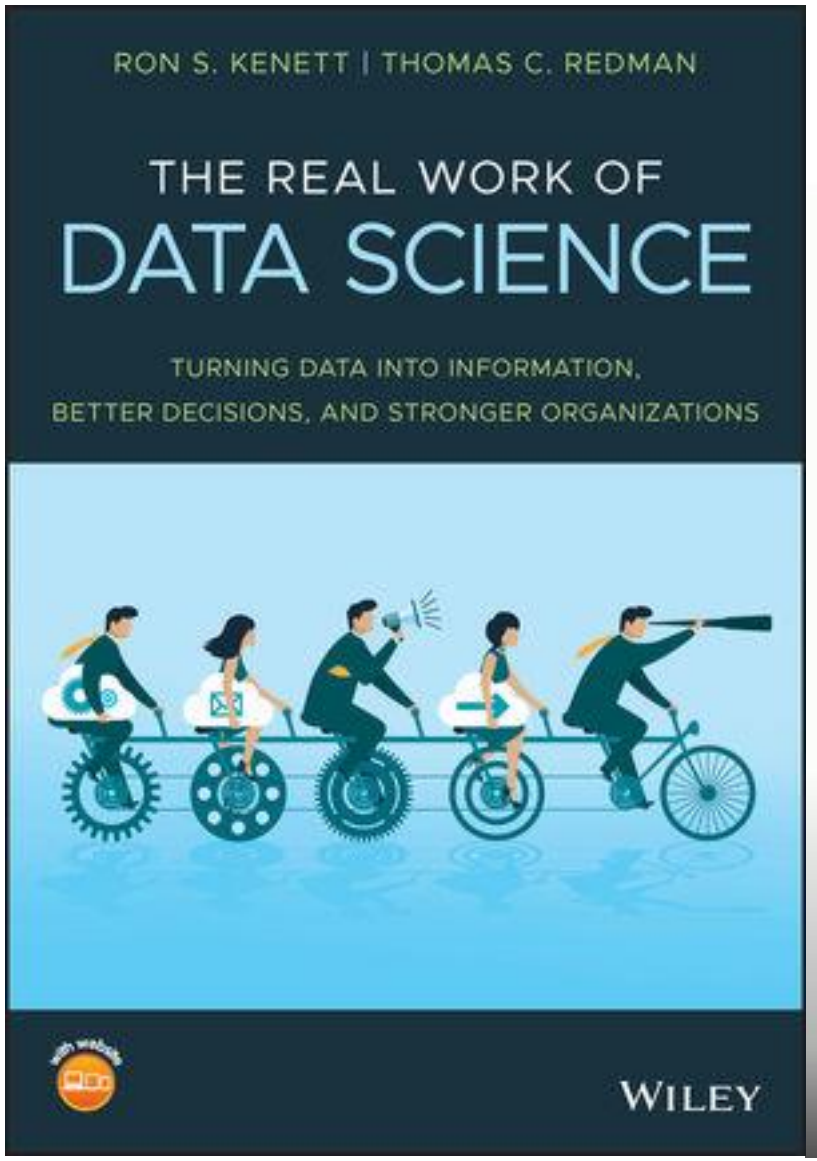
Seminar on 'Statistics at a Crossroad' *[Materials : Kenett – Seminar]*

10.45-11.45 Via Santa Sofia, 9 – aula M203

https://www.linkedin.com/pulse/statistics-crossroad-generating-information-quality-ron-s-kenett/

List of references

Kenett, R. and Zacks, S. (2021) Modern Industrial Statistics: With Applications in R, MINITAB, and JMP, 3rd Edition, Wiley, UK, https://www.wiley.com/en-il/Modern+Industrial+Statistics%3A+With+Applicationsin+R%2C+MINITAB%2C+and+JMP%2C+3rd+Edition-p-9781119714903



https://www.youtube.com/watch?v=gHoeeuuwcPs

https://cran.r-project.org/web/packages/mistat

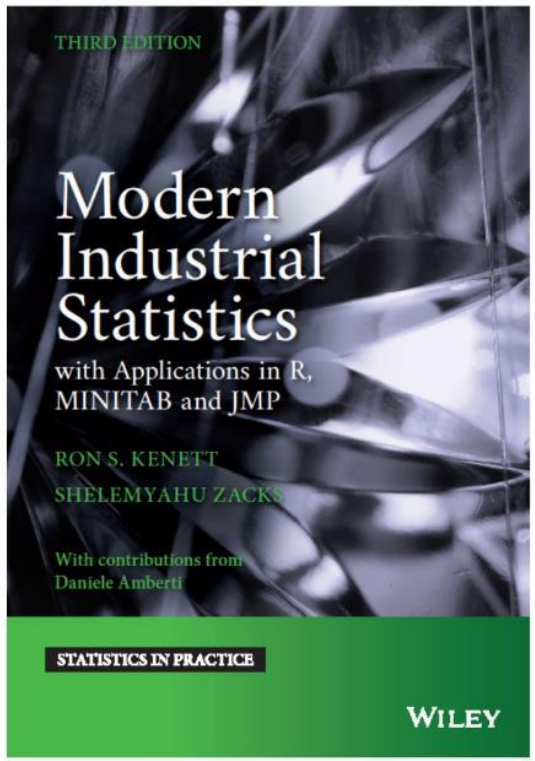
mistat: Data Sets, Functions and Examples from the Book: "Modern Industrial Statistics" by Kenett, Zacks and Amberti

Provide all the data sets and statistical analysis applications used in "Modern Industrial Statistics: with applications in R, MINITAB and JMP" by R.S. Kenett and S. Zacks with contributions by D. Amberti, John Wiley and Sons, 2021, which is a third revised and expanded revision of "Modern Industrial Statistics: Design and Control of Quality and Reliability", R. Kenett and S. Zacks, Duxbury/Wadsworth Publishing, 1998.

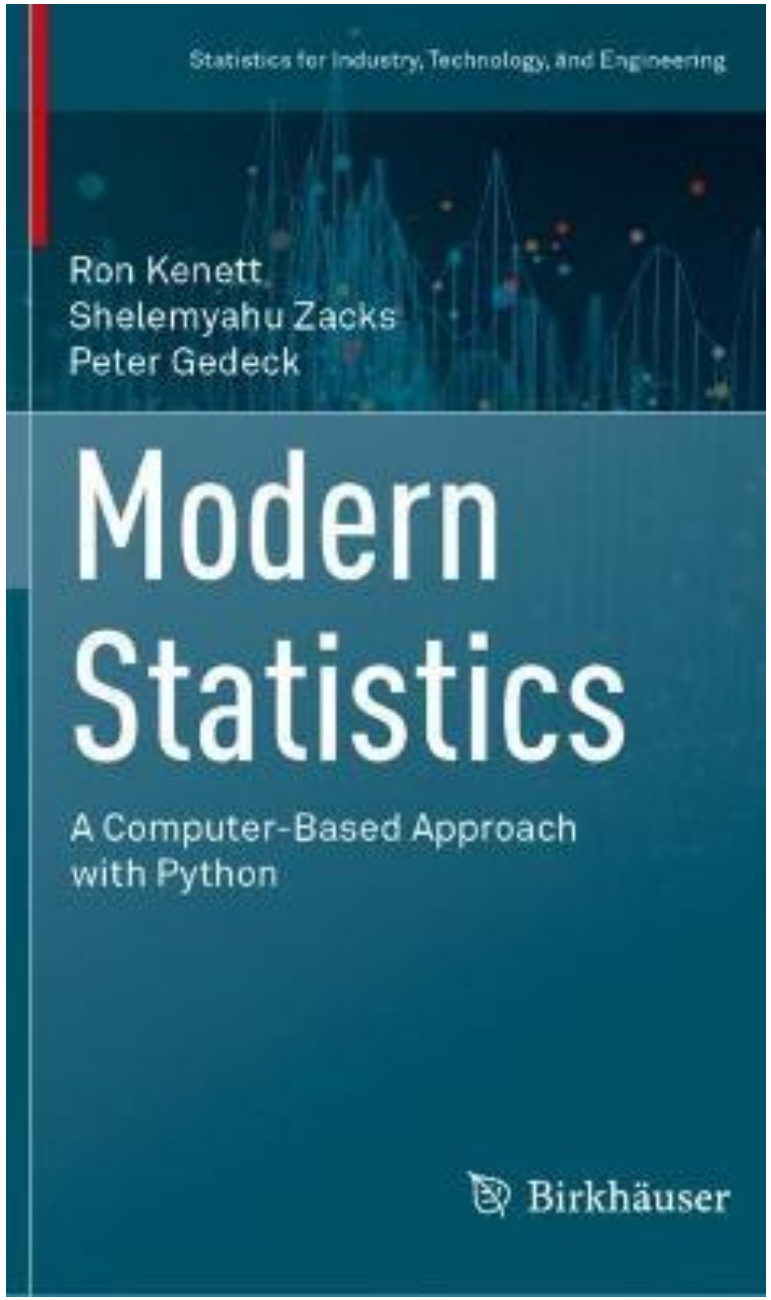
Version:	2.0.4
Depends:	R (≥ 3.5)
Imports:	graphics, methods, stats, utils
Suggests:	e1071, mvtnorm, AcceptanceSampling, boot, car, Dodge, tseries, qcc, DoE.base, FrF2, rsm, LearnBayes, ggplot2, grid, DiceEval, DiceKriging, DiceDesign, DiceView, lhs, survival, rpart, fdapace, randomForest, xgboost
Published:	2023-02-17
Author:	Daniele Amberti
Maintainer:	Daniele Amberti <daniele.amberti at gmail.com>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
NeedsCompilation:	no
Materials:	README ChangeLog
CRAN checks:	mistat results

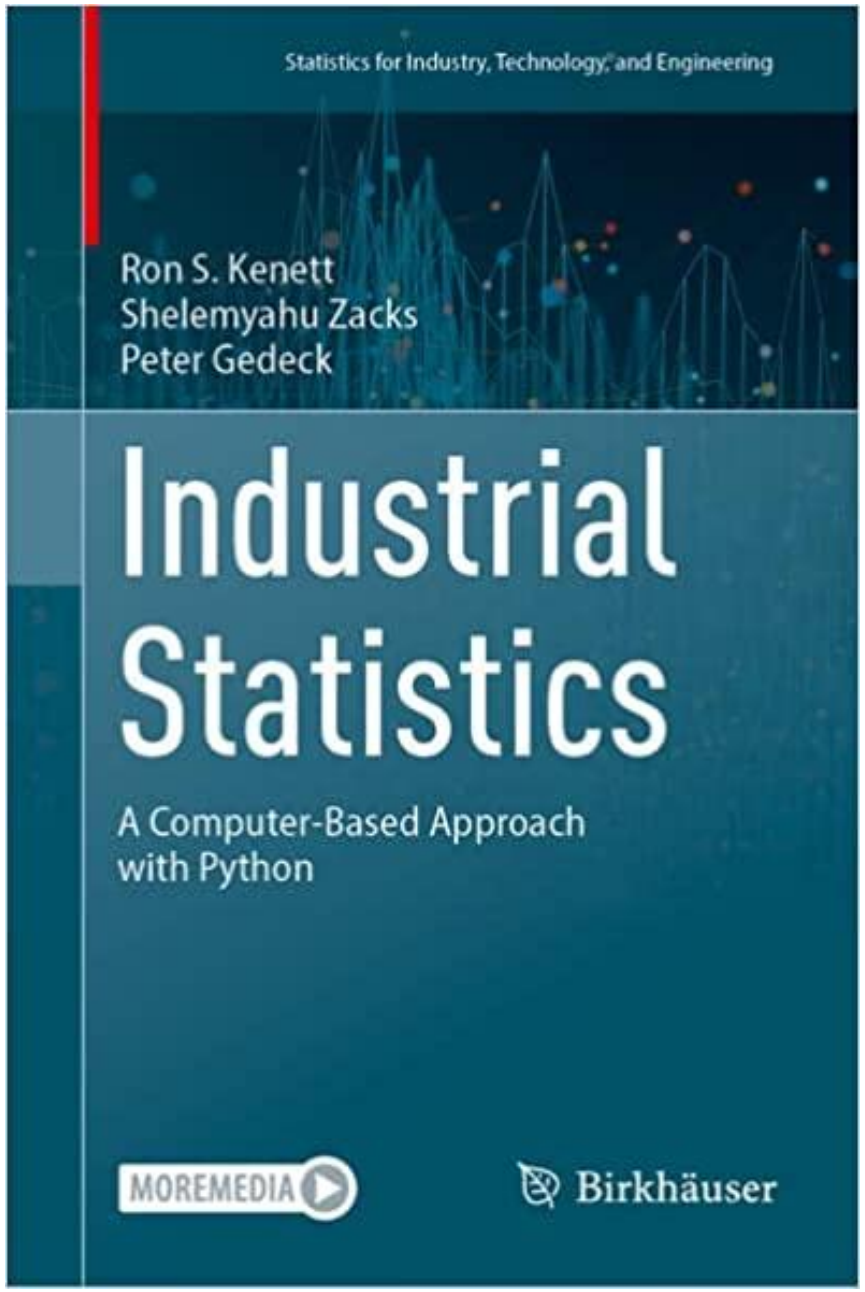
Documentation:

Reference manual: mistat.pdf



https://gedeck.github.io/mistat-code-solutions/ModernStatistics/





https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/

Help

ANACONDA.NAVIGATOR

python™

Home

Environments

Learning

Community

A full Python IDE directly

Documentation

Anaconda Blog

Applications on base (root) Channels

JupyterLab

3.3.2

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch

Notebook

6.4.8

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch

Log in to Overleaf - Overleaf, Onl | mistat-code-solutions | Code rep

gedeck.github.io/mistat-code-solutions/ModernStatistics/

Email link | RonLinks | Facebook | LinkedIn | News Ynet | הארץ | בנק הפועלים – לקו... | Demo: Industry 4.0 | קורונה - לוח בקרה | Wiley StatsRef: Stati... | Other bookmarks

View on GitHub

Modern Statistics

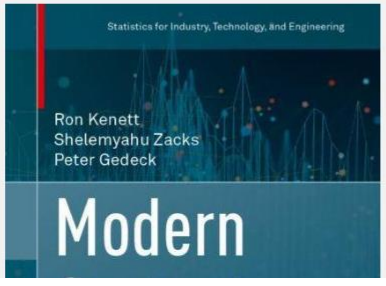
A Computer Based Approach with Python

Home >**README.md**

https://gedeck.github.io/mistat-code-solutions/ModernStatistics/

Test passing | launch binder

Code repository



Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing;
1st edition (September 15, 2022)
ISBN-13: 978-3-031-07565-0 (print)
ISBN-13: 978-3-031-07566-7 (eBook).
Buy at Springer, Amazon, Barnes & Noble

Python notebooks

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

Python packages and Python itself change over time. This can cause warnings or errors. We "Warnings" are f
execution and need to be fixed in order to get results.

If you come across an issue with the code, please follow these steps

- Check the repository (https://gedeck.github.io/mistat-code-solutions/) to see if the code has been upgr
- Check the errata at (https://gedeck.github.io/mistat-code-solutions/ModernStatistics/) and the issue tra
  to see if the issue is known and a solution available.
- Report the problem using the issue tracker at https://github.com/gedeck/mistat-code-solutions/issues

binder

Starting repository: gedeck/mistat-code-solutions/binder-modern-statistics

Your session is taking longer than usual to start!
Check the log messages below to see what is happening.

Binder

https://mybinder.org/v2/gh/gedeck/mistat-code-solutions/binder-modern-statistics?labpath=ModernStatistics%2Fnotebooks%2Findex.ip...

Email link | RonLinks | Facebook | LinkedIn | News Ynet | הארץ | בנק הפועלים – לקו... | Demo: Industry 4.0 | קורונה - לוח בקרה | Wiley StatsRef: Stati... | Other bookmarks

Thanks to Google Cloud, OVH, GESIS Notebooks and the Turing Institute for supporting us!

Donate to mybinder.org!

https://mybinder.org/v2/gh/gedeck/mistat-code-solutions/binder-modern-statistics?labpath=ModernStatistics%2Fnotebooks%2Findex.ipynb

Starting repository: gedeck/mistat-code-solutions/binder-modern-statistics

The tool that powers this page is called BinderHub. It is an open source tool that you can deploy yourself.

Build logs view raw show

Python notebooks

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

The code needs to be executed in sequence.

Python packages and Python itself change over time. This can cause warnings or errors. We "Warnings" are for information only and can usually be ignored. "Errors" will stop execution and need to be fixed in order to get results.

If you come across an issue with the code, please follow these steps

- Check the repository (https://gedeck.github.io/mistat-code-solutions/) to see if the code has been upgraded. This might solve the problem.
- Check the errata at (https://gedeck.github.io/mistat-code-solutions/ModernStatistics/) and the issue tracker at https://github.com/gedeck/mistat-code-solutions/issues to see if the issue is known and a solution available.
- Report the problem using the issue tracker at https://github.com/gedeck/mistat-code-solutions/issues
- Paste the error message into Google and see if someone else already found a solution

https://gedeck.github.io/mistat-code-solutions/ModernStatistics/

https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/

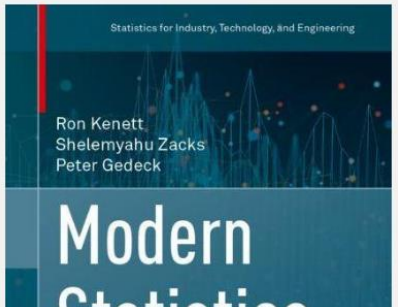
Modern Statistics

A Computer Based Approach with Python

Home >**README.md**

Test passing launch binder

Code repository



Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3-031-07565-0 (hardcover)
ISBN-13: 978-3-031-07568-1 (softcover)
ISBN-13: 978-3-031-28482-3 (eBook).
Buy at Springer, Amazon, Barnes & Noble

Industrial Statistics

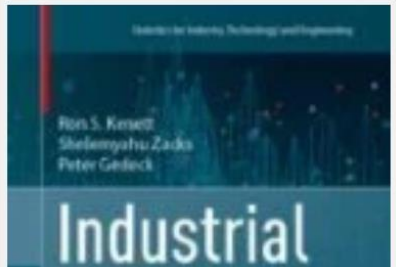
A Computer Based Approach with Python

Home >**README.md**

Test passing

launch binder

Code repository



Industrial Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (August 5, 2023)
ISBN-13: 978-3-031-28481-6 (hardcover)
ISBN-13: 978-3-031-28484-7 (softcover)
ISBN-13: 978-3-031-28482-3 (eBook).
Buy at Springer, Amazon.

https://gedeck.github.io/mistat-code-solutions/ModernStatistics/

Modern Statistics: A Computer Based Approach with Python is a companion volume to the book Industrial Statistics: A Computer Based Approach with Python.

This part of the repository contains:

- `notebooks` : Python code of individual chapters in Jupyter notebooks - download all as notebooks.zip
- `code` : Python code for solutions as plain Python files - download all as code.zip
- `solutions manual` : Solutions_Modernstatistics.pdf: solutions of exercises
- `solutions` : Python code for solutions in Jupyter notebooks - download all as solutions.zip
- `all` : zip file with all files combined - download all as all.zip
- `datafiles` : zip file with all data files - download all as data_files.zip - the `mistat` package gives you already access to all datafiles, you only need to download this file if you want to use it with different software

All the Python applications referred to in this book are contained in a package called `mistat` available for installation from the Python package index https://pypi.org/project/mistat/. The `mistat` packages is maintained in a GitHub repository at https://github.com/gedeck/mistat.

Try the code

You can explore the code on Binder launch binder.

Installation instructions

Instructions on installing Python and required packages are here.

These Python packages are used in the code examples of *Modern Statistics*:

- mistat (for access to data sets and additional functionality)
- numpy
- scipy

https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/

Industrial Statistics: A Computer Based Approach with Python is a companion volume to the book Modern Statistics: A Computer Based Approach with Python.

This part of the repository contains:

- `notebooks` : Python code of individual chapters in Jupyter notebooks - download all as notebooks.zip
- `code` : Python code for solutions as plain Python files - download all as code.zip
- `solutions manual` : Solutions_IndustrialStatistics.pdf: solutions of exercises
- `solutions` : Python code for solutions in Jupyter notebooks - download all as solutions.zip
- `all` : zip file with all files combined - download all as all.zip
- `datafiles` : zip file with all data files - download all as data_files.zip - the `mistat` package gives you already access to all datafiles, you only need to download this file if you want to use it with different software

All the Python applications referred to in this book are contained in a package called `mistat` available for installation from the Python package index https://pypi.org/project/mistat/. The `mistat` packages is maintained in a GitHub repository at https://github.com/gedeck/mistat.

Table of contents (with sample excerpts from chapters)

https://gedeck.github.io/mistat-code-solutions/ModernStatistics/

https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/

Table of contents

Introductory videos

Chapter 1: Analyzing Variability: Descriptive Statistics

The chapter focuses on statistical variability and on various methods of analyzing random data. Random results of experiments are illustrated with distinction between deterministic and random components of variability. The difference between accuracy and precision is explained. Frequency distributions are defined to represent random phenomena. Various characteristics of location and dispersion of frequency distributions are defined. The elements of exploratory data analysis are presented.

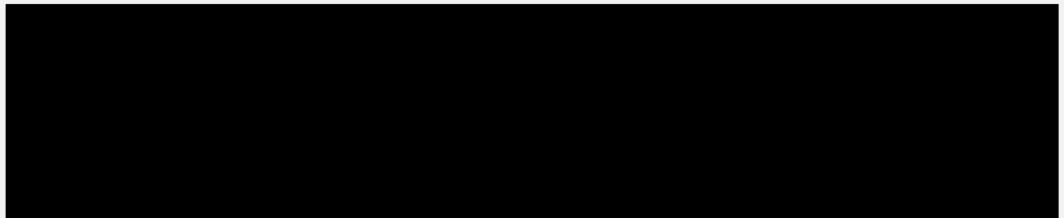


Table of contents (with sample excerpts from chapters)

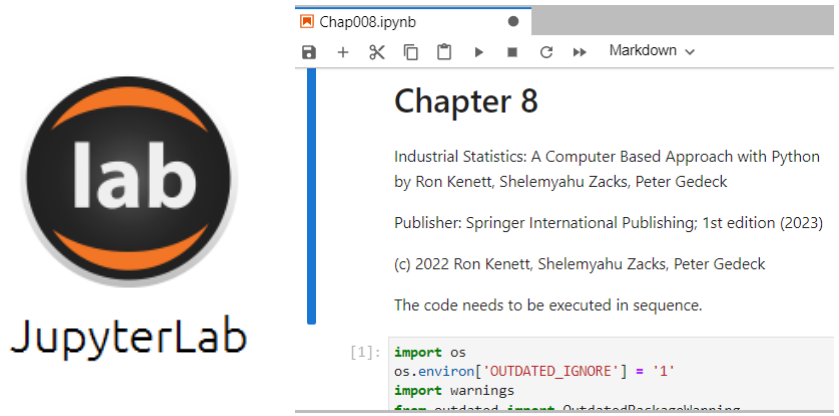
Installation instructions

Instructions on installing Python and required packages are here.

These Python packages are used in the code of *Industrial Statistics*:

- mistat (for access to data sets and additional functionality)
- numpy
- pandas
- scipy
- statsmodels
- matplotlib==3.6.0
- seaborn
- pingouin
- lifelines


(1) Facebook
https://user-images.githubuserco
https://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4?fbclid=IwAR08kOITz...
Email link
RonLinks
Facebook
LinkedIn
News Ynet
הארץ
בנק הפועלים – לקו...
Demo: Industry 4.0
קורונה - לוח בקרה
Wiley StatsRef: Stati...
Other bookmarks
https://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4?fbclid=IwAR08kOITzOhBUIYZhyclkrh_p7Ad9q53rnrwUicOmiUBHsbd4TwJgzkHepo
0:01 / 4:01



https://link.springer.com/book/10.1007/978-3-031-07566-7#toc

Chapter 6

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

The code needs to be executed in sequence.

Time Series Analysis and Prediction

Ron Kenett, Shelemyahu Zacks, Peter Gedeck
Pages 329-360

In [1]:
```
import os
os.environ['OUTDATED_IGNORE'] = '1'
import warnings
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

Time Series Analysis and Prediction

In [2]:
```
import datetime
import statsmodels.formula.api as smf
from statsmodels.tools.sm_exceptions import ValueWarning
import pandas as pd

import random
import numpy as np
import pingouin as pg
from scipy import stats
import matplotlib.pyplot as plt
import mistat
```

Chapter 7

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

The code needs to be executed in sequence.

In [1]:
```
import warnings
import os
os.environ['OUTDATED_IGNORE'] = '1'
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

Modern analytic methods: Part I

In [2]:
```
import warnings
import random
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.impute import SimpleImputer
from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
```

Modern Analytic Methods: Part I

Ron Kenett, Shelemyahu Zacks, Peter Gedeck
Pages 361-393

Chapter 8

Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)
ISBN-13: 978-3031075650

The code needs to be executed in sequence.

In [1]:
```
import os
os.environ['OUTDATED_IGNORE'] = '1'
import warnings
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

Modern analytic methods: Part II

In [2]:
```
import networkx as nx

import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import mistat
```

Modern Analytic Methods: Part II

Ron Kenett, Shelemyahu Zacks, Peter Gedeck
Pages 395-419

A Biomed Data Analyst Training Program

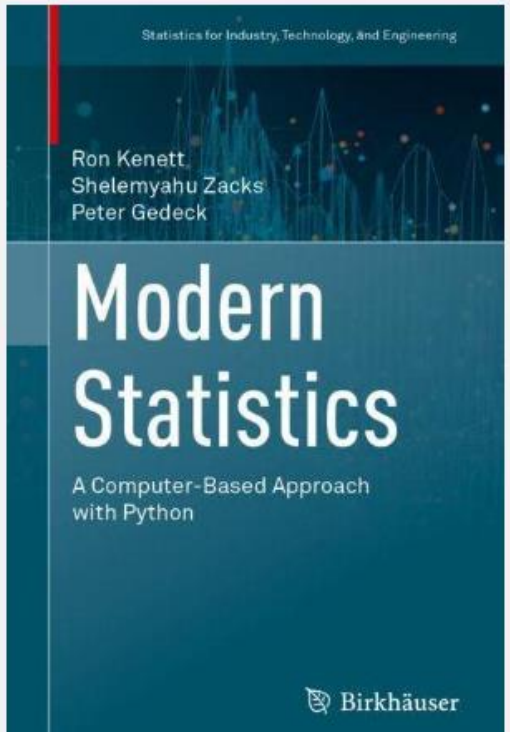
Modern Statistics: A Computer Based Approach with Python

Home >**README.md**

Test passing launch binder

https://gedeck.github.io/mistat-code-solutions/BioMed_DataAnalyst_Course/

A Biomed Data Analyst Training Program



Modern Statistics: A Computer Based Approach with Python
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (Septemb
2022)
ISBN-13: 978-3-031-07565-0 (hardcover)
ISBN-13: 978-3-031-07568-1 (softcover)
ISBN-13: 978-3-031-28482-3 (eBook).
Buy at Springer, Amazon, Barnes & Noble

Errata: See known errata here

Slides

1. Introduction
2. Data types and data integration
3. Supervised learning
4. Model performance
5. Time series
6. Data visualization
7. Causality and experimental design

https://gedeck.github.io/mistat-code-solutions/BioMed_DataAnalyst_Course/

Code and data files

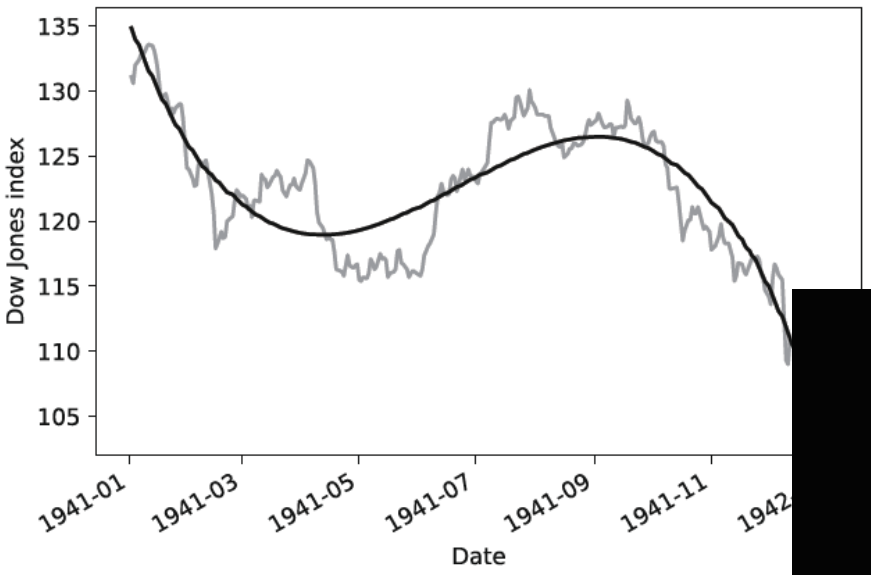
This part of the repository contains:

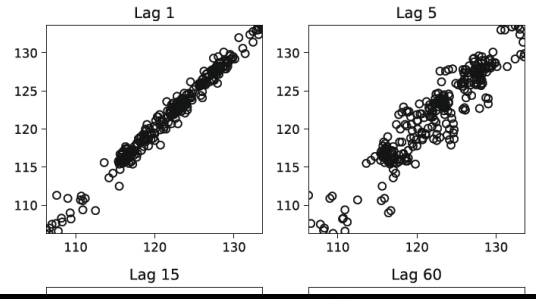
- `notebooks` : Python code of individual chapters in Jupyter notebooks - download all as notebooks.zip

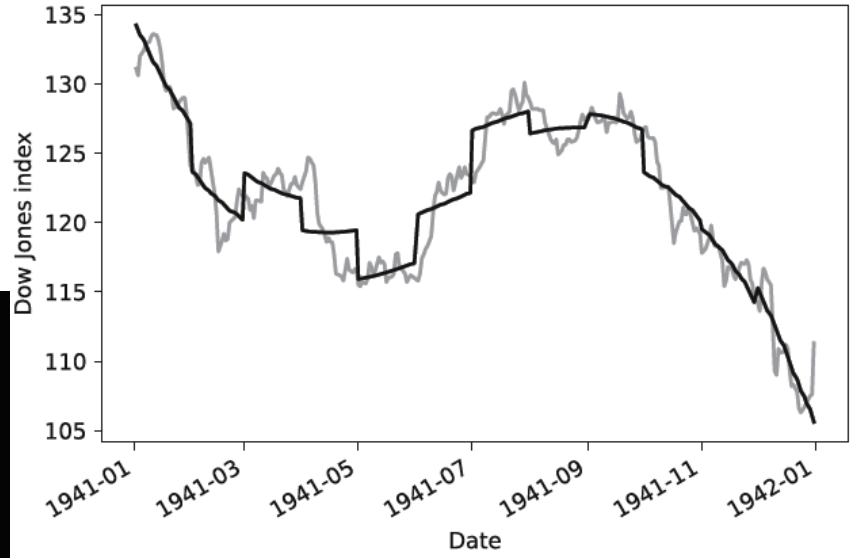
The Python package `mistat` contains datafiles and utility functions referred to in the Modern Statistics book. It is available for installation from the Python package index https://pypi.org/project/mistat/. The `mistat` packages is maintained in a GitHub repository at https://github.com/gedeck/mistat.

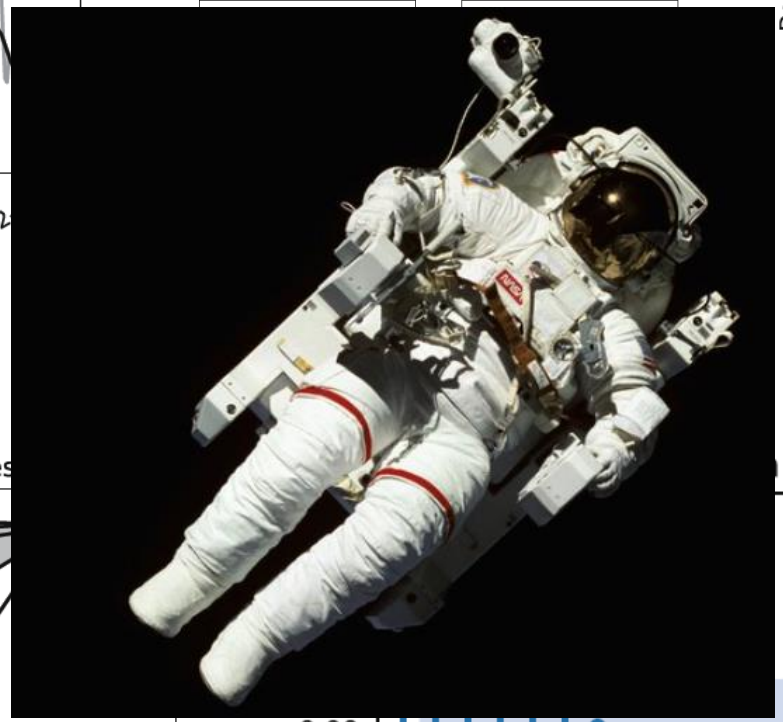
Try the code

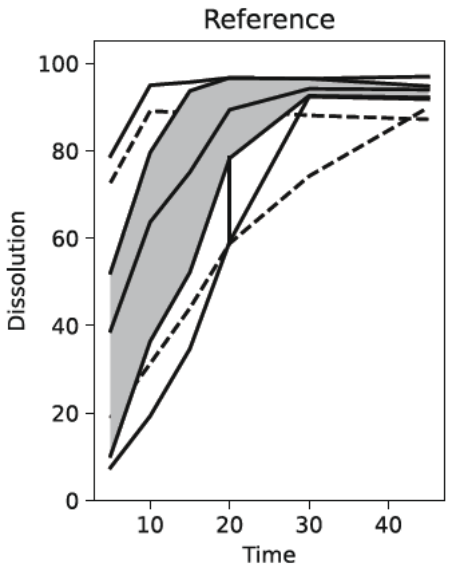
You can explore the code on Binder launch binder.

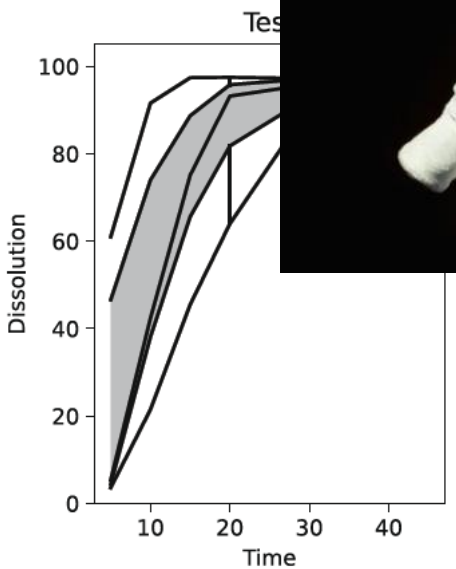

Dow Jones index
Date
1941-01
1941-03
1941-05
1941-07
1941-09
1941-11

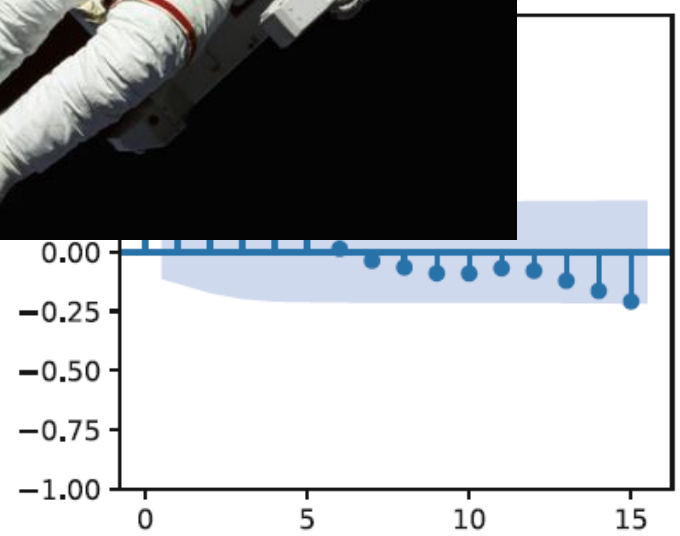

Lag 1
Lag 5
Lag 15
Lag 60

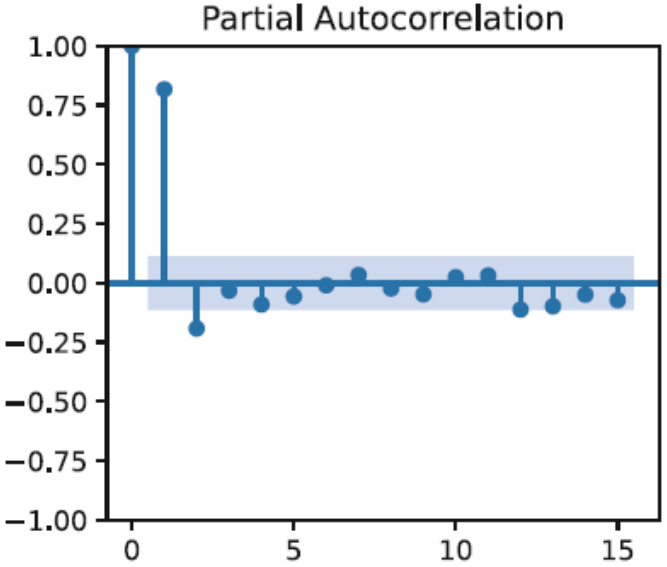

Dow Jones index
Date
1941-01
1941-03
1941-05
1941-07
1941-09
1941-11
1942-01




Reference
Dissolution
Time


Dissolution
Time




Partial Autocorrelation